§9.6 Singular Value Decomposition

Let $A \in \mathbb{R}^{m \times n}$. The SVD of A is

$$A = U\,S\,V^T$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices; and $S \in \mathbb{R}^{m \times n}$ is non-zero with non-negative entries along the main diagonal.

Rank and linear independence

- **Def:** The RANK of $A \in \mathbb{R}^{m \times n}$, denoted **Rank**$(A)$, is the number of linearly independent ROWS in A.
- **Thm:** **Rank**$(A) =$ the number of linearly independent COLUMNS in A.
- **Def:** The NULLITY of A, denoted **Nullity**(A), is $n -$ **Rank**(A), and describes the largest set of linearly independent vectors $\mathbf{v}$ in $\mathbb{R}^n$ for which $A\mathbf{v} = \mathbf{0}$.

Thm: Let $A \in \mathbb{R}^{m \times n}$

- Matrices $A^T A$ and $A A^T$ are symmetric.
- $\textbf{Nullity}(A) = \textbf{Nullity}(A^T A)$.
- $\textbf{Rank}(A) = \textbf{Rank}(A^T A)$
- Eigenvalues of $A^T A$ and $A A^T$ are real and non-negative.
- The nonzero eigenvalues of $A^T A$ and $A A^T$ are the same.

Proof on Nullity(A): For any $\mathbf{v}$ in $\mathbb{R}^n$:

$$\begin{aligned} A\mathbf{v} &= \mathbf{0}, \quad \longrightarrow \quad A^T A\mathbf{v} = \mathbf{0}, \\ A^T A\mathbf{v} &= \mathbf{0}, \quad \longrightarrow \quad \|A\mathbf{v}\|_2^2 = \mathbf{v}^T A^T A\mathbf{v} = 0, \quad \longrightarrow \quad A\mathbf{v} = \mathbf{0}. \end{aligned}$$

Proof on Rank(A):

$$\begin{aligned} \textbf{Rank}(A) + \textbf{Nullity}(A) &= n, \\ \textbf{Rank}(A^T A) + \textbf{Nullity}(A^T A) &= n. \end{aligned}$$

Thm: Let $A \in \mathbb{R}^{m \times n}$

- Matrices $A^T A$ and $A A^T$ are symmetric.
- $\textbf{Nullity}(A) = \textbf{Nullity}(A^T A)$.
- $\textbf{Rank}(A) = \textbf{Rank}(A^T A)$
- Eigenvalues of $A^T A$ and $A A^T$ are real and non-negative.
- The nonzero eigenvalues of $A^T A$ and $A A^T$ are the same.

Proof on eigenvalues: Let λ be non-zero eigenvalue of $A^T A$. Then λ is real. Let $\mathbf{v} \in \mathbb{R}^n$ be unit eigenvector:

- $$A^T A \mathbf{v} = \lambda \mathbf{v}.$$

  Therefore $\lambda = \mathbf{v}^T (\lambda \mathbf{v}) = \mathbf{v}^T \left(A^T A \mathbf{v}\right) = \|A \mathbf{v}\|_2^2 > 0.$

- $$\left(A A^T\right)(A \mathbf{v}) = A \left(A^T A \mathbf{v}\right) = \lambda \, (A \mathbf{v}).$$

  But $A \mathbf{v} \neq \mathbf{0}$. Thus λ is eigenvalue of $A A^T$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

$$D = \mathbf{diag}(d_1, d_2, \cdots, d_n) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

$$\text{So} \quad A^T A \mathbf{v}_j = d_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \textbf{diag}\left(d_1, d_2, \cdots, d_n\right) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = d_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $s_j = \sqrt{d_j}$ for $j = 1, \cdots, n$. Let k be such that $s_k > 0$ and $s_{k+1} = 0$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V\, D\, V^T$ be eigendecomposition, with

  $$D = \textbf{diag}\left(d_1, d_2, \cdots, d_n\right) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A\, \mathbf{v}_j = d_j\, \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $s_j = \sqrt{d_j}$ for $j = 1, \cdots, n$. Let k be such that $s_k > 0$ and $s_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{s_j} A\, \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T\, \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A\, \mathbf{v}_j\right) = \frac{d_j}{s_j^2} = 1$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V D V^T$ be eigendecomposition, with

