Dr A. Iserles
Easter 1997

Numerical Analysis – Lecture 4

. . . proof. We consider next the case of $\rho(H) < 1$, assuming in addition that H possesses n linearly independent eigenvectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n$, say. Hence $H\boldsymbol{w}_j = \lambda_j \boldsymbol{w}_j$, $|\lambda_j| < 1$, $j = 1, 2, \ldots, n$. Linear independence means that every $\boldsymbol{v} \in \mathbb{R}^n$ can be expressed as a linear combination of the eigenvectors. Hence, given $\boldsymbol{x}_0 \in \mathbb{R}^n$, there exist $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{C}$ such that $\boldsymbol{v}_0 = \boldsymbol{x}_0 - \boldsymbol{x}^* = \sum_{j=1}^n \alpha_j \boldsymbol{w}_j$. Thus,

$$\boldsymbol{v}_1 = H\boldsymbol{v}_0 = \sum_{j=1}^n \alpha_j \lambda_j \boldsymbol{w}_j \qquad \text{and, by induction,} \qquad \boldsymbol{v}_m = \sum_{j=1}^n \alpha_j \lambda_j^m \boldsymbol{w}_j$$

for all $m = 0, 1, \ldots$. Since $\rho(H) < 1$, it follows that $\lim_{m\to\infty} \boldsymbol{w}_m = \boldsymbol{0}$, as required. □

The 'missing' case Suppose that $\rho(H) < 1$ but that H does not have n linearly independent eigenvalues. This occurs, for example, for the matrix

$$H = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix},$$

where $b \neq 0$ and $|a| < 1$. The eigenvalues of H are both a, but it is an easy exercise to verify that all eigenvectors are necessarily multiples of $\boldsymbol{e}_1$.

4 QR factorization of matrices

4.1 Scalar products, norms and orthogonality

We revise few definitions. $\mathbb{R}^n$ is the linear space of all real n-tuples. For all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ we define the *scalar product*

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{j=1}^n u_j v_j = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v}.$$

The *norm* (a.k.a. the *Euclidean length*) of $\boldsymbol{u} \in \mathbb{R}^n$ is $\|\boldsymbol{u}\| := \left(\sum_{j=1}^n u_j^2\right)^{1/2} = \langle \boldsymbol{u}, \boldsymbol{u} \rangle^{1/2}$. Note that $\|\boldsymbol{u}\| = 0$ iff $\boldsymbol{u} = \boldsymbol{0}$.

We say that $\boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^n$ are *orthogonal* to each other if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.

The vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m \in \mathbb{R}^n$ are *orthonormal* if

$$\langle \boldsymbol{q}_k, \boldsymbol{q}_\ell \rangle = \begin{cases} 1, & k = \ell, \\ 0, & k \neq \ell, \end{cases} \qquad k, \ell = 1, 2, \ldots, m.$$

An $n \times n$ real matrix Q is *orthogonal* if all its columns are orthonormal. This is equivalent to $Q^{\mathrm{T}}Q = I$ (I is the *unit matrix*), because $(Q^{\mathrm{T}}Q)_{k,\ell} = \langle \boldsymbol{q}_k, \boldsymbol{q}_\ell \rangle$, hence to $Q^{-1} = Q^{\mathrm{T}}$. We conclude that $QQ^{\mathrm{T}} = QQ^{-1} = I$ and also the rows of an orthogonal matrix are orthonormal. As a consequence of $1 = \det I = \det(QQ^{\mathrm{T}}) = \det Q \det Q^{\mathrm{T}} = (\det Q)^2$, we deduce that $\det Q = \pm 1$ and an orthogonal matrix is nonsingular.

Proposition If P, Q are orthogonal then so is PQ.

