Numerical Analysis Hilary Term 2018
Lecture 7: Matrix Eigenvalues

Background: first, an important result from analysis (not proved or examinable!), which will be useful.

Theorem. (Ostrowski) The eigenvalues of a matrix are continuously dependent on the entries. That is, suppose that $\{\lambda_i, i = 1, \ldots, n\}$ and $\{\mu_i, i = 1, \ldots, n\}$ are the eigenvalues of $A \in \mathbb{R}^{n\times n}$ and $A + B \in \mathbb{R}^{n\times n}$ respectively. Given any $\varepsilon > 0$, there is a $\delta > 0$ such that $|\lambda_i - \mu_i| < \varepsilon$ whenever $\max_{i,j} |b_{ij}| < \delta$, where $B = \{b_{ij}\}_{1\le i,j\le n}$.

Aim: estimate the eigenvalues of a matrix.

Theorem. Gerschgorin's theorem: Suppose that $A = \{a_{ij}\}_{1\le i,j\le n} \in \mathbb{R}^{n\times n}$, and λ is an eigenvalue of A. Then, λ lies in the union of the **Gerschgorin discs**

$$D_i = \left\{ z \in \mathbb{C} \,\middle|\, |a_{ii} - z| \le \sum_{\substack{j\ne i\\ j=1}}^{n} |a_{ij}| \right\}, \quad i = 1, \ldots, n.$$

Proof. If λ is an eigenvalue of $A \in \mathbb{R}^{n\times n}$, then there exists an eigenvector $x \in \mathbb{R}^n$ with $Ax = \lambda x$, $x \ne 0$, i.e.,

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i, \quad i = 1, \ldots, n.$$

Suppose that $|x_k| \ge |x_\ell|$, $\ell = 1, \ldots, n$, i.e.,

$$\text{“}x_k \text{ is the largest entry”}. \tag{1}$$

Then certainly $\displaystyle\sum_{j=1}^{n} a_{kj} x_j = \lambda x_k$, or

$$(a_{kk} - \lambda) x_k = -\sum_{\substack{j\ne k\\ j=1}}^{n} a_{kj} x_j.$$

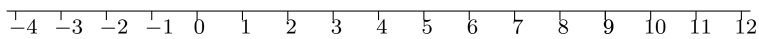
Dividing by x_k, (which, we know, is $\ne 0$) and taking absolute values,

$$|a_{kk} - \lambda| = \left| \sum_{\substack{j\ne k\\ j=1}}^{n} a_{kj} \frac{x_j}{x_k} \right| \le \sum_{\substack{j\ne k\\ j=1}}^{n} |a_{kj}| \left| \frac{x_j}{x_k} \right| \le \sum_{\substack{j\ne k\\ j=1}}^{n} |a_{kj}|$$

by (1). $\square$

Example.

$$A = \begin{bmatrix} 9 & 1 & 2 \\ -3 & 1 & 1 \\ 1 & 2 & -1 \end{bmatrix}$$



With Matlab calculate `>> eig(A) = 8.6573, -2.0639, 2.4066`

Theorem. Gerschgorin's 2nd theorem: If any union of ℓ (say) discs is disjoint from the other discs, then it contains ℓ eigenvalues.

Proof. Consider $B(\theta) = \theta A + (1-\theta)D$, where $D = \mathrm{diag}(A)$, the diagonal matrix whose diagonal entries are those from A. As θ varies from 0 to 1, $B(\theta)$ has entries that vary continuously from $B(0) = D$ to $B(1) = A$. Hence the eigenvalues $\lambda(\theta)$ vary continuously by Ostrowski's theorem. The Gerschgorin discs of $B(0) = D$ are points (the diagonal entries), which are clearly the eigenvalues of D. As θ increases the Gerschgorin discs of $B(\theta)$ increase in radius about these same points as centres. Thus if $A = B(1)$ has a disjoint set of ℓ Gerschgorin discs by continuity of the eigenvalues it must contain exactly ℓ eigenvalues (as they can't jump!). $\square$

Iterative Methods: methods such as LU or QR factorizations are *direct*: they compute a certain number of operations and then finish with "the answer". Another class of methods are *iterative*:

- construct a sequence;
- truncate that sequence "after convergence";
- typically concerned with fast convergence rate (rather than operation count).