  $$D = \mathbf{diag}\left(d_1, d_2, \cdots, d_n\right) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

  $$\text{So} \quad A^T A \mathbf{v}_j = d_j \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $s_j = \sqrt{d_j}$ for $j = 1, \cdots, n$. Let k be such that $s_k > 0$ and $s_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{s_j} A \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A \mathbf{v}_j\right) = \frac{d_j}{s_j^2} = 1$.
  - $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T \mathbf{u}_j = \frac{1}{s_i s_j} \mathbf{v}_i^T \left(A^T A \mathbf{v}_j\right) = 0, \; i \neq j$.

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V\, D\, V^T$ be eigendecomposition, with

$$D = \mathbf{diag}\left(d_1, d_2, \cdots, d_n\right) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

$$\text{So} \quad A^T A\, \mathbf{v}_j = d_j\, \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $s_j = \sqrt{d_j}$ for $j = 1, \cdots, n$. Let k be such that $s_k > 0$ and $s_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{s_j} A\, \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T\, \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A\, \mathbf{v}_j\right) = \frac{d_j}{s_j^2} = 1.$
  - $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T\, \mathbf{u}_j = \frac{1}{s_i\, s_j} \mathbf{v}_i^T \left(A^T A\, \mathbf{v}_j\right) = 0, \; i \neq j.$
- Choose $U = (\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}, \cdots \mathbf{u}_m) \in \mathbb{R}^{m \times m}$ orthogonal.

$$\text{Then } A\, V = U\, S \text{ with } \; S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Constructing the SVD for $A \in \mathbb{R}^{m \times n}$ with $m \geq n$.

- Let $A^T A = V\, D\, V^T$ be eigendecomposition, with

$$D = \mathbf{diag}\,(d_1, d_2, \cdots, d_n) \quad \text{with} \quad d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$$

  be eigenvalues and $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$ be orthogonal.

$$\text{So} \quad A^T A\, \mathbf{v}_j = d_j\, \mathbf{v}_j \quad \text{for} \quad j = 1, \cdots, n.$$

- Define $s_j = \sqrt{d_j}$ for $j = 1, \cdots, n$. Let k be such that $s_k > 0$ and $s_{k+1} = 0$.
- For $j = 1, \cdots, k$, define $\mathbf{u}_j = \frac{1}{s_j} A\, \mathbf{v}_j$.
  - $\mathbf{u}_j$ is unit vector: $\mathbf{u}_j^T\, \mathbf{u}_j = \frac{1}{s_j^2} \mathbf{v}_j^T \left(A^T A\, \mathbf{v}_j\right) = \frac{d_j}{s_j^2} = 1.$
  - $(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ column orthogonal: $\mathbf{u}_i^T\, \mathbf{u}_j = \frac{1}{s_i\, s_j} \mathbf{v}_i^T \left(A^T A\, \mathbf{v}_j\right) = 0, \; i \neq j.$
- Choose $U = (\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}, \cdots \mathbf{u}_m) \in \mathbb{R}^{m \times m}$ orthogonal.

$$\text{Then } A\, V = U\, S \text{ with } S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}. \quad \text{So } A = U\, S\, V^T.$$

Eckart-Young Theorem: For $A \in \mathbb{R}^{m \times n}$ with $m \geq n$

$$\text{Let the SVD of } A = U\,S\,V^T, \quad \text{where}$$

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m \times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n \times n}$

be orthogonal; and $S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m \times n}$ be diagonal

with $s_1 \geq s_2 \cdots \geq s_n \geq 0$.

Eckart-Young Theorem: For $A \in \mathbb{R}^{m\times n}$ with $m \geq n$

Let the SVD of $A = U\,S\,V^T$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $s_1 \geq s_2 \cdots \geq s_n \geq 0$.