Proof. Since $P^{\mathrm{T}}P = Q^{\mathrm{T}}Q = I$, we have $(PQ)^{\mathrm{T}}(PQ) = (Q^{\mathrm{T}}P^{\mathrm{T}})(PQ) = Q^{\mathrm{T}}(P^{\mathrm{T}}P)Q = Q^{\mathrm{T}}Q = I$, hence PQ is orthogonal. □

Proposition Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m \in \mathbb{R}^n$ be orthonormal. Then $m \le n$.

Proof. Suppose that $m \ge n+1$ and let Q be the orthogonal matrix whose columns are $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$. Since Q is nonsingular and $\boldsymbol{q}_m \ne \boldsymbol{0}$, there exists a nonzero solution to the linear system $Q\boldsymbol{a} = \boldsymbol{q}_m$, hence $\boldsymbol{q}_m = \sum_{j=1}^n a_j \boldsymbol{q}_j$. But

$$0 = \langle \boldsymbol{q}_\ell, \boldsymbol{q}_m \rangle = \left\langle \boldsymbol{q}_\ell, \sum_{j=1}^n a_j \boldsymbol{q}_j \right\rangle = \sum_{j=1}^n a_j \langle \boldsymbol{q}_\ell, \boldsymbol{q}_j \rangle = a_j, \qquad \ell = 1, 2, \ldots, n,$$

hence $\boldsymbol{a} = \boldsymbol{0}$, a contradiction. We deduce that $m \le n$. □

Lemma Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m \in \mathbb{R}^n$ be orthonormal and $m \le n-1$. Then there exists $\boldsymbol{q}_{m+1} \in \mathbb{R}^n$ such that $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{m+1}$ are orthonormal.

Proof. Let Q be the $n \times m$ matrix whose columns are $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$. Since $\sum_{k=1}^n \sum_{j=1}^m Q_{k,j}^2 = \sum_{j=1}^m \|\boldsymbol{q}_j\|^2 = m < n$, it follows that $\exists \ell \in \{1, 2, \ldots, n\}$ such that $\sum_{j=1}^m Q_{\ell,j}^2 < 1$. We let $\boldsymbol{w} := \boldsymbol{e}_\ell - \sum_{j=1}^m \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle \boldsymbol{q}_j$. Since $Q_{\ell,j} = \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle$, we have $\|\boldsymbol{w}\|^2 = \langle \boldsymbol{w}, \boldsymbol{w} \rangle = 1 - \sum_{j=1}^m Q_{\ell,j}^2 > 0$ and, by construction, $\boldsymbol{w}$ is orthogonal to $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$. We set $\boldsymbol{q}_{m+1} = \boldsymbol{w}/\|\boldsymbol{w}\|$. □

4.2 The QR factorization

The QR factorization of an $m \times n$ matrix A has the form $A = QR$, where Q is an $m \times m$ *orthogonal* matrix and R is an $m \times n$ *upper triangular* matrix (i.e., $R_{i,j} = 0$ for $i > j$). We will demonstrate in the sequel that every matrix has a (non-unique) QR factorization.

An application Let $m = n$ and A be nonsingular. We can solve $A\boldsymbol{x} = \boldsymbol{b}$ by calculating the QR factorization of A and solving first $Q\boldsymbol{y} = \boldsymbol{b}$ (hence $\boldsymbol{y} = Q^{\mathrm{T}}\boldsymbol{b}$) and then $R\boldsymbol{x} = \boldsymbol{y}$ (a triangular system!).

Interpretation of the QR factorization Let $m \ge n$ and denote the columns of A and Q by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m$ respectively. Since

$$\left[\ \boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_n\ \right] = \left[\ \boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \cdots \quad \boldsymbol{q}_m\ \right] \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,n} \\ 0 & R_{2,2} & & \vdots \\ \vdots & \ddots & \ddots & \\ & & 0 & R_{n,n} \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix},$$

we have $\boldsymbol{a}_k = \sum_{j=1}^k R_{j,k} \boldsymbol{q}_j$, $k = 1, 2, \ldots, n$. In other words, Q has the property that each kth column of A can be expressed as a linear combination of the first k columns of Q.