Lecture 7: Singular Value Decomposition - II

November 8, 2017

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1}} \mathbf{v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2}} \mathbf{u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t} \mathbf{u_t}^T$, Caution: $\mathbf{v_{t_1}} \mathbf{v_{t_2}^T} \neq$ all-zero matrix !

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1}} \mathbf{v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2}} \mathbf{u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t} \mathbf{u_t}^T$,
  Caution: $\mathbf{v_{t_1}} \mathbf{v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t} \mathbf{u_t}^T$. [check by multiplying.]

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1}} \mathbf{v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2}} \mathbf{u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t} \mathbf{u_t}^T$,
  Caution: $\mathbf{v_{t_1}} \mathbf{v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t} \mathbf{u_t}^T$. [check by multiplying.]
- A singular: "pseudo-inverse" $= A^\dagger = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t} \mathbf{u_t}^T$
  $AA^\dagger = \sum_{t=1}^{r} \mathbf{u_t} \mathbf{u_t}^T =$ identity on space of $\mathbf{u_t}$.

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1}} \mathbf{v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2}} \mathbf{u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t} \mathbf{u_t}^T$,
  Caution: $\mathbf{v_{t_1}} \mathbf{v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t} \mathbf{u_t}^T$. [check by multiplying.]
- A singular: "pseudo-inverse" $= A^\dagger = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t} \mathbf{u_t}^T$
  $AA^\dagger = \sum_{t=1}^{r} \mathbf{u_t} \mathbf{u_t}^T =$ identity on space of $\mathbf{u_t}$.
- A symmetric $\rightarrow$ Spectral Decomposition (SD): $A = \sum_{t=1}^{r} \lambda_t \mathbf{u_t} \mathbf{u_t}^T$:

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1} v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2} u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t u_t}^T$,
  Caution: $\mathbf{v_{t_1} v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$. [check by multiplying.]
- A singular: "pseudo-inverse" $= A^\dagger = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$
  $AA^\dagger = \sum_{t=1}^{r} \mathbf{u_t u_t}^T =$ identity on space of $\mathbf{u_t}$.
- A symmetric $\rightarrow$ Spectral Decomposition (SD): $A = \sum_{t=1}^{r} \lambda_t \mathbf{u_t u_t}^T$:
- **Proof** For A with no ties (except $\pm$) in SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t}^T$.
  $A = A^T \Rightarrow A = \sum_{t=1}^{r} \sigma_t \mathbf{v_t u_t}^T$. "No ties" :$\mathbf{u_1} = \pm \mathbf{v_1}$. Repeat. [The $\pm$ means $\lambda_t = \pm\sigma_t$]

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1} v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2} u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t u_t}^T$, Caution: $\mathbf{v_{t_1} v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$. [check by multiplying.]
- A singular: "pseudo-inverse" $= A^{\dagger} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$
  $AA^{\dagger} = \sum_{t=1}^{r} \mathbf{u_t u_t}^T =$ identity on space of $\mathbf{u_t}$.
- A symmetric $\rightarrow$ Spectral Decomposition (SD): $A = \sum_{t=1}^{r} \lambda_t \mathbf{u_t u_t}^T$:
- **Proof** For A with no ties (except $\pm$) in SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t}^T$. $A = A^T \Rightarrow A = \sum_{t=1}^{r} \sigma_t \mathbf{v_t u_t}^T$. "No ties" :$\mathbf{u_1} = \pm \mathbf{v_1}$. Repeat. [The $\pm$ means $\lambda_t = \pm \sigma_t$]
- $A^s = \sum_{t=1}^{r} \lambda_t^s \mathbf{u_t u_t}^T$. If $|\lambda_1| > |\lambda_t| \forall t \geq 2$ $\lim_{s \to \infty} \frac{1}{|\lambda_1|^s} A^s = \mathbf{u_1 u_1}^T$. Rank 1!

