Lecture 19 Outline

Wednesday, October 30, 2019 2:27 PM

Eigensystem Solutions

SVD

Solution Methods Summary

Eigensystem Solutions

Wednesday, October 30, 2019 2:31 PM

Original: Choose $\underline{v}_0$, then $\underline{A}^k \underline{v}_0$ approaches $\underline{q}_1$

Now, let $\{\underline{v}_0^{(1)} \; \underline{v}_0^{(2)} \; \dots \; \underline{v}_0^{(n)}\}$ be a set of n linearly independent vectors close to the eigenvectors of the n largest eigenvalues:

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > |\lambda_{n+1}| \geq \dots \geq |\lambda_m|$$

then it should be expected that

$$\{\underline{A}^k \underline{v}_0^{(1)} \; \underline{A}^k \underline{v}_0^{(2)} \; \dots \; \underline{A}^k \underline{v}_0^{(n)}\}$$

will approach $\{\underline{q}_1 \; \underline{q}_2 \; \dots \; \underline{q}_n\}$ as $k \to \infty$

Let $\underline{V}_0 = [\underline{v}_0^{(1)} \; \underline{v}_0^{(2)} \; \dots \; \underline{v}_0^{(n)}]$, then

$$\underline{V}_k = \underline{A}^k \underline{V}_0$$

and as $k \to \infty$, then the QR of $\underline{V}_k$

i.e., $\underline{V}_k = \underline{Q}_k \underline{R}_k$ converges to the eigenvectors

if $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n| > |\lambda_{n+1}| \geq \ldots$

holds and $\hat{Q}^T \underline{V}_0$ is non-singular

The true matrix of eigenvectors, <u>not</u> the QR decomposition

Problems: 1) Converges only linearly

2) Stability

To fix stability, orthogonalize every iteration

Algorithm: Simultaneous Iteration

Let $\underline{A}$ be any matrix (or the result of upper Hessenberg)

Let $\underline{Q}_0 \in \mathbb{R}^{m \times n}$ with orthonormal columns for $k = 1, 2, \ldots$

$$\underline{Z} = \underline{A}\,\underline{\hat{Q}}_{k-1} \qquad \text{(make orthonormal)}$$

$$\underline{\hat{Q}}_k \underline{\hat{R}}_k = \underline{Z} \qquad \text{(QR of } \underline{Z}\text{)}$$

end

One can show that if $\hat{Q}_0 = \underline{I}$, then this is the QR method

See Trefethen, Lectures 26-28

Issues with prior methods: Slow convergence

Introduce Inverse Shifts

Focus on symmetric $\underline{A}$

Let $\underline{Q}^{(k)} = \underline{Q}_1 \underline{Q}_2 \cdots \underline{Q}_k$

$$\underline{R}^{(k)} = \underline{R}_k \underline{R}_{k-1} \cdots \underline{R}_1$$

from the QR method for eigenproblems

One can show that

$$\underline{A}^k = \underline{A}\,\underline{Q}^{(k-1)}\underline{R}^{(k-1)} = Q^{(k)}\underline{R}_k\underline{R}^{(k-1)} = \underline{Q}^{(k)}\underline{R}^{(k)}$$

Since $\underline{A} = \underline{A}^T$, then

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

$$= \left((R^{(k)})^T (Q^{(k)})^T \right)^{-1} = (Q^{(k)})^{-T}(R^{(k)})^{-T}$$

$$= \underline{Q}^{(k)}(\underline{R}^{(k)})^{-T}$$

$$= (\underline{A}^{-k})^T$$

Now let

$$\underline{P} = \begin{bmatrix} 0 & & 1 \\ & 1 & \\ 1 & & 0 \end{bmatrix}$$

Thus, $\underline{P}$ flips rows + columns

$$\underline{P}\,\underline{A} = \underline{A}^T$$

then

$$\underline{P}^2 = \underline{I}\ , \quad (\underline{A}^T)^T = \underline{A}$$

$$\Rightarrow (\underline{A}^{-k})^T = \underline{A}^{-k}\,\underline{P} = \underline{Q}^{(k)}\,\underline{P}^2\,(\underline{R}^{(k)})^{-T}\,\underline{P}$$

$$= \underbrace{(\underline{Q}^{(k)}\underline{P})}_{\text{Orthogonal}}\underbrace{[\underline{P}(\underline{R}^{(k)})^{-T}\underline{P}]}_{\text{Upper Triangular}}$$

