Dr C.-B. Schönlieb

Mathematical Tripos Part II: Michaelmas Term 2015

Numerical Analysis – Lecture 23

Theorem 5.8 *Let A and S be $n \times n$ matrices, S being nonsingular. Then $\boldsymbol{w}$ is an eigenvector of A with eigenvalue λ if and only if $\widehat{\boldsymbol{w}} = S\boldsymbol{w}$ is an eigenvector of $\widehat{A} = SAS^{-1}$ with the same eigenvalue.*

Proof. $A\boldsymbol{w} = \lambda\boldsymbol{w} \quad \Leftrightarrow \quad AS^{-1}(S\boldsymbol{w}) = \lambda\boldsymbol{w} \quad \Leftrightarrow \quad (SAS^{-1})(S\boldsymbol{w}) = \lambda(S\boldsymbol{w}).$ □

Definition 5.9 (Deflation) Suppose that we have found one solution of the eigenvector equation $A\boldsymbol{w} = \lambda\boldsymbol{w}$, where A is again $n \times n$. Then *deflation* is the task of constructing an $(n-1) \times (n-1)$ matrix, B say, whose eigenvalues are the other eigenvalues of A. Specifically, we apply a similarity transformation S to A such that the first column of $\widehat{A} = SAS^{-1}$ is λ times the first coordinate vector $\boldsymbol{e}_1$, because it follows from the characteristic equation for eigenvalues and from Theorem 5.8 that we can let B be the bottom right $(n-1) \times (n-1)$ submatrix of $\widehat{A} = SAS^{-1}$.

We write the condition on S as $(SAS^{-1})\boldsymbol{e}_1 = \lambda\boldsymbol{e}_1$. Then the last equation in the proof of Theorem 5.8 shows that it is sufficient if S has the property $S\boldsymbol{w} = c\boldsymbol{e}_1$, where c is any nonzero scalar.

Technique 5.10 (Algorithm for deflation for symmetric A) Suppose that A is symmetric and $\boldsymbol{w} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ are given so that $A\boldsymbol{w} = \lambda\boldsymbol{w}$. We seek a nonsingular matrix S such that $S\boldsymbol{w} = c\boldsymbol{e}_1$ and such that SAS^{-1} is also symmetric. The last condition holds if S is orthogonal, since then $S^{-1} = S^T$. It is suitable to pick a *Householder reflection*, which means that S has the form

$$H_u = I - 2\boldsymbol{u}\boldsymbol{u}^T/\|\boldsymbol{u}\|^2, \quad \text{where} \quad \boldsymbol{u} \in \mathbb{R}^n.$$

Specifically, we recall from the Numerical Analysis IB course that Householder reflections are orthogonal and that, because $H_u\boldsymbol{u} = -\boldsymbol{u}$ and $H_u\boldsymbol{v} = \boldsymbol{v}$ if $\boldsymbol{u}^T\boldsymbol{v} = 0$, they reflect any vector in $\mathbb{R}^n$ with respect to the $(n-1)$-dimensional hyperplane orthogonal to $\boldsymbol{u}$. So, for any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of equal lengths,

$$H_{\boldsymbol{u}}\boldsymbol{x} = \boldsymbol{y}, \quad \text{where} \quad \boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y}.$$

Hence,

$$\left(I - 2\frac{\boldsymbol{u}\boldsymbol{u}^T}{\|\boldsymbol{u}\|^2}\right)\boldsymbol{w} = \pm\|\boldsymbol{w}\|\boldsymbol{e}_1. \quad \text{where} \quad \boldsymbol{u} = \boldsymbol{w} \mp \|\boldsymbol{w}\|\boldsymbol{e}_1\,.$$

Since the bottom $n-1$ components of $\boldsymbol{u}$ and $\boldsymbol{w}$ coincide, the calculation of $\boldsymbol{u}$ requires only $\mathcal{O}(n)$ computer operations. Further, the calculation of SAS^{-1} can be done in only $\mathcal{O}(n^2)$ operations, taking advantage of the form $S = I - 2\boldsymbol{u}\boldsymbol{u}^T/\|\boldsymbol{u}\|^2$, even if all the elements of A are nonzero.

After deflation, we may find an eigenvector, $\widehat{\boldsymbol{w}}$ say, of SAS^{-1}. Then the new eigenvector of A, according to Theorem 5.8, is $S^{-1}\widehat{\boldsymbol{w}} = S\widehat{\boldsymbol{w}}$, because Householder matrices, like all symmetric orthogonal matrices, are *involutions*: $S^2 = I$.

