Numerical Analysis

Raphael Hauser
with thanks to Endre Süli

Oxford Mathematical Institute

HT 2019

Givens Rotations

Another useful family of orthogonal matrices are the **Givens' rotation** matrices:

$$J(i,j,\theta) = \begin{pmatrix} 1 & & & & & & \\ & \cdot & & & & & \\ & & c & & s & & \\ & & & \cdot & & & \\ & & -s & & c & & \\ & & & & & \cdot & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \leftarrow \; i\text{th row} \\ \\ \leftarrow \; j\text{th row} \\ \\ \\ \end{matrix}$$

$$\begin{matrix} \uparrow & \uparrow \\ i & j \end{matrix}$$

where $c = \cos\theta$ and $s = \sin\theta$.

Exercise: Prove that $J(i, j, \theta)J(i, j, \theta)^{\mathrm{T}} = I$— obvious though, since the columns form an orthonormal basis.

Exercise: Prove that $J(i,j,\theta)J(i,j,\theta)^{\mathrm{T}} = I$— obvious though, since the columns form an orthonormal basis.
Note that if $x = (x_1,\ x_2,\ \ldots, x_n)^{\mathrm{T}}$ and $y = J(i,j,\theta)x$, then

$$\begin{aligned} y_k &= x_k \text{ for } k \neq i,j \\ y_i &= cx_i + sx_j \\ y_j &= -sx_i + cx_j \end{aligned}$$

and so we can ensure that $y_j = 0$ by choosing $x_i \sin\theta = x_j \cos\theta$, i.e.,

$$\tan\theta = \frac{x_j}{x_i} \text{ or equivalently } s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}} \text{ and } c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}. \quad (7.1)$$

Thus, unlike the Householder matrices, which introduce lots of zeros by pre-multiplication, the Givens' matrices introduce a single zero in a chosen position by pre-multiplication.

Thus, unlike the Householder matrices, which introduce lots of zeros by pre-multiplication, the Givens' matrices introduce a single zero in a chosen position by pre-multiplication.

Since (7.1) can always be satisfied, we only ever think of Givens' matrices $J(i,j)$ for a specific vector or column with the angle chosen to make a zero in the jth position, e.g., $J(1,2)x$ tacitly implies that we choose

$$\theta = \tan^{-1} x_2/x_1$$

so that the second entry of $J(1,2)x$ is zero.

Thus, unlike the Householder matrices, which introduce lots of zeros by pre-multiplication, the Givens' matrices introduce a single zero in a chosen position by pre-multiplication.

Since (7.1) can always be satisfied, we only ever think of Givens' matrices $J(i,j)$ for a specific vector or column with the angle chosen to make a zero in the jth position, e.g., $J(1,2)x$ tacitly implies that we choose

$$\theta = \tan^{-1} x_2/x_1$$

so that the second entry of $J(1,2)x$ is zero.

Similarly, for a matrix $A \in \mathbb{R}^{m\times n}$, $J(i,j)A := J(i,j,\theta)A$, where

$$\theta = \tan^{-1} a_{ji}/a_{ii},$$

i.e., it is the ith column of A, which is used to define θ so that $(J(i,j)A)_{ji} = 0$.

Thus, unlike the Householder matrices, which introduce lots of zeros by pre-multiplication, the Givens' matrices introduce a single zero in a chosen position by pre-multiplication.

Since (7.1) can always be satisfied, we only ever think of Givens' matrices $J(i,j)$ for a specific vector or column with the angle chosen to make a zero in the jth position, e.g., $J(1,2)x$ tacitly implies that we choose

$$\theta = \tan^{-1} x_2/x_1$$

so that the second entry of $J(1,2)x$ is zero.

Similarly, for a matrix $A \in \mathbb{R}^{m\times n}$, $J(i,j)A := J(i,j,\theta)A$, where

$$\theta = \tan^{-1} a_{ji}/a_{ii},$$

i.e., it is the ith column of A, which is used to define θ so that $(J(i,j)A)_{ji} = 0$. We shall return to these in a later lecture.

Matrix Eigenvalues

Background:

first, an important result from analysis (not proved or examinable!).

Matrix Eigenvalues

Background:

first, an important result from analysis (not proved or examinable!).

Theorem (Ostrowski)

The eigenvalues of a matrix are continuously dependent on the entries. I.e., 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.

Gerschgorin Theorems

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**

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

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

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

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

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

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

"x_k is the largest entry". (7.2)

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

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

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

$$\text{“}x_k \text{ is the largest entry”}. \tag{7.2}$$

Since $x \neq 0$, it follows that $x_k \neq 0$.