$\Rightarrow$ This is the QR factorization

$(\underline{A}^{-(k)})^T$ has a QR factorization

$\Rightarrow$ One can use the QR method for eigen problems on $\underline{A}^{-1}$

Algorithm: Shifted QR for Eigenproblems

Let $\underline{A}_0$ be given by $\underline{Q}_0^T\,\underline{A}_0\,\underline{Q}_0 = \underline{A}$

(where $\underline{A}_0$ comes from the Upper Hessenberg algorithm)

for $k = 1, 2, \ldots$

Pick a shift μ_k

$\underline{Q}_k \underline{R}_k = \underline{A}_{k-1} - \mu_k \underline{I}$ (QR of shifted matrix)

$\underline{A}_k = \underline{R}_k \underline{Q}_k + \mu_k \underline{I}$

end

How does one pick μ_k?

Typically, one wants the smallest eigenvalue

Try Rayleigh Quotient on the last column of $\underline{Q}_k$

$\mu_k = \underline{q}_k^{(m)T} \underline{A} \, \underline{q}_k^{(m)}$ with $\underline{q}_k^{(m)}$: last column of $\underline{Q}_k$

It turns out that the value at the (m, m) location at $\underline{A}_k$ is location of $\underline{q}_k$ is

$$\underline{A}_k(m,m) = \underline{q}_k^{(m)T} \underline{A}\, \underline{q}_k^{(m)}$$

$\Rightarrow$ Just set $\mu_k = \underline{A}_k(m,m)$

However sometimes this is not stable, so one can try the Wilkinson Shift

If μ_k is chosen properly, then convergence is 3rd order

See Lecture 29 in Trefethen for details

Singular Value Decomposition (SVD)

Sunday, November 3, 2019 6:30 PM

SVD is an extension of eigensystems to singular and rectangular matrices

Eigenproblems require that $\underline{A}$ be square and defective eigenvalues cause issues for eigen decomposition

Instead, look for the singular values σ and the vectors $\underline{u}$ and $\underline{v}$, such that

$$\underset{m\times n}{\underline{A}}\ \underset{n\times 1}{\underline{v}} = \underset{1\times 1}{\sigma}\ \underset{m\times 1}{\underline{u}}, \quad \underline{A} \in \mathbb{R}^{m\times n}$$

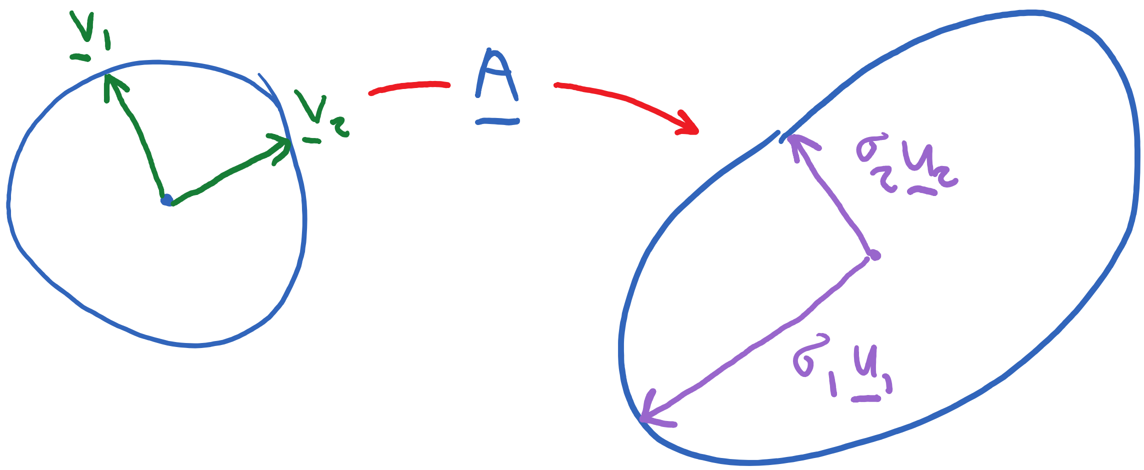
$\underline{v}$ is in the row space of $\underline{A}$