Revision 5.11 (Givens rotations) The notation $\Omega^{[i,j]}$ denotes the following $n \times n$ matrix

$$\Omega^{[i,j]} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & c & & s \\ & & & \ddots & \\ & & -s & & c \\ & & & & & \ddots \\ & & \uparrow_{i} & & \uparrow_{j} & & 1 \end{bmatrix}, \quad c^2 + s^2 = 1.$$

Generally, for any vector $\boldsymbol{a}_k \in \mathbb{R}^n$, we can find a matrix $\Omega^{[i,j]}$ such that

$$\Omega^{[i,j]}\boldsymbol{a} = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & c & & s \\ & & & \ddots & \\ & & -s & & c \\ & & & & & \ddots \\ & & \uparrow_{i} & & \uparrow_{j} & & 1 \end{bmatrix} \begin{bmatrix} a_{1k} \\ \vdots \\ a_{ik} \\ \vdots \\ a_{jk} \\ \vdots \\ a_{nk} \end{bmatrix} = \begin{bmatrix} a_{1k} \\ \vdots \\ r \\ \vdots \\ 0 \\ \vdots \\ a_{nk} \end{bmatrix} \begin{matrix} \\ \\ \leftarrow i \\ \\ \leftarrow j \\ \\ \\ \end{matrix} \qquad \begin{aligned} c &= \frac{a_{ik}}{\sqrt{a_{ik}^2 + a_{jk}^2}}\,, \\ s &= \frac{a_{jk}}{\sqrt{a_{ik}^2 + a_{jk}^2}}\,, \\ r &= \sqrt{a_{ik}^2 + a_{jk}^2}\,. \end{aligned}$$

1) We can choose $\Omega^{[i,j]}$ so that any prescribed element $\widetilde{a}_{jk}$ in the j-th row of $\widetilde{A} = \Omega^{[i,j]} \times A$ is zero.

2) The rows of $\widetilde{A} = \Omega^{[i,j]} \times A$ 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.

3) The columns of $\widehat{A} = \widetilde{A} \times \Omega^{[i,j]T}$ are the same as the columns of $\widetilde{A}$, except that the i-th and j-th columns of $\widehat{A}$ are linear combinations of the i-th and j-th columns of $\widetilde{A}$.

4) $\Omega^{[i,j]}$ is an orthogonal matrix, thus $\widehat{A} = \Omega^{[i,j]} A \Omega^{[i,j]T}$ inherits the eigenvalues of A.

5) If A is symmetric, then so is $\widehat{A}$.

Method 5.12 (Transformation to an upper Hessenberg form) We replace A by $\widehat{A} = SAS^{-1}$, where S is a product of Givens rotations $\Omega^{[i,j]}$ chosen to annihilate subsubdiagonal elements $a_{j,i-1}$ in the (i–1)-st column:

$$\begin{bmatrix} \ast\ast\ast\ast \\ \ast\ast\ast\ast \\ \ast\ast\ast\ast \\ \ast\ast\ast\ast \end{bmatrix} \overset{\Omega^{[2,3]}\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast \\ \bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet \\ \ast\ast\ast\ast \end{bmatrix} \overset{\times\Omega^{[2,3]T}}{\rightarrow} \begin{bmatrix} \ast\bullet\bullet\ast \\ \ast\bullet\bullet\ast \\ 0\bullet\bullet\ast \\ \ast\bullet\bullet\ast \end{bmatrix} \overset{\Omega^{[2,4]}\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast \\ \bullet\bullet\bullet\bullet \\ 0\ast\ast\ast \\ 0\bullet\bullet\bullet \end{bmatrix} \overset{\times\Omega^{[2,4]T}}{\rightarrow} \begin{bmatrix} \ast\bullet\ast\bullet \\ \ast\bullet\ast\bullet \\ 0\bullet\ast\bullet \\ 0\bullet\ast\bullet \end{bmatrix} \overset{\Omega^{[3,4]}\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast \\ \ast\ast\ast\ast \\ 0\bullet\bullet\bullet \\ 00\bullet\bullet \end{bmatrix} \overset{\times\Omega^{[3,4]T}}{\rightarrow} \begin{bmatrix} \ast\ast\bullet\bullet \\ \ast\ast\bullet\bullet \\ 0\ast\bullet\bullet \\ 00\bullet\bullet \end{bmatrix}$$

The $\bullet$-elements have changed through a single transformation while the $\ast$-elements remained the same. It is seen that every element that we have set to zero remains zero, and 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 $\mathcal{O}(n^3)$.