For any $1 \leq k \leq n$, the rank-k TRUNCATED SVD of A is

$$A_k \stackrel{def}{=} (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k) \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_k \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)^T .$$

Eckart-Young Theorem: For $A \in \mathbb{R}^{m\times n}$ with $m \geq n$

Let the SVD of $A = U\,S\,V^T$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $s_1 \geq s_2 \cdots \geq s_n \geq 0$.

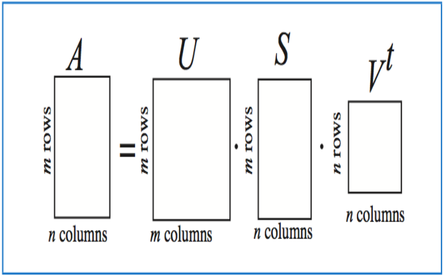
For any $1 \leq k \leq n$, the rank-k TRUNCATED SVD of A is

$$A_k \stackrel{def}{=} (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k) \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_k \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)^T.$$

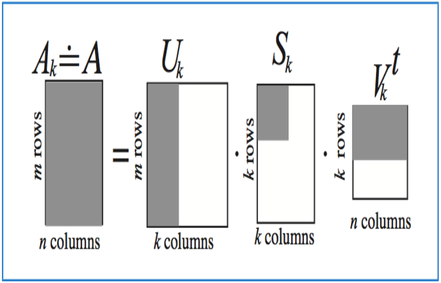
Then

$$\mathbf{min}_{B\in\mathbb{R}^{m\times n},\ \mathbf{Rank}(B)\leq k} \|A - B\|_2 = \|A - A_k\|_2 = s_{k+1}.$$

- SVD



- TRUNCATED SVD



Prf: $\min_{B\in\mathbb{R}^{m\times n},\ \mathbf{Rank}(B)\le k} \|A-B\|_2 = \|A-A_k\|_2 = s_{k+1}$

- Let $B\in\mathbb{R}^{m\times n}$ with $\mathbf{Rank}(B)\le k$.
  Then Null(B) has dimension $p = n - \mathbf{Rank}(B) \ge n-k$.

Prf: $\min_{B \in \mathbb{R}^{m\times n},\ \textbf{Rank}(B)\le k} \|A - B\|_2 = \|A - A_k\|_2 = s_{k+1}$

- Let $B \in \mathbb{R}^{m\times n}$ with $\textbf{Rank}(B) \le k$.
  Then Null(B) has dimension $p = n - \textbf{Rank}(B) \ge n - k$.
- Let $(\mathbf{w}_1, \cdots \mathbf{w}_p)$ be a basis of Null(B) and $\mathbf{x}$ be ANY unit vector in Null(B). There exists a vector $\mathbf{y} \in \mathbb{R}^p$ so that

$$\mathbf{x} = (\mathbf{w}_1, \cdots \mathbf{w}_p)\ \mathbf{y}, \quad B\,\mathbf{x} = 0.$$

Prf: $\min_{B\in\mathbb{R}^{m\times n},\ \mathbf{Rank}(B)\leq k} \|A - B\|_2 = \|A - A_k\|_2 = s_{k+1}$

- Let $B \in \mathbb{R}^{m\times n}$ with $\mathbf{Rank}(B) \leq k$.
  Then $\text{Null}(B)$ has dimension $p = n - \mathbf{Rank}(B) \geq n - k$.
- Let $(\mathbf{w}_1, \cdots \mathbf{w}_p)$ be a basis of $\text{Null}(B)$ and $\mathbf{x}$ be ANY unit vector in $\text{Null}(B)$. There exists a vector $\mathbf{y} \in \mathbb{R}^p$ so that

$$\mathbf{x} = (\mathbf{w}_1, \cdots \mathbf{w}_p)\ \mathbf{y}, \quad B\,\mathbf{x} = 0.$$