Using Orthogonality of singular vectors

- SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t^T}$. $\mathbf{v_t}$ orthonormal. Many uses.
- $AA^T = \left(\sum_{t_1=1}^{r} \sigma_{t_1} \mathbf{u_{t_1} v_{t_1}^T}\right) \left(\sum_{t_2=1}^{r} \sigma_{t_2} \mathbf{v_{t_2} u_{t_2}^T}\right) = \sum_{t=1}^{r} \sigma_t^2 \mathbf{u_t u_t}^T$, Caution: $\mathbf{v_{t_1} v_{t_2}^T} \neq$ all-zero matrix !
- A invertible $\Rightarrow A^{-1} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$. [check by multiplying.]
- A singular: "pseudo-inverse" $= A^{\dagger} = \sum_{t=1}^{r} \frac{1}{\sigma_t} \mathbf{v_t u_t}^T$
  $AA^{\dagger} = \sum_{t=1}^{r} \mathbf{u_t u_t}^T =$ identity on space of $\mathbf{u_t}$.
- A symmetric $\rightarrow$ Spectral Decomposition (SD): $A = \sum_{t=1}^{r} \lambda_t \mathbf{u_t u_t}^T$:
- **Proof** For A with no ties (except $\pm$) in SVD: $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t v_t}^T$. $A = A^T \Rightarrow A = \sum_{t=1}^{r} \sigma_t \mathbf{v_t u_t}^T$. "No ties" :$\mathbf{u_1} = \pm \mathbf{v_1}$. Repeat. [The $\pm$ means $\lambda_t = \pm \sigma_t$]
- $A^s = \sum_{t=1}^{r} \lambda_t^s \mathbf{u_t u_t}^T$. If $|\lambda_1| > |\lambda_t| \forall t \geq 2$ $\lim_{s \to \infty} \frac{1}{|\lambda_1|^s} A^s = \mathbf{u_1 u_1}^T$. Rank 1!
- **Power Method, Fundamental Thm of Markov Chains**

Best Low-rank approximation

- SVD of $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$. Define for each k, $A_k = \sum_{t=1}^{k} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$ (called the **truncated SVD**). Will show: for each k, A_k is the best approximation to A of rank at most k.

Best Low-rank approximation

- SVD of $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$. Define for each k, $A_k = \sum_{t=1}^{k} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$ (called the **truncated SVD**). Will show: for each k, A_k is the best approximation to A of rank at most k.
- **Lemma** The rows of A_k are the projections of the (corres.) rows of A onto V_k, the space spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$.

Best Low-rank approximation

- SVD of $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$. Define for each k, $A_k = \sum_{t=1}^{k} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$ (called the **truncated SVD**). Will show: for each k, A_k is the best approximation to A of rank at most k.
- **Lemma** The rows of A_k are the projections of the (corres.) rows of A onto V_k, the space spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$.
- Proof: For any row vector $\mathbf{a}$, the proj of $\mathbf{a}$ onto V_k is given by $\sum_{t=1}^{k} (\mathbf{a} \cdot \mathbf{v_t}) \mathbf{v_t}$. So, projecting each row of A into V_k, we get $\sum_{t=1}^{k} A \mathbf{v_i} \mathbf{v_i}^T$.

Best Low-rank approximation

- SVD of $A = \sum_{t=1}^{r} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$. Define for each k, $A_k = \sum_{t=1}^{k} \sigma_t \mathbf{u_t} \mathbf{v_t}^T$ (called the **truncated SVD**). Will show: for each k, A_k is the best approximation to A of rank at most k.
- **Lemma** The rows of A_k are the projections of the (corres.) rows of A onto V_k, the space spanned by $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_k}$.
- Proof: For any row vector **a**, the proj of **a** onto V_k is given by $\sum_{t=1}^{k} (\mathbf{a} \cdot \mathbf{v_t}) \mathbf{v_t}$. So, projecting each row of A into V_k, we get $\sum_{t=1}^{k} A \mathbf{v_i} \mathbf{v_i}^T$.
- $\sum_{i=1}^{k} A \mathbf{v_i} \mathbf{v_i}^T = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T = A_k$, proving Lemma.

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$
- Let B minimize $\|A - B\|_F^2$ among all rank k or less matrices. $V =$ row space of B.

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$
- Let B minimize $\|A - B\|_F^2$ among all rank k or less matrices. $V =$ row space of B.
- Each row of B is the projection of the corresponding row of A onto V (why?):

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$
- Let B minimize $\|A - B\|_F^2$ among all rank k or less matrices. $V =$ row space of B.
- Each row of B is the projection of the corresponding row of A onto V (why?):
- Otherwise replace row of B with the projection of the corresponding row of A onto V. Keeps row space of $B \subseteq V$, so $\text{rank}(B) \leq k$. But, $\|A - B\|_F^2$ reduced. $\rightarrow\leftarrow$ ($||A - B||_F$ best).

