# Numerical Analysis – Lecture 7[1]

**Revision 2.14** (Givens transformations). The notation $\Omega^{(i,j)}$ denotes an $n \times n$ matrix whose elements are those of the identity matrix, except that

$$\Omega^{(i,j)}_{i,i} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}}, \qquad \Omega^{(i,j)}_{i,j} = \frac{\beta}{\sqrt{\alpha^2+\beta^2}},$$
$$\Omega^{(i,j)}_{j,i} = -\frac{\beta}{\sqrt{\alpha^2+\beta^2}}, \qquad \Omega^{(i,j)}_{j,j} = \frac{\alpha}{\sqrt{\alpha^2+\beta^2}},$$

where $\alpha, \beta \in \mathbb{R}$, $\alpha^2+\beta^2>0$.

- We can choose $\alpha, \beta$ so that any prescribed element in the $i$th or $j$th row of $\Omega^{(i,j)}A$ is zero.
- The rows of $\Omega^{(i,j)}A$ are the same as the rows of $A$, except that the $i$th and $j$th rows of the product are linear combinations of the $i$th and $j$th rows of $A$.
- $\Omega^{(i,j)}$ is an orthogonal matrix, thus $\tilde{A} = \Omega^{(i,j)}A\Omega^{(i,j)\top}$ inherits the eigenvalues of $A$.
- The only difference between $\tilde{A}$ and $\Omega^{(i,j)}A$ is that the $i$th and $j$th columns of $\tilde{A}$ are linear combinations of the $i$th and $j$th columns of $\Omega^{(i,j)}A$.
- If $A$ is symmetric, then so is $\tilde{A}$.

**Method 2.15** (Transformation to an upper Hessenberg form). The following technique replaces $A$ by $SAS^{-1}$, where $S$ is a product of Givens rotations, chosen so that $A$ becomes *upper Hessenberg.*

**0.** Given an $n \times n$ matrix $A$, $n \geq 3$, set $p=1$, $q=3$.

**1.** Choose the Givens rotation $\Omega^{(p+1,q)}$ such that the $(q,p)$th element of $\Omega^{(p+1,q)}A$ is zero. Overwrite $A$ by $\Omega^{(p+1,q)}A\Omega^{(p+1,q)\top}$.

**2.** If $q<n$, then increase $q$ by 1 and go to Step **1.** If $q=n$ and $p<n-2$, increase $p$ by 1, reset $q=p+2$ go to Step **1.** Otherwise terminate.

Since every element that we have set to zero remains zero, the final outcome is indeed an upper Hessenberg matrix. If A is symmetric then so will be the outcome of the calculation, hence it will be tridiagonal. In general, the cost of this procedure is $\mathcal{O}(n^3)$.

Note that we can transform $A$ to upper Hessenberg using *Householder reflections,* rather than Givens rotations. In that case we deal with a column at a time, seeking $\boldsymbol{u} \in \mathbb{R}^n$ s.t. $Q = I - 2\boldsymbol{u}\boldsymbol{u}^\top/\|\boldsymbol{u}\|^2$ and the $k$th column of $QA$ is consistent with the upper Hessenberg form (cf. Example 12).

**Algorithm 2.16** (The QR algorithm). The "plain vanilla" version of the QR algorithm is as follows. Set $A_0 = A$. For $k=0,1,\ldots$ calculate the QR factorization $A_k = Q_kR_k$ (here $Q_k$ is $n\times n$ orthogonal and $R_k$ is $n \times n$ upper triangular) and set $A_{k+1} = R_kQ_k$.

The eigenvalues of $A_{k+1}$ are the same as the eigenvalues of $A_k$, since

$$A_{k+1} = R_kQ_k = Q_k^{-1}(Q_kR_k)Q_k = Q_k^{-1}A_kQ_k, \tag{2.3}$$

a similarity transformation. Moreover, $Q_k^{-1} = Q_k^\top$, therefore if $A_k$ is symmetric then so is $A_{k+1}$.

Suppose that for some $k \geq 0$

$$A_{k+1} = \begin{bmatrix} B & C \\ D & E \end{bmatrix},$$

[1] Please email all corrections and suggestions to these notes to A.Iserles@damtp.cam.ac.uk. All handouts are available on the WWW at the URL http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html.

where $B, E$ are square and $D \approx O$. Because of (2.3), we can *deflate* $A_{k+1}$ and calculate the eigenvalues of $B$ and $E$ separately (again, with QR, except that there is nothing to calculate for $1 \times 1$ and $2 \times 2$ blocks). As it turns out, this state of affairs occurs surprisingly often.

**Technique 2.17** (The QR iteration for upper Hessenberg matrices). In this case the factorization $A_k = Q_k R_k$ is calculated as follows. Let $H = A_k$. For $i = 1, 2, \ldots, n-1$, replace $H$ by $\Omega^{(i,i+1)} H$, where $\Omega^{(i,i+1)}$ is the Givens rotation that renders the $(i+1, i)$ element of the new $H$ matrix zero. Thus the product $\Omega^{(n-1,n)}\Omega^{(n-2,n-1)} \cdots \Omega^{(1,2)} A_k = \hat{H}$, say, is upper triangular. This means that $R_k = \hat{H}$, $Q_k = \Omega^{(1,2)\top}\Omega^{(2,3)\top} \cdots \Omega^{(n-1,n)\top}$ and the QR iteration sets $A_{k+1} = R_k Q_k = \hat{H}\Omega^{(1,2)\top}\Omega^{(2,3)\top} \cdots \Omega^{(n-1,n)\top}$. It follows that $A_{k+1}$ is also upper Hessenberg, because, for $j = 1, 2, \ldots, n-2$, the $j$th column of $A_{k+1}$ is a linear combination of the first $j+1$ columns of $\hat{H}$. Thus a strong advantage of applying Method 2.15 initially is that, in every iteration, $Q_k$ in Algorithm 2.16 is a product of just $n-1$ Givens rotations. Hence each iteration of the QR algorithm requires just $\mathcal{O}(n^2)$ operations.

