The QR Algorithm
Past and Present

Marco Latini

Applied & Computational Mathematics

February 26, 2004

Outline of the talk

1. Introduction to eigenvalue problems
2. Power and inverse power iteration
3. Explicit shifted and non-shifted QR Algorithm
4. Implicit multi-shift QR algorithm
5. The QZ algorithm for generalized eigenvalue problems

Eigenvalue algorithms are iterative

- To every polynomial $p(x)$ there corresponds a companion matrix C_p

$$
\begin{aligned}
p(x) &= x^n - a_{n-1}x^{n-1} - \ldots - a_1 x - a_0 \\
C_p &= \begin{bmatrix} a_{n-1} & a_{n-2} & \ldots & a_1 & a_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}
\end{aligned}
$$

- Therefore by Abel's proof (1824) that there are no algebraic formulae for the roots of a general polynomial with $n > 4$ we cannot have a non-iterative eigenvalue algorithm
- This is in contrast to solving linear systems

The Power Iteration is based on a simple observation

- Let A be a matrix with a complete set of eigenvalues and eigenvectors pairs (λ_i, q_i)
- Let the eigenvalues be ordered such that $|\lambda_1| > |\lambda_2| > ... > |\lambda_n|$
- Then for any vector u, $A^k u = \gamma_1 \lambda_1^k q_1 + \sum_i \gamma_i \lambda_i^k q_i$.
- As $k \to \infty$ the term containing λ_1 will dominate and $A^k u$ approaches a multiple of the dominant eigenvector q_1

Based on the above we can write the Power iteration

- Choose $v^{(0)}$ such that $||v^{(0)}|| = 1$, then for $k = 1, 2, ...$ we do the following steps

$$\begin{aligned} w &= Av^{(k-1)} \\ v^{(k)} &= \frac{w}{||w||} \\ \lambda^{(k)} &= (v^{(k)})^T A v^{(k)}. \end{aligned}$$

- The rate of convergence to the eigenvectors

$$\begin{aligned} ||v^{(k)} - q_1|| &= O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right) \\ |\lambda^{(k)} - \lambda_1| &= O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right). \end{aligned}$$

- The power method goes back to a work by Muntz of 1913. The method was then used by Hotelling in 1933 and by Aitken in 1937.

- Before the QR era the power method was the standard tool to compute eigenvalues.

The Inverse power iteration solves some problems with the power iteration

- Suppose μ is NOT an eigenvalue of A
- The matrix $B = (A - \mu I)^{-1}$ has the same eigenvectors as A and the eigenvalues are $(\lambda_j - \mu)^{-1}$
- If we choose a value of μ close to an eigenvalue then $(\lambda_j - \mu)^{-1}$ will be an enormous number
- We can thus apply the power iteration to the matrix $(A - \mu I)^{-1}$ we will converge rapidly to q_j and to $(\lambda_j - \mu)^{-1}$.

Based on the above we can write the inverse power iteration

- Choose $v^{(0)}$ such that $||v^{(0)}|| = 1$ and for $k = 1, 2, ...$ we have

$$\begin{aligned} w &= (A - \mu I)^{-1} v^{(k-1)} \\ v^{(k)} &= \frac{w}{||w||} \\ \lambda^{(k)} &= (v^{(k)})^T A v^{(k)}. \end{aligned}$$

- The convergence for this method are as follows

$$\begin{aligned} \left|\left| v^{(k)} - q_j \right|\right| &= O\left(\left| \frac{\mu - \lambda_j}{\mu - \lambda_k} \right|^k \right) \\ \left| \lambda^{(k)} - \lambda_j \right| &= O\left(\left| \frac{\mu - \lambda_j}{\mu - \lambda_k} \right|^{2k} \right), \end{aligned}$$

- λ_j is the closest eigenvalue to μ and λ_k is the second closest.

- The inverse power method is due to Wieland in 1944.

