April 2, 2020

Numerical Analysis

HW 4 will be released at 12am Monday April 6
due 12am Tuesday April 7

Content: everything between HW3 and Jacobi's Method (i.e. lecture on Tuesday March 31).

Jacobi's Method

If A is real symmetric 2×2 matrix:

Then $A = VDV^T$ where V is an orthogonal matrix

$$\Rightarrow \quad V = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

eigenvalues of A.

If $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$, compute φ using a, b, d to diagonalize A.

If A is real symmetric $n\times n$ matrix, then use this idea to zero-out off-diagonal elements.

$$\text{Let } R^{pq}(\varphi) = \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \cos\varphi & & \sin\varphi & & \\ & & & \ddots & & & \\ & & -\sin\varphi & & \cos\varphi & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{array}{l} \\ \\ \text{row } p \\ \\ \text{row } q. \\ \\ \\ \end{array}$$

column p column q

$B = R^{pq}(\varphi)^T A R^{pq}(\varphi)$ has elements $a_{pq} = a_{qp} = 0$ if φ was computed using elements a_{pp}, a_{pq}, a_{qq}

Jacobi's Algorithm

Apply a sequence of $R(\varphi_k)$'s to A that zero out all off-diagonal elements.

Question: Do elements that were set to zero stay zero forever in the Jacobi method? <u>No</u>

<u>Idea behind convergence</u>: Every Jacobi rotation moves some "mass" of the matrix from off-diagonal positions to diagonal positions.

apply R^T and R

What can we say about the convergence of Jacobi's Method?

First a Lemma:

<u>Lemma</u>: If R is an orthogonal transformation, then

$$\|A\|_F = \|R^T A R\|_F$$

Frobenius Norm

$$\|A\|_F = \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2}$$

<u>Proof</u>: Let $B = R^T A R$. Then A and B have the same eigenvalues, and

$$B^2 = (R^T A R)(R^T A R) = R^T A^2 R.$$

$\Rightarrow$ A^2 and B^2 have the same eigenvalues, and therefore $\operatorname{trace}(A^2) = \operatorname{trace}(B^2)$.

<u>But</u> $\|A\|_F^2 = \operatorname{trace}(A^T A) = \operatorname{trace}(\underbrace{AA}_{A^2}) = \operatorname{trace}(B^2) = \|B\|_F^2$.

(This was proven in Homework.) $\square$.

Some notation:

Split the Frobenius norm into two pieces:

Let $\|A\|_F^2 = S(A) = \sum_{i,j} |a_{ij}|^2$

$D(A) = \sum_i |a_{ii}|^2$ diagonal part

$L(A) = \sum_{i\neq j} |a_{ij}|^2$ off-diagonal part

$\Rightarrow \quad S(A) = D(A) + \underline{L(A)} = \|A\|_F^2.$

Theorem Let $A^{(k)}$ be the k^{th} iterate in the Jacobi Algorithm.

Then $\lim_{k\to\infty} L(A^{(k)}) = 0$

$\lim_{k\to\infty} D(A^{(k)}) = \text{trace}(A^2).$

Proof: Let a_{pq} be the off-diagonal element of A with the largest absolute value.

Let $B = R^{pq}(\varphi)^T A R^{pq}(\varphi)$ (a single Jacobi Rotation)

Then $\underbrace{\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix}}_{\tilde{B}} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}^T \underbrace{\begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix}}_{\tilde{A}} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$

and don't forget $b_{pq} = b_{qp} = 0$. (by construction).

But from the lemma, $\|\tilde{B}\|_F^2 = \|R^T \tilde{A} R\|_F^2 = \|\tilde{A}\|_F^2$.

$\Rightarrow$ This implies that $b_{pp}^2 + b_{qq}^2 = a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2$.

