July 18, 2014

Theorem: rang(A) = span $\{u_1, \ldots, u_r\}$ and null(A) = span $\{v_{r+1}, \ldots, v_n\}$.

Theorem: $||A||_2 = \sigma_1$ and $||A||_F = \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}$.
(Note: $||A||^2 = \lambda_{max}(A^T A)$, $||A||_F^2 = \text{sum } a_{ij}^2$)

Pf: $A^T A = (V \Sigma U^T)(U \Sigma V^T) = V \Sigma^2 V^T \sim \Sigma^2$

$\therefore \quad \lambda_{max}(A^T A) = \sigma_1^2 \Rightarrow ||A||_2 = \sigma_1$.

$$||A||_F^2 = tr(A^T A) = tr(V \Sigma^2 V^T) = tr((V\Sigma)(V\Sigma)^T)$$
$$= tr((V\Sigma)^T(V\Sigma)) = tr(\Sigma V^T V \Sigma) = tr(\Sigma^2)$$
$$= \sigma_1^2 + \ldots + \sigma_r^2$$

Theorem: The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^TA or AA^T.

Pf: A^TA and AA^T are similar to Σ^2.

Theorem: If $A = A^T$, then $\sigma(A) = \{ |\lambda| : \lambda \in \Lambda(A) \}$. In particular, if A is SPD, then $\sigma(A) = \Lambda(A)$.

Theorem: The condition number of $A = \sigma_1/\sigma_n$, $A \in R^{n\times n}$.

Pf: $\kappa_2(A) \equiv ||A||_2||A^{-1}||_2 = \sigma_1 ||A^{-1}||_2$

$A = U \Sigma V^T \Rightarrow A^{-1} = V \Sigma^{-1} U^T \Rightarrow$ SVD of A^{-1}

$||A^{-1}||_2 = 1/\sigma_n$.

Computing the SVD

- $A = U \Sigma V^T$

  $\therefore\ A^T A = (V \Sigma^T U^T)(U \Sigma V^T)$

  $= V \Sigma^T \Sigma V^T$

  $\therefore$ Eigenvalues of $A^T A = \{ \sigma_i^2 \}$ of A.

- An algorithm:

  (1) Form $A^T A$

  (2) Compute the eigen-decomposition $A^T A = V \Lambda V^T$

  (3) Compute

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & & \end{bmatrix}, \qquad \sigma_i = \sqrt{\lambda_i}, \qquad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

  (4) Solve the equation:

$$U \Sigma = A V$$

  for orthogonal U (by QR factorization)

- Unstable algorithm:

$$\left| \tilde{\sigma}_k - \sigma_k \right| = O(\varepsilon \| A \| \kappa(A))$$

Example: Find the SVD of $A = \begin{bmatrix} 0 & \frac{-1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$

Method 1: $$A^T A = \begin{bmatrix} 9 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

Eigenvalues and eigenvectors are:

$\lambda_1 = 9$, $v_1 = (1,0)^T$, $\lambda_2 = 1/4$, $v_2 = (0,1)^T$.

$\therefore$ $\sigma_1 = 3$, $\sigma_2 = 1/2$,

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \frac{-1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Rightarrow \quad u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

Method 2: Consider AA^T.

Method 3: By inspection, range(A) = span $\{u_1, u_2\}$

where $u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, $u_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$

u_1, u_2 are orthogonal and they have norm 1.

Also the length of principal axes are 3 and 1/2.

$\Rightarrow$ $\sigma_1 = 3$, $\sigma_2 = 1/2$.

By definition, $A v_1 = \sigma_1 u_1$

$$\begin{bmatrix} 0 & \frac{-1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \times \\ \times \end{bmatrix} = 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \Rightarrow \quad v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$A v_2 = \sigma_2 u_2$

$$\begin{bmatrix} 0 & \frac{-1}{2} \\ 3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \times \\ \times \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow v_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\therefore \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Alternative formulation

- Assume A is square, i.e. $m = n$.
- Consider the $2n \times 2n$ symmetric matrix:

$$H = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$$

Since $A = U \Sigma V^T$, $A V = U \Sigma$, $A^T U = V \Sigma^T = V \Sigma$

$$\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} V & V \\ U & -U \end{bmatrix} = \begin{bmatrix} A^T U & -A^T U \\ AV & AV \end{bmatrix}$$

$$= \begin{bmatrix} V\Sigma & -V\Sigma \\ U\Sigma & U\Sigma \end{bmatrix}$$

$$= \begin{bmatrix} V & V \\ U & -U \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}$$

i.e. $H Q = Q \Lambda$ -> eigen-decomposition of H.

- Compute an eigen-decomposition of H. Then $\sigma_A = |\lambda_H|$. U, V can be extracted from the eigenvectors Q.
- Stable algorithm.

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) ~ $4mn^2 - 4/3\, n^3$.

Low-rank approximation

Theorem: A is the sum of r rank-one matrices:

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T$$

Pf:

$$A = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$$= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \\ 0 \end{bmatrix}$$

$$= \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$$

Theorem: 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 $A - A_k$

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

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 $||Av|| \ge \sigma_{k+1} \; ||v||$.

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

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

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

Notes

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 $\sigma_1 v_1, \ldots, \sigma_k v_k$ -> (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%