July 23, 2014

<u>Theorem</u>: For any k, $0 \le k \le r$, define

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$$

Then $\|A - A_k\|_2 = \inf_{rank(B)\le k} \|A - B\|_2 = \sigma_{k+1}$

Pf: First, note that

$$A - A_k = \sum_{j=k+1}^{r} \sigma_j u_j v_j^T = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} 0 & & & & \\ & \sigma_{k+1} & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

It is the SVD of $A - A_k$

$\therefore$ $\|A-A_k\|_2 = \sigma_{k+1}$

Suppose $\exists$ B rank(B) $\le$ k such that

$$\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$$

Then $\exists$ (n-k)-dim subspace W such that

$$w \in W \Rightarrow B w = 0$$

Note $A w = (A-B) w$. Then

$$\|Aw\|_2 = \|(A-B)w\|_2 \le \|A-B\|_2 \, \|w\|_2 < \sigma_{k+1} \|w\|_2.$$

But $\exists$ (k+1)-dim subspace V_{k+1} such that $||Av|| \geq \sigma_{k+1} ||v||$.

e.g. $V_{k+1} = \text{span}\{ v_1, v_2, \ldots, v_{k+1} \}$.

(Note: $A v_j = \sigma_j u_j$, $||A v_j|| = \sigma_j \geq \sigma_{k+1} ||v_j||$.)

But $\dim(W) + \dim(V_{k+1}) > n \rightarrow$ contradiction.

<u>Notes</u>

1) $$A_k = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & \\ & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}$$

$$= U_k \Sigma_k V_k^T$$

2) A_k is the best rank-k approximation of A. The error of approximation is σ_{k+1} (in L_2-norm).

Application: Image compression

- An m×n image can be represented by m×n matrix A where A_{ij} = pixel value at (i,j).

- Compress the image by storing less than mn entries.

- Let $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, the best rank-k approximation of A.

  Keep the first k singular values and use A_k to approx. A; i.e. A_k = compressed image.

- E.g. m = 320, n = 200. To store A_k, only need store $u_1, \ldots, u_k$ and $\sigma_1 v_1, \ldots, \sigma_k v_k \rightarrow$ (m+n)k words.

- To store A, one needs mn words.

- Compression ratio: (m+n)k / mn ≈ k/123 (if m=320, n=200)

k	Rel error σ_{k+1}/σ_1	Compression ratio
3	0.155	2.4%
10	0.077	8.1%
20	0.040	16.3%

Two-phase process

Idea: First reduce the matrix to bidiagonal form. Then it is diagonalized.

$$\underset{A}{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix}} \xrightarrow{\text{Phase 1}} \underset{B}{\begin{bmatrix} \times & \times & & \\ & \times & \times & \\ & & \times & \times \\ & & & \times \\ & & & \end{bmatrix}} \xrightarrow{\text{Phase 2}} \underset{\Sigma}{\begin{bmatrix} \times & & & \\ & \times & & \\ & & \times & \\ & & & \times \\ & & & \end{bmatrix}}$$

Golub-Kahn Bidiagonalization

- Apply Householder reflectors on the left and the right.

$$\underset{A}{\begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix}} \longrightarrow \underset{U_1^T A}{\begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}} \longrightarrow \underset{U_1^T A V_1}{\begin{bmatrix} \times & \times & 0 & 0 \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}} \longrightarrow$$

$$\underset{U_2^T U_1^T A V_1}{\begin{bmatrix} \times & \times & 0 & 0 \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}} \longrightarrow \underset{U_2^T U_1^T A V_1 V_2}{\begin{bmatrix} \times & \times & 0 & 0 \\ 0 & \times & \times & 0 \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}} \longrightarrow \cdots$$

- n reflectors on the left, n-2 on the right.

- flops(bidiag) = 2 × flops(QR) ~ $4mn^2 - 4/3\, n^3$.

Convergence of Iterative Methods

Richardson

The iteration matrix is given by:

$$G^{Rich} \equiv I - M_{Rich}^{-1} A = I - (\theta I)A = I - \theta A$$

Suppose (λ, v) is an eigenpair of A. Then

$$G^{Rich} v = (I - \theta A)v = v - \theta\lambda v = (1 - \theta\lambda)v$$

Hence $\mu \equiv 1 - \theta\lambda$ is an eigenvalue of G^{Rich}.

Lemma: Let λ_{min} and λ_{max} be the smallest and largest eigenvalue of A. Then

$$\rho(G^{Rich}) = \max\{|1 - \theta\lambda_{\min}|, |1 - \theta\lambda_{\max}|\}$$

Pf:

$$\lambda_{min} \leq \lambda \leq \lambda_{max}$$

$$1 - \theta\lambda_{max} \leq 1 - \theta\lambda \leq 1 - \theta\lambda_{min}$$

$$|\mu| \leq \max\{\, |1 - \theta\lambda_{min}|, |1 - \theta\lambda_{max}| \,\}$$

Note

If $\lambda_{min} < 0$ and $\lambda_{max} > 0$, then

either $1 - \theta\lambda_{min} > 1$ $(\theta>0)$ or $1-\theta\lambda_{max} > 1$ $(\theta<0)$

$\Rightarrow$ $\rho(G^{Rich}) > 1$

$\Rightarrow$ Richardson method diverges.