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$
- Let B minimize $\|A - B\|_F^2$ among all rank k or less matrices. $V =$ row space of B.
- Each row of B is the projection of the corresponding row of A onto V (why?):
- Otherwise replace row of B with the projection of the corresponding row of A onto V. Keeps row space of $B \subseteq V$, so $\text{rank}(B) \leq k$. But, $\|A - B\|_F^2$ reduced. $\rightarrow\leftarrow$ ($||A - B||_F$ best).
- So, $||A - B||_F^2 =$ sum of squared distances of rows of A to V, so $||A - B||_F^2 \geq ||A - A_k||_F^2$. QED

Best rank k approximation

- **Theorem** For any matrix B of rank at most k, $\|A - A_k\|_F \leq \|A - B\|_F$
- Let B minimize $\|A - B\|_F^2$ among all rank k or less matrices. $V =$ row space of B.
- Each row of B is the projection of the corresponding row of A onto V (why?):
- Otherwise replace row of B with the projection of the corresponding row of A onto V. Keeps row space of $B \subseteq V$, so rank(B) $\leq k$. But, $\|A - B\|_F^2$ reduced. $\rightarrow\leftarrow$ ($||A - B||_F$ best).
- So, $||A - B||_F^2 =$ sum of squared distances of rows of A to V, so $||A - B||_F^2 \geq ||A - A_k||_F^2$. QED
- Will also prove A_k is best rank k approx in spectral norm.

Left singular vectors are orthogonal

- **Theorem** The left singular vectors are pairwise orthogonal.

Left singular vectors are orthogonal

- **Theorem** The left singular vectors are pairwise orthogonal.
- **Proof** Suppose not. Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$.

Left singular vectors are orthogonal

- **Theorem** The left singular vectors are pairwise orthogonal.
- **Proof** Suppose not. Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$.
- Wlg, $\mathbf{u_i}^T\mathbf{u_j} = \delta > 0$. If $\mathbf{u_i}^T\mathbf{u_j} < 0$.....

Left singular vectors are orthogonal

- **Theorem** The left singular vectors are pairwise orthogonal.
- **Proof** Suppose not. Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$.
- Wlg, $\mathbf{u_i}^T\mathbf{u_j} = \delta > 0$. If $\mathbf{u_i}^T\mathbf{u_j} < 0$.....
- $j > i$ since i smallest violation.

Left singular vectors are orthogonal

- **Theorem** The left singular vectors are pairwise orthogonal.
- **Proof** Suppose not. Let i be the smallest integer such that $\mathbf{u_i}$ is not orthogonal to some other $\mathbf{u_j}$.
- Wlg, $\mathbf{u_i}^T\mathbf{u_j} = \delta > 0$. If $\mathbf{u_i}^T\mathbf{u_j} < 0$.....
- $j > i$ since i smallest violation.
- For $\varepsilon > 0$, let $\mathbf{v_i'} = \frac{\mathbf{v_i}+\epsilon\mathbf{v_j}}{|\mathbf{v_i}+\epsilon\mathbf{v_j}|}$. $|\mathbf{v_i'}| = 1$ and $A\mathbf{v_i'} = \frac{\sigma_i\mathbf{u_i}+\varepsilon\sigma_j\mathbf{u_j}}{\sqrt{1+\varepsilon^2}}$ has length at least as large as its component along $\mathbf{u_i}$ which is $\mathbf{u_i^T}\left(\frac{\sigma_i\mathbf{u_1}+\varepsilon\sigma_i\mathbf{u_i}}{\sqrt{1+\varepsilon^2}}\right) > (\sigma_i + \varepsilon\sigma_i\delta)\left(1 - \frac{\varepsilon^2}{2}\right) > \sigma_i - \frac{\varepsilon^2}{2}\sigma_i + \varepsilon\sigma_i\delta - \frac{\varepsilon^3}{2}\sigma_i\delta > \sigma_i$, for sufficiently small ϵ, a contradiction since $\mathbf{v_i} + \varepsilon\mathbf{v_j}$ is orthogonal to $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{i-1}}$ since $j > i$ and σ_i is defined to be the maximum of $|A\mathbf{v}|$ over such vectors. □

A Lemma

- **Lemma** $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.

A Lemma

- **Lemma** $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Proof: SVD: $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. item <3-> $A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. Let $\mathbf{v}$ be the top singular vector of $A - A_k$. Express $\mathbf{v}$ is a lin. combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ (Why?): $\mathbf{v} = \sum_{j=1}^{r} c_j \mathbf{v_j}$.

A Lemma

- **Lemma** $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Proof: SVD: $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. item <3-> $A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. Let $\mathbf{v}$ be the top singular vector of $A - A_k$. Express $\mathbf{v}$ is a lin. combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ (Why?): $\mathbf{v} = \sum_{j=1}^{r} c_j \mathbf{v_j}$.