$\underline{u}$ is in the column space of $\underline{A}$

with $r = \text{rank}(\underline{A})$

What do σ, $\underline{u}$ + $\underline{v}$ represent?

Consider application of $\underline{A}$ to the unit circle



The singular decomposition gives the principal directions of the hyperellipses of $\underline{A}$ applied to the unit circle

Let $\underset{n \times r}{\hat{\underline{V}}} = [\underline{v}_1 \ \underline{v}_2 \cdots \underline{v}_r]$, $\underset{m \times r}{\hat{\underline{U}}} = [\underline{u}_1 \ \underline{u}_2 \cdots \underline{u}_r]$

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & 0 \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix} \qquad r = \text{rank}(\underline{A})$$

$$\underline{A}\,\hat{\underline{V}} = \hat{\underline{U}}\,\underline{\Sigma}$$

where both $\hat{\underline{U}}$ and $\hat{\underline{V}}$ are unitary

$$\underset{r\times n}{\hat{\underline{V}}^T}\ \underset{n\times r}{\hat{\underline{V}}} = \underset{r\times r}{\underline{I}}, \quad \underset{r\times m}{\hat{\underline{U}}^T}\ \underset{m\times r}{\hat{\underline{U}}} = \underset{r\times r}{\underline{I}}$$

$$\Rightarrow \underline{A} = \hat{\underline{U}}\,\hat{\underline{\Sigma}}\,\hat{\underline{V}}^T \leftarrow \text{the reduced SVD}$$

If $r < \min(m,n)$

$\Rightarrow$ Non-zero null space

$\Rightarrow$ There is a set of vectors that correspond to the singular values $\sigma = 0$

$$\underline{A}\,\underline{v} = \underline{\sigma}\,\underline{u} = 0\,\underline{u} = \underline{0}$$

$\Rightarrow$ $\underline{v}$ is the null space of $\underline{A}$

The full SVD of $\underline{A}$ is then

$$\underline{A}\,[\underbrace{\underline{v}_1\ \underline{v}_2 \cdots \underline{v}_r}_{\substack{r \text{ vectors}\\ \text{in}\\ \text{row space}}}\ \underbrace{\underline{v}_{r+1} \cdots \underline{v}_n}_{\substack{n-r\\ \text{vectors}\\ \text{in null}}}] = [\underbrace{\underline{u}_1 \cdots \underline{u}_r}_{\substack{r \text{ vectors}\\ \text{in}\\ \text{column}}}\ \underbrace{\underline{u}_{r+1} \cdots \underline{u}_m}_{\substack{m-r\\ \text{vectors}\\ \text{in}}}]\,\underline{\Sigma}$$

space space $\mathrm{Null}(\underline{A}^T)$

$\Rightarrow$ $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$ contains the orthonormal basis for all four matrix subspaces

Formal Definition

Let $\underline{A} \in \mathbb{R}^{m \times n}$ $m \geq n$ not required

$\underline{A}$ might not be full rank

SVD of $\underline{A}$ is given by $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$

$\underline{U} \in \mathbb{R}^{m \times m}$ is unitary

$\underline{V} \in \mathbb{R}^{n \times n}$ is unitary

$\underline{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal

It is also assumed that all σ_j in $\underline{\Sigma}$ are real, non-negative and in non-increasing order