**Technique 2.18** (The QR iteration for symmetric matrices). Again we begin by using Method 2.15, in order that Algorithm 2.16 commences from a symmetric tridiagonal matrix. Then Technique 2.17 is applied as before, except that we take advantage of some helpful features. Firstly, not just the upper Hessenberg structure, also symmetry is retained, so each $A_{k+1}$ is *both symmetric and tridiagonal.* It follows that, whenever a Givens rotation combines either two adjacent rows or two adjacent columns of a matrix, the total number of nonzero elements in the new combination of rows or columns is at most five. Thus there is a bound on the work of each rotation that is independent of $n$. Hence each QR iteration requires just $\mathcal{O}(n)$ operations!

**Notation 2.19** To analyse the matrices $A_k$ that occur in the QR algorithm 2.16, we introduce

$$\bar{Q}_k = Q_0 Q_1 \cdots Q_k, \qquad \bar{R}_k = R_k R_{k-1} \cdots R_0, \qquad k = 0, 1, \ldots. \tag{2.4}$$

Note that $\bar{Q}_k$ is orthogonal and $\bar{R}_k$ upper triangular.

***Lemma 2.20*** *(Fundamental properties of $\bar{Q}_k$ and $\bar{R}_k$). $A_{k+1}$ is related to the original matrix $A$ by the similarity transformation $A_{k+1} = \bar{Q}_k^\top A \bar{Q}_k$. Further, $\bar{Q}_k \bar{R}_k$ is the QR factorization of $A^{k+1}$.*

**Proof** We prove the first assertion by induction. (2.3) implies that $A_1 = Q_0^\top A_0 Q_0 = \bar{Q}_0^\top A \bar{Q}_0$. Assuming $A_k = \bar{Q}_{k-1}^\top A \bar{Q}_{k-1}$, equations (2.3) and (2.4) imply that

$$A_{k+1} = Q_k^\top A_k Q_k = Q_k^\top (\bar{Q}_{k-1}^\top A \bar{Q}_{k-1}) Q_k = \bar{Q}_k^\top A \bar{Q}_k,$$

as claimed.
The second assertion is true for $k = 0$, since $\bar{Q}_0 \bar{R}_0 = Q_0 R_0 = A_0 = A$. Again, we use induction, assuming $\bar{Q}_{k-1}\bar{R}_{k-1} = A^k$. Thus, using the definition (2.4) and the first statement of the lemma, we deduce that

$$\begin{aligned}\bar{Q}_k \bar{R}_k &= (\bar{Q}_{k-1} Q_k)(R_k \bar{R}_{k-1}) = \bar{Q}_{k-1} A_k \bar{R}_{k-1} = \bar{Q}_{k-1}(\bar{Q}_{k-1}^\top A \bar{Q}_{k-1})\bar{R}_{k-1} \\ &= A\bar{Q}_{k-1}\bar{R}_{k-1} = A \cdot A^k = A^{k+1}\end{aligned}$$

and the lemma is true. □

**Property 2.21** (Relation between QR and the power method). We consider the first column of the matrix equation $\bar{Q}_{k-1}\bar{R}_{k-1} = A^k$, when we assume that the eigenvalues of $A$ have different magnitudes, say $0 \le |\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$. Let $\boldsymbol{e}_1 = \sum_{j=1}^n \psi_j \boldsymbol{w}_j$ be the expansion of the first coordinate vector in terms of the normalized eigenvectors of $A$, and let $l \in \{1, 2, \ldots, n\}$ be the greatest integer such that $\psi_l \neq 0$. Then $A^k \boldsymbol{e}_1$ is a multiple of $\sum_{j=1}^l \psi_j (\lambda_j/\lambda_l)^k \boldsymbol{w}_j$, so the first column of $A^k$ tends to be a multiple of $\boldsymbol{w}_l$ for $k \gg 1$. Let $\bar{\boldsymbol{q}}_k$ be the first column of $\bar{Q}_{k-1}$. Because $\bar{R}_{k-1}$ is upper triangular, the first column of $A^k = \bar{Q}_{k-1}\bar{R}_{k-1}$ is a multiple of $\bar{\boldsymbol{q}}_k$. Therefore $\bar{\boldsymbol{q}}_k$ tends to be a multiple of $\boldsymbol{w}_l$. Further, because both $\bar{\boldsymbol{q}}_k$ and and $\boldsymbol{w}_l$ have unit length, we deduce that $\bar{\boldsymbol{q}}_k = \pm \boldsymbol{w}_l + \boldsymbol{h}_k$, where $\boldsymbol{h}_k$ tends to zero as $k \to \infty$, and where the $\pm$ alternative may depend on $k$.