A Lemma

- **Lemma** $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Proof: SVD: $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. item <3-> $A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. Let $\mathbf{v}$ be the top singular vector of $A - A_k$. Express $\mathbf{v}$ is a lin. combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ (Why?): $\mathbf{v} = \sum_{j=1}^{r} c_j \mathbf{v_j}$.
- $|(A - A_k)\mathbf{v}| = \left| \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \sum_{j=1}^{r} c_j \mathbf{v_j} \right| = \left| \sum_{i=k+1}^{r} c_i \sigma_i \mathbf{u_i} \mathbf{v_i}^T \mathbf{v_i} \right| = \left| \sum_{i=k+1}^{r} c_i \sigma_i \mathbf{u_i} \right| = \sqrt{\sum_{i=k+1}^{r} c_i^2 \sigma_i^2}$.

A Lemma

- **Lemma** $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Proof: SVD: $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. item <3-> $A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$. Let $\mathbf{v}$ be the top singular vector of $A - A_k$. Express $\mathbf{v}$ is a lin. combination of $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_r}$ (Why?): $\mathbf{v} = \sum_{j=1}^{r} c_j \mathbf{v_j}$.
- $|(A - A_k)\mathbf{v}| = \left| \sum_{i=k+1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \sum_{j=1}^{r} c_j \mathbf{v_j} \right| = \left| \sum_{i=k+1}^{r} c_i \sigma_i \mathbf{u_i} \mathbf{v_i}^T \mathbf{v_i} \right| = \left| \sum_{i=k+1}^{r} c_i \sigma_i \mathbf{u_i} \right| = \sqrt{\sum_{i=k+1}^{r} c_i^2 \sigma_i^2}$.
- For $\mathbf{v}$ maximizing this, subject to $|\mathbf{v}|^2 = \sum_{i=1}^{r} c_i^2 = 1$, have $c_{k+1} = 1$, rest $c_i = 0$. QED

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.
- Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$.

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.
- Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$.
- Assume rank(A) $\geq k$. $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Null space of B, has dimension $\geq d - k$. Dimension counts imply $\exists \mathbf{z} \neq 0$ in Null $(B) \cap$ Span $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+1}}\}$ .. Scale $\mathbf{z}$ to be of length one.

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.
- Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$.
- Assume rank(A) $\geq k$. $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Null space of B, has dimension $\geq d - k$. Dimension counts imply $\exists \mathbf{z} \neq 0$ in Null$\,(B) \cap$ Span$\,\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+1}}\}$ .. Scale $\mathbf{z}$ to be of length one.
- $\|A - B\|_2^2 \geq |(A - B)\,\mathbf{z}|^2$.

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.
- Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$.
- Assume rank(A) $\geq k$. $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Null space of B, has dimension $\geq d - k$. Dimension counts imply $\exists \mathbf{z} \neq 0$ in $\text{Null}\,(B) \cap \text{Span}\,\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+1}}\}$ .. Scale $\mathbf{z}$ to be of length one.
- $\|A - B\|_2^2 \geq |(A - B)\,\mathbf{z}|^2$.
- Since $B\mathbf{z} = 0$, $\|A - B\|_2^2 \geq |A\mathbf{z}|^2$.

Finally: A_k best in spectral norm too

- For B of rank $\leq k$ $\|A - A_k\|_2 \leq \|A - B\|_2$.
- Proof: If A is of rank k or less, the theorem is obviously true since $\|A - A_k\|_2 = 0$.
- Assume rank(A) $\geq k$. $\|A - A_k\|_2^2 = \sigma_{k+1}^2$.
- Null space of B, has dimension $\geq d - k$. Dimension counts imply $\exists \mathbf{z} \neq 0$ in Null $(B) \cap$ Span $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+1}}\}$ .. Scale $\mathbf{z}$ to be of length one.
- $\|A - B\|_2^2 \geq |(A - B)\,\mathbf{z}|^2$.
- Since $B\mathbf{z} = 0$, $\|A - B\|_2^2 \geq |A\mathbf{z}|^2$.
- Since $\mathbf{z} \in$ Span $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{k+1}}\}$ $|A\mathbf{z}|^2 = \left|\sum_{i=1}^{n} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \mathbf{z}\right|^2 = \sum_{i=1}^{n} \sigma_i^2 \left(\mathbf{v_i}^T \mathbf{z}\right)^2 = \sum_{i=1}^{k+1} \sigma_i^2 \left(\mathbf{v_i}^T \mathbf{z}\right)^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \left(\mathbf{v_i}^T \mathbf{z}\right)^2 = \sigma_{k+1}^2$. QED