Mathematical Tripos Part IB: Easter 2001

Numerical Analysis – Lecture 4[1]

3.3 Spectral radius and Gershgorin theorem

How can one compute the spectral radius of a matrix? The simplest tool for estimating the position of the eigenvalues is the following theorem.

Theorem [Gerschgorin]. Let A be a $n \times n$ matrix (with real or complex coefficients). Then all of its eigenvalues are contained in the union $\cup_{i=1}^n \Gamma_i$ where Γ_i's are the circles in the complex plane

$$\Gamma_i = \{z \in \mathbb{C} : |z - a_{ii}| \le r_i\}, \quad r_i := \textstyle\sum_{j=1, j\ne i}^n |a_{ij}|, \qquad i = 1, \ldots, n. \tag{3.6}$$

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A, and let $\boldsymbol{w}$ be the corresponding eigenvector. Let $\boldsymbol{w}_i$ be the largest component in absolute value. Consider the i-th equation of the relation $A\boldsymbol{w} = \lambda\boldsymbol{w}$:

$$a_{ii}\boldsymbol{w}_i + \textstyle\sum_{j=1, j\ne i}^n a_{ij}\boldsymbol{w}_j = \lambda\boldsymbol{w}_i.$$

This implies the inequality

$$|\lambda - a_{ii}| \cdot |\boldsymbol{w}_i| \le \textstyle\sum_{j=1, j\ne i}^n |a_{ij}||\boldsymbol{w}_j| \le |\boldsymbol{w}_i| \sum_{j=1, j\ne i}^n |a_{ij}| =: |\boldsymbol{w}_i| r_i$$

which, after dividing by $|\boldsymbol{w}_i|$, gives (3.6). □

A matrix A is said to be *strictly diagonal dominant* if $|a_{ii}| > r_i$ for all i. From Gerschgorin theorem, it follows that a strictly diagonal dominant matrix is nonsingular.

Theorem. Let A be strictly diagonal dominant. Then the Jacobi iteration converges.

Proof. We prove that the spectral radius of the Jacobi iteration matrix $H = -A_{\mathrm{D}}^{-1}(A_{\mathrm{L}} + A_{\mathrm{U}})$ satisfies $\rho(H) < 1$. One computes $h_{ii} = 0$, $h_{ij} = -a_{ij}/a_{ii}$, and if A is strictly diagonal dominant, then $r_i = \sum_{i \ne j} |h_{ij}| < 1$. By Gerschgorin theorem, the eigenvalues of H lie in the union of the discs $\Gamma_i := \{z \in \mathbb{C} : |z| < 1\}$, i.e. $\rho(H) < 1$ □

4 QR factorization of matrices

4.1 Scalar products, norms and orthogonality

We first revise a few definitions. In what follows, $\mathbb{R}^n$ is the linear space of all real n-tuples.

- For all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ we define the *scalar product*

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

- If $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$ then $\langle \alpha\boldsymbol{u} + \beta\boldsymbol{w}, \boldsymbol{v}\rangle = \alpha\langle \boldsymbol{u}, \boldsymbol{v}\rangle + \beta\langle \boldsymbol{w}, \boldsymbol{v}\rangle$.
- The *norm* (a.k.a. the *Euclidean length*) of $\boldsymbol{u} \in \mathbb{R}^n$ is $\|\boldsymbol{u}\| = \left(\sum_{j=1}^n u_j^2\right)^{1/2} = \langle \boldsymbol{u}, \boldsymbol{u}\rangle^{1/2} \ge 0$.
- For $\boldsymbol{u} \in \mathbb{R}^n$, $\|\boldsymbol{u}\| = 0$ 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

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

- An $n \times n$ real matrix Q is *orthogonal* if all its columns are orthonormal. Since $(Q^{\mathrm{T}}Q)_{k,\ell} = \langle \boldsymbol{q}_k, \boldsymbol{q}_\ell\rangle$, this implies that $Q^{\mathrm{T}}Q = I$ (I is the *unit matrix*). Hence $Q^{-1} = Q^{\mathrm{T}}$ and $QQ^{\mathrm{T}} = QQ^{-1} = I$. We conclude that the rows of an orthogonal matrix are also orthonormal, and that Q^{T} is an orthogonal matrix. Further, $1 = \det I = \det(QQ^{\mathrm{T}}) = \det Q \det Q^{\mathrm{T}} = (\det Q)^2$, and thus we deduce that $\det Q = \pm 1$, and that an orthogonal matrix is nonsingular.

[1]Corrections and suggestions to these notes should be emailed to `a.shadrin@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartIB/Handouts.html`.

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. We argue by contradiction. 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

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

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. We construct $\boldsymbol{q}_{m+1}$. 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 $\boldsymbol{w} = \boldsymbol{e}_\ell - \sum_{j=1}^{m} \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle \boldsymbol{q}_j$. Then for $i = 1, 2, \ldots, m$

$$\langle \boldsymbol{q}_i, \boldsymbol{w} \rangle = \langle \boldsymbol{q}_i, \boldsymbol{e}_\ell \rangle - \sum_{j=1}^{m} \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle \langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = 0,$$

i.e. by design $\boldsymbol{w}$ is orthogonal to $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m$. Further, since $Q_{\ell,j} = \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle$, we have

$$\|\boldsymbol{w}\|^2 = \langle \boldsymbol{w}, \boldsymbol{w} \rangle = \langle \boldsymbol{e}_\ell, \boldsymbol{e}_\ell \rangle - 2\sum_{j=1}^{m} \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle \langle \boldsymbol{e}_\ell, \boldsymbol{q}_j \rangle + \sum_{j=1}^{m} \langle \boldsymbol{q}_j, \boldsymbol{e}_\ell \rangle \sum_{k=1}^{m} \langle \boldsymbol{q}_k, \boldsymbol{e}_\ell \rangle \langle \boldsymbol{q}_j, \boldsymbol{q}_k \rangle = 1 - \sum_{j=1}^{m} Q_{\ell,j}^2 > 0.$$

Thus we define $\boldsymbol{q}_{m+1} = \boldsymbol{w}/\|\boldsymbol{w}\|$. □

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

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

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.