Honors Numerical Analysis

Lecture 14

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:

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

$$= S(A)$$

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

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

$$\Rightarrow \quad L(B) < L(A)$$

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

What gurantees that $\lim_{k\to\infty} D(A^{(k)}) = \text{trace}(A^2)$
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 $n = 1000$, $1-\frac{2}{n(n-1)} = .99999799799\ldots$

A 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 $L(A^{(k)}) \le \epsilon$

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

$\approx$ diagonal

$$\Rightarrow \quad A = \underbrace{R}\, A^{(k)} \underbrace{R^T}$$

$\uparrow$

$\approx$ diagonal

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 all eigenvalues and eigenvectors at the same time.

The QR Method

For general matrices, the QR Method can be used to find all eigenvalues. Examine algorithm, then analyze:

Algorithm

$A^{(0)} = A$

FOR $k = 1, 2, \ldots$

Factor $A^{(k-1)} = Q^{(k)} R^{(k)}$

Compute $A^{(k)} = R^{(k)} Q^{(k)}$

Under certain assumptions,

$A^{(k)} \rightarrow$ upper triangular,

and since $R^{(k)} = Q^{*(k)} A^{(k-1)}$

$$\Rightarrow \quad A^{(k)} = Q^{*(k)} A^{(k-1)} Q^{(k)}$$

so $A^{(k)}$ and A have the same eigenvalues.

Three things must be done in order for this to be a usable, accelerated algorithm:

1. First reduce A to tridiagonal, or Hessenberg, form using Householder reflections
2. Apply to shifted matrix, $A^{(k)} - \mu^{(k)} I$, with $\mu^{(k)}$ an estimate for some λ.
3. Use deflation — decoupling into smaller problems

To understand the "pure" QR algorithm above, first study subspace iteration (basically power method on groups of vectors):

Subspace Iteration

Pick $\hat{Q}^{(0)} \in \mathbb{R}^{m \times n}$, orthonormal

For $k = 1, 2, \ldots$

Set $Z = A \hat{Q}^{(k-1)}$

Factor $Z = \hat{Q}^{(k)} \hat{R}^{(k)}$

Then $\hat{Q}^{(k)} \rightarrow$ top eigenvalues of A at a rate of

$$\max_{p \in 1,2,\ldots,n} \left(\frac{\lambda_{p+1}}{\lambda_p} \right)^k .$$

More details later...