The Rayleigh Quotient Iteration is a natural union of the power and inverse power iterations

- The main insight is to use the estimates from the power iteration as shifts for the inverse power method
- Choose $v^{(0)}$ as any vector such that $||v^{(0)}|| = 0$. Let $\lambda^{(0)} = (v^{(0)})^T A v^{(0)}$, then for $k = 1, 2, ...$ we have

$$
\begin{aligned}
w &= (A - \lambda^{(k-1)} I)^{-1} v^{(k-1)} \\
v^{(k)} &= \frac{w}{||w||} \\
\lambda^{(k)} &= (v^{(k)})^T A v^{(k)}.
\end{aligned}
$$

- The convergence for this method is given by

$$
\begin{aligned}
||v^{(k+1)} - q_j|| &= O\left(\left|\left|v^{(k)} - q_j\right|\right|^3\right) \\
\left|\lambda^{(k+1)} - \lambda_j\right| &= O\left(\left|\lambda^{(k)} - \lambda_j\right|^3\right).
\end{aligned}
$$

- The Rayleigh quotient iteration is due to Ostroski in 1958

The Explicit QR Algorithm with Shift can be easily written but it is harder to understand

- Let A be any matrix (real or complex).
- Let $A_0 = A$, for $k = 1, 2, 3, \ldots$ then given shifts κ_i we have

$$\begin{aligned} A_k - \kappa_k I &= Q_k R_k \\ A_{k+1} &= R_k Q_k + \kappa_k I. \end{aligned}$$

- Notice that the iterates satisfy

$$\begin{aligned} A_{k+1} &= R_k Q_k + \kappa_k I \\ &= Q_k^*(A_k - \kappa_k I) Q_k + \kappa_k I \\ &= Q_k^* A Q_k, \end{aligned}$$

  and therefore are related through a similarity transformation.
- The algorithm works because of its connection to the inverse power method with shift κ_k.

The connection between QR and the inverse power method goes through the Schur factorization

- The main idea for QR is related to deflation of a Schur decomposition in the presence of an eigenvalue, eigenvector pair (λ, q)
- Let $Q = (Q_p q)$ be unitary. Then

$$Q^* A Q = \begin{pmatrix} Q_p^* A Q_p & Q_p^* A q \\ q^* A Q_p & q^* A q \end{pmatrix} = \begin{pmatrix} B & h \\ g & \mu \end{pmatrix},$$

where

$$\begin{aligned} g &= q^* A Q_p = \lambda q^* Q_p = 0 \\ \mu &= q^* A q = \lambda q^* q = \lambda. \end{aligned}$$

We want to see Schur and inverse iteration in QR algorithm

- Consider the following decomposition in a step of QR

$$\begin{pmatrix} Q_p^* \\ q^* \end{pmatrix} (A - \kappa I) = \begin{pmatrix} R_p \\ r^* \end{pmatrix}.$$

  where $r^* = r_{nn} e_n^*$ and $q^* = r_{nn} e_n^* (A - \kappa I)^{-1}$.

- Thus the last column of the matrix Q (which we call q^*) is the result of the inverse power method with shift κ applied to the vector e_n^*.

- For the decomposition of $A = \begin{pmatrix} B & h \\ g^* & \mu \end{pmatrix}$. the analysis suggests that applying QR is like a step of Schur factorization with the eigenvalue, eigenvector pair (μ, e_n)

- $\mu = e_n^* A e_n$ is a good candidate for a shift and in fact this is called the Rayleigh quotient shift.

To show convergence we need to show that $g \to 0$ and $\mu \to \lambda$

- Consider the following partition

$$A - \kappa I = \begin{pmatrix} B - \kappa I & h \\ g^* & \mu - \kappa \end{pmatrix} \quad = \quad QR = \begin{pmatrix} P & f \\ e^* & \pi \end{pmatrix} \begin{pmatrix} S & r \\ 0 & \rho \end{pmatrix}$$

$$RQ = \begin{pmatrix} S & r \\ 0 & \rho \end{pmatrix} \begin{pmatrix} P & f \\ e^* & \pi \end{pmatrix} \quad = \quad \hat{A} - \kappa I = \begin{pmatrix} \hat{B} - \kappa I & \hat{h} \\ \hat{g}^* & \hat{\mu} - \kappa \end{pmatrix}$$