Furthermore, $S(\tilde{B}) = D(\tilde{B}) + \underbrace{L(\tilde{B})}_{=0}$

$= S(\hat{A})$

$= D(\hat{A}) + L(\hat{A})$

So this means that $D(\tilde{B}) = D(\tilde{A}) + L(\tilde{A})$

$$= D(\tilde{A}) + 2a_{pq}^2.$$

$$\Rightarrow \quad L(\tilde{B}) = L(\tilde{A}) - 2a_{pq}^2 = 0 \quad \text{for } \hat{A}, \hat{B}.$$

But, the same argument works for A and B, the original $n\times n$ matrices:

$$S(B) = D(B) + \underbrace{L(B)}_{\neq 0 \text{ in general}}$$

$$= S(A)$$

$$= D(A) + L(A)$$

$$\Rightarrow \quad \underset{\uparrow \geq 0}{L(B)} = \underset{\uparrow \geq 0}{L(A)} - 2\underbrace{a_{pq}^2}_{\uparrow > 0}.$$

$$\Rightarrow$$

$$L(B) < L(A)$$

Continuing: Since a_{pq} was the largest off diagonal element of A, we have that $L(A) \leq n(n-1)\, a_{pq}^2$

$$\Leftrightarrow \quad a_{pq}^2 \geq \frac{L(A)}{n(n-1)}$$

Therefore, $L(B) = L(A) - 2a_{pq}^2$

$$\leq L(A) - \frac{2\,L(A)}{n(n-1)}$$

$$= L(A)\left(1 - \frac{2}{n(n-1)}\right)$$

Re-label the matrices:

$$A^{(0)} = A$$

$$A^{(1)} = B$$

$$\Rightarrow \quad L(A^{(1)}) \leq \left(1 - \frac{2}{n(n-1)}\right) L(A^{(0)})$$

$$\Rightarrow \quad L(A^{(2)}) \leq \left(1 - \frac{2}{n(n-1)}\right) L(A^{(1)}) \leq \left(1 - \frac{2}{n(n-1)}\right)^2 L(A^{(0)}).$$

$\Rightarrow$ After k steps, $L(A^{(k)}) \le \underbrace{\left(1 - \frac{2}{n(n-1)}\right)^k}_{<1} L(A^{(0)})$

$\Rightarrow$ So therefore, $L(A^{(k)}) \to 0$ as $k \to \infty$.

And since $S(A^{(k)}) = D(A^{(k)}) + L(A^{(k)})$

$$= \operatorname{trace}(A^2)$$

$$\lim_{k\to\infty} \left(D(A^{(k)}) + L(A^{(k)}) \right) = \lim_{k\to\infty} D(A^{(k)}) = \operatorname{trace}(A^2). \quad \square$$

What gurantees that $\lim_{k\to\infty} D(A^{(k)}) = \operatorname{trace}(A^2)$

implies that $A^{(k)} \to \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$?

<u>Idea</u> Apply Gerschgorin's Theorem :

$A^{(k)}$ and A have the same eigenvalues, and $L(A^{(k)}) \to 0$.

Therefore the Gerschgorin disks of $A^{(k)}$ have radii going to 0 as well. Therefore, the eigenvalues of $A^{(k)}$, and therefore A, are the limit of the diagonal of $A^{(k)}$.

What about the rate of convergence?

We showed that $L(A^{(k)}) \le \left(1 - \frac{2}{n(n-1)}\right)^k L(A^{(0)})$

if $n = 1000$, $1 - \frac{2}{n(n-1)} = .99999799799\ldots$

& if $k = 100$, $\left(1 - \frac{2}{n(n-1)}\right)^k \sim .999799$

$k = 10000$, $(\quad)^k \sim .98$

Real-life convergence is often <u>much</u> faster than indicated in the proof.

One final note:

Jacobi's Algorithm can be terminated when $L(A^{(k)}) \leq \epsilon$

$$\Rightarrow \quad A^{(k)} = \underbrace{R^{pq}(\varphi_k)^T \cdots R^{pq}(\varphi_1)^T}_{R^T} A \underbrace{R^{pq}(\varphi_1) \cdots R^{pq}(\varphi_k)}_{R}$$

$\approx$ diagonal

$$\Rightarrow \quad A = \underbrace{R}\ \underset{\substack{\uparrow \\ \approx \text{ diagonal}}}{A^{(k)}}\ \underbrace{R^T}$$

R diagonalizes A.

$\Rightarrow$ R is the matrix of eigenvectors of A (approximate eigenvectors).

$\Rightarrow$ $A^{(k)}$ has eigenvalues (approximations of) on the diagonal

This means that Jacobi algorithm computes <u>all</u> eigenvalues and eigenvectors at the same time.

Topics on HW4

- matrix and vector norms
- matrix condition numbers
- least squares
  - QR factorization
  - Normal equations
  - SVD factorization (pseudo-inverse)
- eigenvalue computations
  - Gerschgorin Thm
  - power method (w/shifts, inverse power method with shifts)

(No questions on homework about Jacobi's Algorithm).