Notation: for $x \in \mathbb{R}^n$, $\|x\| = \sqrt{x^{\mathrm{T}}x}$ is the (Euclidean) length of x.

Notation: in iterative methods, x_k usually means the vector x at the kth iteration (rather than kth entry of vector x). Some sources use x^k or $x^{(k)}$ instead.

Power Iteration: a simple method for calculating a single (largest) eigenvalue of a square matrix A (and its associated eigenvector). For arbitrary $y \in \mathbb{R}^n$, set $x_0 = y/\|y\|$ to calculate an initial vector, and then for $k = 0, 1, \ldots$

Compute $y_k = Ax_k$

and set $x_{k+1} = y_k/\|y_k\|$.

This is the **Power Method** or **Iteration**, and computes unit vectors in the direction of $x_0, Ax_0, A^2x_0, A^3x_0, \ldots, A^kx_0$.

Suppose that A is diagonalizable so that there is a basis of eigenvectors of A:

$$\{v_1, v_2, \ldots, v_n\}$$

with $Av_i = \lambda_i v_i$ and $\|v_i\| = 1$, $i = 1, 2, \ldots, n$, and assume that

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Then we can write

$$x_0 = \sum_{i=1}^{n} \alpha_i v_i$$

for some $\alpha_i \in \mathbb{R}$, $i = 1, 2, \ldots, n$, so

$$A^k x_0 = A^k \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} \alpha_i A^k v_i.$$

However, since $Av_i = \lambda_i v_i \implies A^2 v_i = A(Av_i) = \lambda_i A v_i = \lambda_i^2 v_i$, inductively $A^k v_i = \lambda_i^k v_i$. So

$$A^k x_0 = \sum_{i=1}^{n} \alpha_i \lambda_i^k v_i = \lambda_1^k \left[\alpha_1 v_1 + \sum_{i=2}^{n} \alpha_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i \right].$$

Since $(\lambda_i/\lambda_1)^k \to 0$ as $k \to \infty$, $A^k x_0$ tends to look like $\lambda_1^k \alpha_1 v_1$ as k gets large. The result is that by normalizing to be a unit vector

$$\frac{A^k x_0}{\|A^k x_0\|} \to \pm v_1 \quad \text{and} \quad \frac{\|A^k x_0\|}{\|A^{k-1} x_0\|} \approx \left| \frac{\lambda_1^k \alpha_1}{\lambda_1^{k-1} \alpha_1} \right| = |\lambda_1|$$

as $k \to \infty$, and the sign of λ_1 is identified by looking at, e.g., $(A^k x_0)_1/(A^{k-1} x_0)_1$.

Essentially the same argument works when we normalize at each step: the Power Iteration may be seen to compute $y_k = \beta_k A^k x_0$ for some β_k. Then, from the above,

$$x_{k+1} = \frac{y_k}{\|y_k\|} = \frac{\beta_k}{|\beta_k|} \cdot \frac{A^k x_0}{\|A^k x_0\|} \to \pm v_1.$$

Similarly, $y_{k-1} = \beta_{k-1} A^{k-1} x_0$ for some β_{k-1}. Thus

$$x_k = \frac{\beta_{k-1}}{|\beta_{k-1}|} \cdot \frac{A^{k-1} x_0}{\|A^{k-1} x_0\|} \quad \text{and hence} \quad y_k = A x_k = \frac{\beta_{k-1}}{|\beta_{k-1}|} \cdot \frac{A^k x_0}{\|A^{k-1} x_0\|}.$$

Therefore, as above,

$$\|y_k\| = \frac{\|A^k x_0\|}{\|A^{k-1} x_0\|} \approx |\lambda_1|,$$

and the sign of λ_1 may be identified by looking at, e.g., $(x_{k+1})_1/(x_k)_1$.

Hence the largest eigenvalue (and its eigenvector) can be found.

Note: it is possible for a chosen vector x_0 that $\alpha_1 = 0$, but rounding errors in the computation generally introduce a small component in v_1, so that in practice this is not a concern!

This simplified method for eigenvalue computation is the basis for effective methods, but the current state of the art is the **QR Algorithm** which was invented by John Francis in London in 1959/60. We consider the **QR Algorithm** only in the case when A is symmetric.