- Notice that from $||e||^2 + \pi^2 = ||f||^2 + \pi^2 = 1$ we have that $||e|| = ||f||$
- We want to show that e is small when g is small. We get that $g^* = e^* S$. Let $\sigma = ||S^{-1}||$ and $\rho = f^* h + \pi(\mu - \kappa)$. We obtain

$$\begin{aligned} ||e|| &\leq \sigma\, ||g|| \\ ||\rho|| &\leq \sigma ||g|| ||h|| + |\mu - k| \end{aligned}$$

- From the result for ρ and for e we obtain

$$||\hat{g}|| = ||\rho e^*|| \leq \sigma^2 ||h|| ||g|| + \sigma(\mu - \kappa)||g||.$$

- For the bounding $h_k \leq \eta$ and $\sigma_k \leq \sigma$ then

$$||\hat{g}|| \leq \sigma^2 \eta ||g||^2$$

- For A_0 hermitian, then $h_k = g_k$ and we have that

$$||\hat{g}|| \leq \sigma^2 ||g||^3.$$

The QR Algorithm is also a Power Iteration

- As the shifted QR algorithm is applied, better results are observed
- This is a manifestation of the Wilkinson theorem (1965) for un-shifted QR algorithm.
- **Theorem**: Let $A = X\Lambda X^{-1}$ where the matrix possess a complete set of distinct eigenvalues. Let $X^{-1} = LU$ and $X = QR$. If A_k, the kth step in the QR iteration has decomposition $A_k = Q_k R_k$ then there exist diagonal matrices D_k where $|D_k| = I$ such that $Q_k D_k \to Q$.
- Thus the shifted QR algorithm is like applying the un-shifted QR algorithm to the matrix $A - \kappa_k I$.

The Origin of the QR algorithm dates to 1961

- Rutishauser in 1955 was the first to propose what is now known as the LU algorithm to determine the eigenvalues of a matrix. He proposed that for a tridiagonal matrix T one could do an LU factorization and multiply back in reverse order. In 1958 Rutishauser introduced a shift to speed the convergence. However, this algorithm is unstable if the matrices are not positive definite.

- In 1961 Kublanovskaya and Francis independently proposed to substitute the stable QR decomposition instead of the LU decomposition. Francis went on to propose the shift and the reduction to Hessenberg form.

- The connection between the QR algorithm and the inverse power iteration is first mentioned in Stewart 1973. The connection with the power method was known right from the beginning.

- The overall stability theory for the QR algorithm was given by Wilkinsons in 1965 in his work “the algebraic eigenvalue problem”.

The QR algorithm is very expensive if we do not go to an upper-Hessenberg form

- It is now to focus on the cost of the QR algorithm.
- The algorithm requires a decomposition $A = QR$ which takes $O(n^3)$ operations. Because this needs to be repeated for each eigenvalue we obtain an overall cost of $O(n^4)$ which is prohibitively expensive.
- The cure to this problem is the transformation of the matrix to upper-Hessenberg form H. When this is done the total cost of the algorithm is $O(n^3)$.

Householder transformations can reduce A to upper Hessenberg form via similarity transformations

- Householder reflections were introduced by Householder in 1958. However, their applications to eigenvalue problems is due to Wilkins 1960.
- We defined a Householder reflection as

$$H = I - uu^*,$$

  where $||u|| = \sqrt{2}$.
- A householder matrix is Hermitian and unitary.
- Householder reflections are used to introduce zeros into vectors (or matrices).
- Consider the following application. Choose $u = (a + e_1)/\sqrt{1 + a_1}$ for any vector a such that $||a|| = 1$. Then when we apply the Householder reflection we obtain $Ha = -e_1$.

Givens rotations are used for the QR factorization

- Givens rotations are defined as follows

$$P = \begin{pmatrix} c & s \\ -\bar{s} & \bar{c} \end{pmatrix}$$

where c and s are the cosine and sine of a rotation by θ degrees.

- Givens rotations are used to introduce a zero in a two component vector such that

$$P \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} g \\ 0 \end{pmatrix},$$

where $g = \nu a/|a|$, $\nu = \sqrt{a^2 + b^2}$ and $c = |a|/|\nu|$ and $s = a\bar{b}/(|a|\nu)$.