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0 \quad \text{for } p = \min(m,n)$$

To show real + non-negative consider $\underline{A}^T \underline{A}$

$$\underline{A}^T \underline{A} = (\underline{U} \underline{\Sigma} \underline{V}^T)^T (\underline{U} \underline{\Sigma} \underline{V}^T)$$

$$= \underline{V} \underline{\Sigma}^T \underline{U}^T \underline{U} \underline{\Sigma} \underline{V}^T$$

$$= \underline{V} \underline{\Sigma}^2 \underline{V}^T$$

$\rightarrow$ Looks like an eigendecomposition of $\underline{A}^T \underline{A}$

Since $\underline{A}^T \underline{A}$ is normal $\Rightarrow$ $\underline{V}$ is unitary

Now consider $\underline{x}^T (\underline{A}^T \underline{A}) \underline{x}$ for any $\underline{x}$

$$\underline{x}^T (\underline{A}^T \underline{A}) \underline{x} = (\underline{A} \underline{x})^T (\underline{A} \underline{x}) = \underline{y}^T \underline{y} > 0$$

$\Rightarrow$ $\underline{A}^T \underline{A}$ is positive definite

$\Rightarrow$ $\underline{A}^T \underline{A}$ only has positive eigenvalues

Since $\underline{\Sigma}^2$ is the matrix of eigenvalues of $\underline{A}^T\underline{A}$

$\Rightarrow \sigma_j = \sqrt{\lambda_j} \Rightarrow$ will be positive + real

Theorem: Every matrix $\underline{A} \in \mathbb{R}^{m \times n}$ has an SVD and the singular values $\{\sigma_j\}$ are all uniquely determined

If $\underline{A}$ is square and all $\{\sigma_j\}$ are distinct, then $\{\underline{u}_j\}$ and $\{\underline{v}_j\}$ are uniquely determined up to a sign.

Properties:

Let $\underline{A} \in \mathbb{R}^{m \times n}$ with $p = \min(m,n)$

$r = \#$ of singular values $\leq p$

① $\text{rank}(\underline{A}) = r$

② $\text{range}(\underline{A}) = \text{span}(\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r)$

$\text{null}(\underline{A}) = \text{span}(\underline{v}_{r+1}, \underline{v}_{r+2}, \ldots, \underline{v}_n)$

3 $\|\underline{A}\|_2 = \sigma_1$

$\|\underline{A}\|_F = (\sigma_1^2 + \sigma_2^2 + ... + \sigma_r^2)^{1/2}$

4 Non-zero singular values of $\underline{A}$ are the square roots of the eigenvalues of

$$\underline{A}^T\underline{A} \text{ or } \underline{A}\,\underline{A}^T$$

5 If $\underline{A} = \underline{A}^T$, then σ_j is $|\lambda_j|$ of $\underline{A}$

6 If $\underline{A} \in \mathbb{R}^{m \times m}$, then $|\det(\underline{A})| = \prod_{i=1}^{m} \sigma_i$

$\Rightarrow$ If $\underline{A}$ is square, but one $\sigma_i = 0$, then $\underline{A}^{-1}$ does not exist

Because of 2 above, the SVD says that any matrix can be made diagonal if one uses the proper row & column space basis

Consider $\underline{A}\,\underline{x} = \underline{b}$ $\qquad \underline{A} \in \mathbb{R}^{m \times n}$