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

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

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

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

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

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

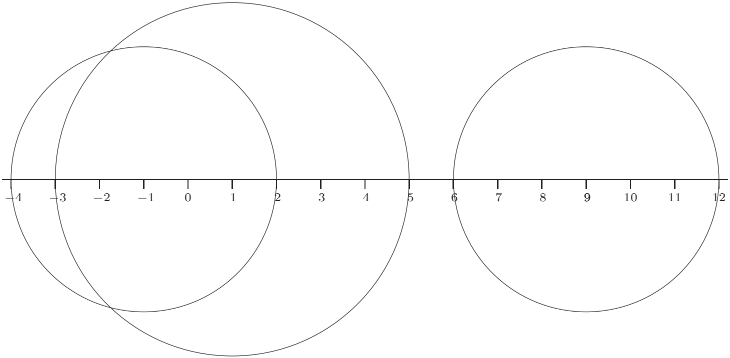
by (7.2). $\square$

Example.

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

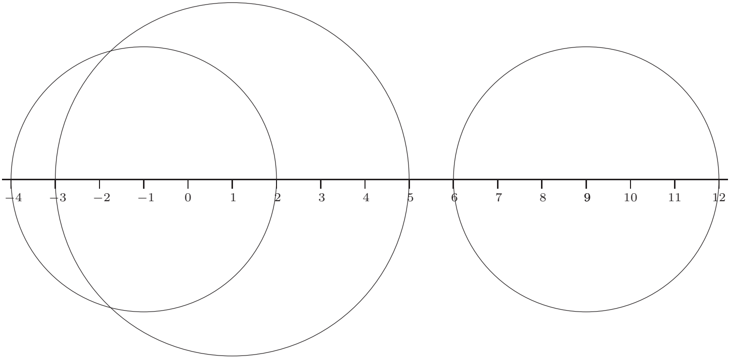
Example.

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



Example.

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



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

Theorem (Gerschgorin's 2nd theorem)

If any union of ℓ Gerschgorin discs is disjoint from the other Gerschgorin discs, then this union contains ℓ eigenvalues.

Proof. Consider $B(\theta) = \theta A + (1 - \theta)D$, where $D = \text{diag}(A)$, 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 $B(1) = A$.

Proof. Consider $B(\theta) = \theta A + (1 - \theta)D$, where $D = \text{diag}(A)$, 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.

Proof. Consider $B(\theta) = \theta A + (1 - \theta)D$, where $D = \text{diag}(A)$, 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.

Proof. Consider $B(\theta) = \theta A + (1-\theta)D$, where $D = \text{diag}(A)$, 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.

Proof. Consider $B(\theta) = \theta A + (1-\theta)D$, where $D = \text{diag}(A)$, 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!). □

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

Power Method/Iteration

Power Method/Iteration is a simple method for calculating a single (largest) eigenvalue of a square matrix A:

Power Method/Iteration

Power Method/Iteration is a simple method for calculating a single (largest) eigenvalue of a square matrix A:

choose arbitrary $y \in \mathbb{R}^n$

set $x_0 = y/\|y\|$ to calculate an initial vector;

then, for $k = 0, 1, \ldots$

compute $y_k = Ax_k$

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

Power Method/Iteration

Power Method/Iteration is a simple method for calculating a single (largest) eigenvalue of a square matrix A:

choose arbitrary $y \in \mathbb{R}^n$
set $x_0 = y/\|y\|$ to calculate an initial vector;
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 **Power Iteration**.

Power Method/Iteration

Power Method/Iteration is a simple method for calculating a single (largest) eigenvalue of a square matrix A:

choose arbitrary $y \in \mathbb{R}^n$
set $x_0 = y/\|y\|$ to calculate an initial vector;
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 **Power Iteration**.

It computes unit vectors in the direction of $x_0, Ax_0, A^2x_0, A^3x_0, \ldots, A^kx_0$.

Suppose 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

$$\boxed{|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|}\ .$$

Suppose 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

$$\boxed{|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|} \; .$$

Then we can write

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

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

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

However, since

$$Av_i = \lambda_i v_i \quad \Longrightarrow \quad A^2 v_i = A(Av_i) = \lambda_i A v_i = \lambda_i^2 v_i,$$

inductively

$$A^k v_i = \lambda_i^k v_i.$$

However, since

$$Av_i = \lambda_i v_i \qquad \Longrightarrow \qquad A^2 v_i = A(Av_i) = \lambda_i A v_i = \lambda_i^2 v_i,$$

inductively

$$A^k v_i = \lambda_i^k v_i.$$

So

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

However, since

$$Av_i = \lambda_i v_i \qquad \Longrightarrow \qquad A^2 v_i = A(Av_i) = \lambda_i A v_i = \lambda_i^2 v_i,$$

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 $k \to \infty$.

However, since

$$Av_i = \lambda_i v_i \quad \Longrightarrow \quad A^2 v_i = A(Av_i) = \lambda_i A v_i = \lambda_i^2 v_i,$$

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 $k \to \infty$. The result is that by normalizing to be a unit vector

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

The sign is identified by looking at, e.g., $(A^k x_0)_1 / (A^{k-1} x_0)_1$.

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!

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 we consider only in the case when A is symmetric (see next lecture).