Many techniques can be used to accelerate convergence of the QR algorithm

- Deflation
- Aggressive deflation (due to Francis 1961, Watkins 1995)
- Wilkinsons shift defined as the eigenvalue of the lower 2×2 matrix that is closest to the a_{nn} value
- Ad hoc shifts
- Balancing (due to Osborne 1960)

The Implicitly shifted QR algorithm improves on the explicit QR algorithm

- Complex arithmetic is expensive and can be avoided via the “real” Schur decomposition
- To remove complex values we apply a double shift every time the Wilkinsons shifts produces a pair of complex conjugate eigenvalues

$$QR = (H - \kappa I)(H - \bar{\kappa} I).$$

- Now $(H - \kappa I)(H - \bar{\kappa} I) = H^2 - 2Re(\kappa)H + |\kappa|^2 I$ is a real matrix

Outline of the implicit shifted QR algorithm

1. We determine the first column c of $C = H^2 - 2Re(\kappa)H + |\kappa|^2 I$.
2. Next let Q_0 be a Householder transformation such that $Q_0^* c = \sigma e_1$.
3. Next let $H_1 = Q_0^* A Q_0$.
4. Use Householder transformations to reduce H_1 into upper Hessenberg form and call it $\hat{H}$. Call Q_1 the accumulated transformations.
5. Set $\hat{Q} = Q_0 Q_1$.
6. To show that this algorithm works we need to show that $\hat{Q}$ and $\hat{H}$ are the matrices that one would have obtained using the explicit QR algorithm.
7. We also need to show that c and the reduction of H_1 to upper Hessenberg can be done rapidly

The Implicit Q theorem can be used to show that the implicit QR algorithm works

- The implicit Q theorem states that for a matrix A of order n, let $H = Q^* A Q$ be a reduction of A to Hessenberg form. If the elements in the lower diagonal of H are non-zero (that is the H is un-reduced) then Q and H are uniquely determined by the first or last column of Q.
- First perform the QR decomposition of the matrix C as follows

$$(c C_p) = (q Q_p) \begin{pmatrix} \rho & r \\ 0 & R_p \end{pmatrix}.$$

- The Q matrix in the QR decomposition of C is the matrix that we want to show to be equal to $\hat{Q}$
- Since $c = q\rho$ and $c = \sigma q^{(0)}$ then the first column of Q_0 and the first column of $\hat{Q}$ are up to scaling the same.

The implicit QR is far superior to the explicit QR algorithm

- The computation of c takes $O(1)$ operations because only the first three components of c are non-zero
- The reduction of H_1 to upper Hessenberg can be done in $O(n^2)$
- Total cost for implicit QR for k iterations is

$$2kn^3\text{additions} + 2kn^3\text{multiplications}$$

- Total cost for explicit QR for k iterations is

$$8kn^3\text{additions} + 12kn^3\text{multiplications}$$

- Further notice that one step of implicit $=$ 2 steps of explicit

We conclude with the generalized eigenvalue problem

- The generalized eigenvalue problem determines non-trivial solutions to the system $Ax = \lambda Bx$
- The pencil (A, B) is regular if $\det(A - \lambda B)$
- Schur form: There exist a pair of U and V such that $S = U^*AV$ and $T = U^*BV$ are triangular for the regular pencil (A, B)
- Real Schur form: $S = U^*AV$ and $T = U^*BV$ are quasi-triangular matrices
- Hessenberg-Tridiagonal form: Procedure applied to a regular pencil (A, B) reduces A to upper Hessenberg form and B to triangular form

The QZ Algorithm of Moler and Stewart (1973) solves the generalized eigenvalue problem

- The main idea is to use the doubly shifted implicit QR on the matrix $C = AB^{-1}$ which is an upper Hessenberg matrix.
- Find the first column of C.
- Let H be the Householder transformation such that $Q_0^* c = \sigma e_1$.
- Find Q and Z such that $Q^*(HA, HB)Z$ is again in Hessenberg-Triangular form.
- This algorithm works because $\hat{C} = Q^* H^* C H Q$, therefore we are applying a double step of the implicit QR algorithm on C.
- Elements in the sub-diagonal of C will converge to zero
- This means that either A is becoming triangular or if the elements or b become zero then we have found an infinite eigenvalue and we can deflate the problem.

The Power Iteration

- hello