Alternatively, 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, taking $\boldsymbol{u}$ such that, with $H_u = I - 2\boldsymbol{u}\boldsymbol{u}^T/\|\boldsymbol{u}\|^2$, the i-th column of $\widetilde{B} = H_u B$ is consistent with the upper Hessenberg form. Such a $\boldsymbol{u}$ has its first i coordinates vanishing, therefore $\widehat{B} = \widetilde{B} H_u^T$ has the first i columns unchanged, and all new and old zeros (which are in the first i columns) stay untouched.

$$\begin{bmatrix} \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \end{bmatrix} \overset{H_1\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast\ast \\ \bullet\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \end{bmatrix} \overset{\times H_1^T}{\rightarrow} \begin{bmatrix} \ast\bullet\bullet\bullet\bullet \\ \ast\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \\ 0\bullet\bullet\bullet\bullet \end{bmatrix} \overset{H_2\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \\ 0\bullet\bullet\bullet\bullet \\ 00\bullet\bullet\bullet \\ 00\bullet\bullet\bullet \end{bmatrix} \overset{\times H_2^T}{\rightarrow} \begin{bmatrix} \ast\ast\bullet\bullet\bullet \\ \ast\ast\bullet\bullet\bullet \\ 0\ast\bullet\bullet\bullet \\ 00\bullet\bullet\bullet \\ 00\bullet\bullet\bullet \end{bmatrix} \overset{H_3\times}{\rightarrow} \begin{bmatrix} \ast\ast\ast\ast\ast \\ \ast\ast\ast\ast\ast \\ 0\ast\ast\ast\ast \\ 00\bullet\bullet\bullet \\ 000\bullet\bullet \end{bmatrix} \overset{\times H_3^T}{\rightarrow} \begin{bmatrix} \ast\ast\ast\bullet\bullet \\ \ast\ast\ast\bullet\bullet \\ 0\ast\ast\bullet\bullet \\ 00\ast\bullet\bullet \\ 000\bullet\bullet \end{bmatrix}$$

Algorithm 5.13 (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 $A_k = Q_k R_k$ (here Q_k is $n \times n$ orthogonal and R_k is $n \times n$ upper triangular) and set $A_{k+1} = R_k Q_k$.

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

$$A_{k+1} = R_k Q_k = Q_k^{-1}(Q_k R_k) Q_k = Q_k^{-1} A_k Q_k, \tag{5.2}$$

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

If for some $k \geq 0$ the matrix A_{k+1} can be regarded as "deflated", i.e. it has the block form

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

where B, E are square and $D \approx O$, then we calculate the eigenvalues of B and E separately (again, with QR, except that there is nothing to calculate for 1×1 and 2×2 blocks). As it turns out, such a "deflation" occurs surprisingly often.

Technique 5.14 (The QR iteration for upper Hessenberg matrices) If A_k is upper Hessenberg, then its QR factorization by means of the Givens rotations produces the matrix

$$R_k = Q_k^T A_k = \Omega^{[n-1,n]} \cdots \Omega^{[2,3]} \Omega^{[1,2]} A_k\,,$$

which is upper triangular. The QR iteration sets $A_{k+1} = R_k Q_k = R_k \Omega^{[1,2]T} \Omega^{[2,3]T} \cdots \Omega^{[n-1,n]T}$, and it follows that A_{k+1} is also upper Hessenberg, because

$$\begin{bmatrix} \ast\ast\ast\ast \\ 0\ast\ast\ast \\ 00\ast\ast \\ 000\ast \end{bmatrix} \overset{\times\Omega^{[1,2]T}}{\rightarrow} \begin{bmatrix} \bullet\bullet\ast\ast \\ \bullet\bullet\ast\ast \\ 00\ast\ast \\ 000\ast \end{bmatrix} \overset{\times\Omega^{[2,3]T}}{\rightarrow} \begin{bmatrix} \ast\bullet\bullet\ast \\ \ast\bullet\bullet\ast \\ 0\bullet\bullet\ast \\ 000\ast \end{bmatrix} \overset{\times\Omega^{[3,4]T}}{\rightarrow} \begin{bmatrix} \ast\ast\bullet\bullet \\ \ast\ast\bullet\bullet \\ 0\ast\bullet\bullet \\ 00\bullet\bullet \end{bmatrix}$$

Thus a strong advantage of bringing A to the upper Hessenberg form initially is that then, in every iteration in QR algorithm, Q_k is a product of just $n-1$ Givens rotations. Hence each iteration of the QR algorithm requires just $\mathcal{O}(n^2)$ operations.