- CHOOSE $\mathbf{x} \in \text{Null}(B) \bigcap \text{Range}(\mathbf{v}_1, \cdots \mathbf{v}_{k+1})$:

$$\mathbf{x} = (\mathbf{w}_1, \cdots \mathbf{w}_p)\ \mathbf{y} = (\mathbf{v}_1, \cdots \mathbf{v}_{k+1})\ \mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^{k+1}, \quad \text{or}$$

$$(\mathbf{w}_1, \cdots \mathbf{w}_p, \mathbf{v}_1, \cdots \mathbf{v}_{k+1}) \begin{pmatrix} \mathbf{y} \\ -\mathbf{z} \end{pmatrix} = \mathbf{0}.$$

Prf: $\min_{B\in\mathbb{R}^{m\times n},\ \mathbf{Rank}(B)\le k} \|A - B\|_2 = \|A - A_k\|_2 = s_{k+1}$

- Let $B \in \mathbb{R}^{m\times n}$ with $\mathbf{Rank}(B) \le k$.
  Then Null(B) has dimension $p = n - \mathbf{Rank}(B) \ge n - k$.
- Let $(\mathbf{w}_1, \cdots \mathbf{w}_p)$ be a basis of Null(B) and $\mathbf{x}$ be ANY unit vector in Null(B). There exists a vector $\mathbf{y} \in \mathbb{R}^p$ so that

$$\mathbf{x} = (\mathbf{w}_1, \cdots \mathbf{w}_p)\ \mathbf{y}, \quad B\,\mathbf{x} = 0.$$

- CHOOSE $\mathbf{x} \in \text{Null}(B) \bigcap \text{Range}(\mathbf{v}_1, \cdots \mathbf{v}_{k+1})$:

$$\mathbf{x} = (\mathbf{w}_1, \cdots \mathbf{w}_p)\ \mathbf{y} = (\mathbf{v}_1, \cdots \mathbf{v}_{k+1})\ \mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^{k+1}, \quad \text{or}$$

$$(\mathbf{w}_1, \cdots \mathbf{w}_p, \mathbf{v}_1, \cdots \mathbf{v}_{k+1}) \begin{pmatrix} \mathbf{y} \\ -\mathbf{z} \end{pmatrix} = \mathbf{0}.$$

  Coefficient matrix has n rows but $p + k + 1 \ge n + 1$ columns. There exists a non-zero solution with $\|\mathbf{x}\|_2 = \|\mathbf{z}\|_2 = 1$.

Prf: $\min_{B\in\mathbb{R}^{m\times n},\ \mathbf{Rank}(B)\le k}\|A-B\|_2=\|A-A_k\|_2=s_{k+1}$

- Let $B\in\mathbb{R}^{m\times n}$ with $\mathbf{Rank}(B)\le k$.
  Then Null(B) has dimension $p=n-\mathbf{Rank}(B)\ge n-k$.
- Let $(\mathbf{w}_1,\cdots\mathbf{w}_p)$ be a basis of Null(B) and $\mathbf{x}$ be ANY unit vector in Null(B). There exists a vector $\mathbf{y}\in\mathbb{R}^p$ so that

$$\mathbf{x}=(\mathbf{w}_1,\cdots\mathbf{w}_p)\ \mathbf{y},\quad B\,\mathbf{x}=0.$$

- CHOOSE $\mathbf{x}\in$ Null$(B)\bigcap$ Range$(\mathbf{v}_1,\cdots\mathbf{v}_{k+1})$:

$$\mathbf{x}=(\mathbf{w}_1,\cdots\mathbf{w}_p)\ \mathbf{y}=(\mathbf{v}_1,\cdots\mathbf{v}_{k+1})\ \mathbf{z},\quad \mathbf{z}\in\mathbb{R}^{k+1},\quad \text{or}$$

$$(\mathbf{w}_1,\cdots\mathbf{w}_p,\mathbf{v}_1,\cdots\mathbf{v}_{k+1})\begin{pmatrix}\mathbf{y}\\-\mathbf{z}\end{pmatrix}=\mathbf{0}.$$

  Coefficient matrix has n rows but $p+k+1\ge n+1$ columns.
  There exists a non-zero solution with $\|\mathbf{x}\|_2=\|\mathbf{z}\|_2=1$.