$\underline{x} \in \mathbb{R}^n, \ \underline{b} \in \mathbb{R}^m$

$\underline{V}$ spans $\mathbb{R}^n$, while $\underline{U}$ spans $\mathbb{R}^m$

$\Rightarrow$ One can write $\underline{x}$ in terms of coordinates of $\underline{V}$

$$\underline{x}' = \underline{V}^T \underline{x} \quad \rightarrow \quad \underline{V}\,\underline{x}' = \underline{V}\,\underline{V}^T \underline{x}$$

$$\therefore \quad \underline{x} = \underline{V}\,\underline{x}'$$

Similarly

$$\underline{b}' = \underline{U}^T \underline{b}$$

$$\underline{A}\,\underline{x} = \underline{b} \Rightarrow \underline{U}^T \underline{A}\,\underline{x} = \underline{U}^T \underline{b}$$

$$\Rightarrow \underline{U}^T \underline{U}\,\underline{\Sigma}\,\underline{V}^T x = \underline{U}^T \underline{b}$$

$$\underline{\Sigma}\,\underline{V}^T \underline{x} = \underline{U}^T \underline{b}$$

$$\underline{\Sigma}\,\underline{x}' = \underline{b}'$$ Coordinates in U

$$\underline{\Sigma}\,\underline{x}' = \underline{b}'$$

Coordinates in $\underline{U}$

Diagonal matrix

Coordinates in $\underline{V}$

Uses of SVD:

① Pseudo-Inverse

All matrices have $\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T$

Define the pseudo-inverse as

$$\underline{A}^+ \underline{A} = \underline{I} = \underline{A}\,\underline{A}^+$$

Note: $\underline{A}^{-1}$ might not exist

Let $\underline{A}^+ = \underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T$ with

$$\underline{\Sigma}^{-1} = \begin{bmatrix} \sigma_1^{-1} & & 0 \\ & \sigma_2^{-1} & \\ 0 & & \ddots \end{bmatrix}$$

Then

$$\underline{A}^+ \underline{A} = (\underline{V}\,\underline{\Sigma}^{-1}\underline{U}^T)(\underline{U}\,\underline{\Sigma}\,\underline{V}^T)$$

Then

$$A^+ A = (V \Sigma^{-1} U^T)(U \Sigma V^T)$$
$$= V \Sigma^{-1} \Sigma V^T = V V^T = I$$

$$A A^+ = (U \Sigma V^T)(V \Sigma^{-1} U^T)$$
$$= U \Sigma \Sigma^{-1} U^T = U U^T = I$$

② Low Rank Approximations

Let

$$\Sigma_j = \begin{bmatrix} 0 & & & & \\ & \ddots & & & 0 \\ & & 0 & & \\ & & & \sigma_j & \\ & 0 & & & 0 \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}$$

Then

$$U \Sigma_j V^T = \sigma_j u_j v_j^T$$ with u_j : jth column of U

v_j^T : jth column of V

$$\Rightarrow U \Sigma V^T = \sum^{r} U \Sigma_j V^T$$

$$\Rightarrow \underline{U}\,\underline{\Sigma}\,\underline{V}^T = \sum_{j=1}^{r} \underline{U}\,\underline{\Sigma}_j\,\underline{V}^T$$

$$= \sum_{j=1}^{r} \sigma_j\,\underline{u}_j\,\underline{v}_j^T$$

$\Rightarrow$ Any matrix $\underline{A}$ can be written as the finite sum of rank 1 matrices

Theorem: Let $\underline{A}_\nu = \sum_{j=1}^{\nu} \sigma_j\,\underline{u}_j\,\underline{v}_j^T$ be a low rank approximation of $\underline{A}$, where $\nu \leq \text{rank}(\underline{A})$

Then, it can be shown that

$$\|\underline{A} - \underline{A}_\nu\|_2 = \inf_{\substack{\underline{B} \in \mathbb{R}^{m \times n} \\ \text{rank}\,\underline{B} \leq \nu}} \|\underline{A} - \underline{B}\|_2 = \sigma_{\nu+1}$$

where $\sigma_{\nu+1} = 0$, if $\nu = p = \min(m,n)$

$\Rightarrow \underline{A}_\nu$ minimizes the error

One also can show that $\underline{A}_\nu$ minimizes the

$\|\underline{A} - \underline{A}_\nu\|_F$ error

To show this in an application, look at compression