Theorem: Assume all eigenvalues of A are positive. Then Richardson converges if and only if

$$0 < \theta < 2 / \lambda_{max}$$

Pf: If $0 < \theta < 2 / \lambda_{max}$, then

$$0 < \theta \lambda_{min} \leq \theta \lambda_{max} < 2$$

$$-2 < -\theta \lambda_{max} \leq -\theta \lambda_{min} < 0$$

$$-1 < 1 - \theta \lambda_{max} \leq 1 - \theta \lambda_{min} < 1$$

$$\Rightarrow \quad |1 - \theta \lambda_{max}| < 1, \ |1 - \theta \lambda_{min}| < 1$$

$$\Rightarrow \quad \rho(G^{Rich}) < 1$$

Now, assume $\rho(G^{Rich}) < 1$. Then

$$-1 < 1 - \theta \lambda_{max} \leq \mu \leq 1 - \theta \lambda_{min} < 1$$

Right inequality $\Rightarrow$ $\theta > 0$.

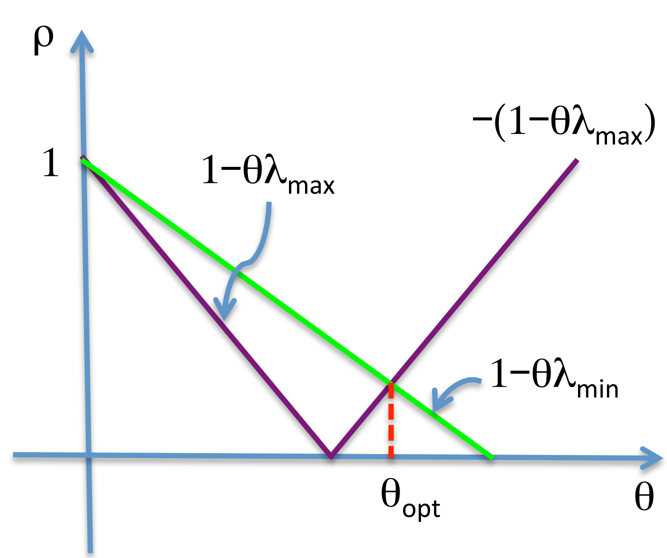
Left inequality $\Rightarrow$ $1 - \theta \lambda_{max} > -1$ $\Rightarrow$ $\theta < 2 / \lambda_{max}$.

Optimal θ

θ_{opt}: $-(1 - \theta \lambda_{max}) = 1 - \theta \lambda_{min}$

$$\theta_{opt} = 2/(\lambda_{min} + \lambda_{max})$$

$$\begin{aligned}\rho_{opt} &= 1 - \theta_{opt} \lambda_{min} \\ &= (\lambda_{min}-\lambda_{max})/(\lambda_{min}+\lambda_{max}) \\ &= (\lambda_{max}/\lambda_{min}-1)/(\lambda_{max}/\lambda_{min}+1) \\ &= (\kappa(A)-1)/(\kappa(A)+1)\end{aligned}$$



Jacobi convergence

Theorem: If A and $2D - A$ are SPD, then Jacobi converges.

Pf: Let μ be an eigenvalue of $I - M_J^{-1} A = I - D^{-1} A$.

$$(I - D^{-1} A)\, v = \mu\, v \qquad \text{for some } v \neq 0$$

$$D^{-1} (D - A)\, v = \mu\, v$$

$$(D - A)\, v = \mu\, D\, v$$

$$v^T (D - A)\, v = \mu\, v^T D\, v$$

$$v^T D\, v - v^T A\, v = \mu\, v^T D\, v$$

$$0 < v^T A\, v = (1 - \mu)\, v^T D\, v \quad \Rightarrow \quad \mu < 1$$

Since $2D - A$ is SPD, $v^T (2D - A)\, v > 0$

$$\therefore \quad 2\, v^T D\, v - v^T A\, v > 0$$

$$v^T D\, v - v^T A\, v > - v^T D\, v$$

$$\mu\, v^T D\, v > -v^T D\, v \qquad (v^T D v > 0 \text{ since } D = \text{SPD})$$

$$\mu > -1$$

Hence $-1 < \mu < 1$

$\Rightarrow \quad \rho(I - D^{-1} A) < 1$

Gauss-Seidel & SOR

Theorem: If A is SPD, then GS & SOR ($0<\omega<2$) converge.

Def: A is an M-matrix if

(i) $a_{ii} > 0$

(ii) $a_{ij} < 0$

(iii) A^{-1} exists and $(A^{-1})_{ij} \geq 0$ $\forall$ i, j.

Theorem: If A is an M-matrix, Jacobi and GS converge. Moreover,

$$\rho(I - M_{GS}^{-1}A) \leq \rho(I - M_J^{-1}A) < 1$$

i.e. the convergence rate of GS is better than that of Jacobi.