$$\begin{aligned}\|A-B\|_2 &\ge \|(A-B)\,\mathbf{x}\|_2=\|A\,\mathbf{x}\|_2\\ &= \left\|(\mathbf{u}_1,\mathbf{u}_2,\cdots,\mathbf{u}_{k+1})\begin{pmatrix}s_1&&\\&\ddots&\\&&s_{k+1}\end{pmatrix}\mathbf{z}\right\|_2\ge s_{k+1}.\end{aligned}$$

Hilbert Matrix $H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n} & \cdots & \frac{1}{2\,n-3} & \frac{1}{2\,n-2} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2\,n-2} & \frac{1}{2\,n-1} \end{pmatrix} \in \mathbb{R}^{n\times n}$

- H is SPD: For any non-zero $\mathbf{x} = (x_0, x_1, \cdots, x_{n-1})^T \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{x}^T H \mathbf{x} &= \sum_{i,j=0}^{n-1} \frac{x_i\, x_j}{i+j+1} = \sum_{i,j=0}^{n-1} x_i\, x_j \int_0^1 y^{i+j}\, d\,y \\ &= \int_0^1 \left(\sum_{i,j=0}^{n-1} x_i\, x_j y^{i+j} \right) d\,y = \int_0^1 \left(\sum_{i=0}^{n-1} x_i\, y^i \right)^2 d\,y > 0. \end{aligned}$$

Hilbert Matrix $H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n} & \cdots & \frac{1}{2n-3} & \frac{1}{2n-2} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-2} & \frac{1}{2n-1} \end{pmatrix} \in \mathbb{R}^{n \times n}$

- H is SPD: For any non-zero $\mathbf{x} = (x_0, x_1, \cdots, x_{n-1})^T \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{x}^T H \mathbf{x} &= \sum_{i,j=0}^{n-1} \frac{x_i x_j}{i+j+1} = \sum_{i,j=0}^{n-1} x_i x_j \int_0^1 y^{i+j} \, d\, y \\ &= \int_0^1 \left(\sum_{i,j=0}^{n-1} x_i x_j y^{i+j} \right) d\, y = \int_0^1 \left(\sum_{i=0}^{n-1} x_i y^i \right)^2 d\, y > 0. \end{aligned}$$

- In eigendecomposition $H = U D U^T$, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) \in \mathbb{R}^{n \times n}$ is orthogonal;

$$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \in \mathbb{R}^{n \times n} \text{ diagonal, } d_1 \geq d_2 \cdots \geq d_n > 0.$$

Hilbert Matrix $H = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n} & \frac{1}{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{n-1} & \frac{1}{n} & \cdots & \frac{1}{2\,n-3} & \frac{1}{2\,n-2} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2\,n-2} & \frac{1}{2\,n-1} \end{pmatrix} \in \mathbb{R}^{n\times n}$

- H is SPD: For any non-zero $\mathbf{x} = (x_0, x_1, \cdots, x_{n-1})^T \in \mathbb{R}^n$,

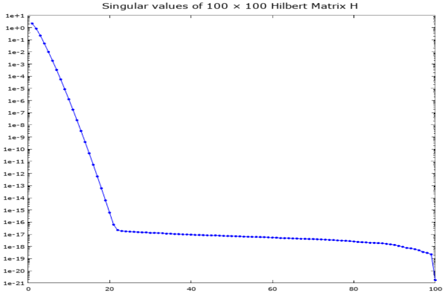
$$\begin{aligned} \mathbf{x}^T H \mathbf{x} &= \sum_{i,j=0}^{n-1} \frac{x_i\, x_j}{i+j+1} = \sum_{i,j=0}^{n-1} x_i\, x_j \int_0^1 y^{i+j}\, d\, y \\ &= \int_0^1 \left(\sum_{i,j=0}^{n-1} x_i\, x_j y^{i+j} \right) d\, y = \int_0^1 \left(\sum_{i=0}^{n-1} x_i\, y^i \right)^2 d\, y > 0. \end{aligned}$$

- In eigendecomposition $H = U\, D\, U^T$, where $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n) \in \mathbb{R}^{n\times n}$ is orthogonal;

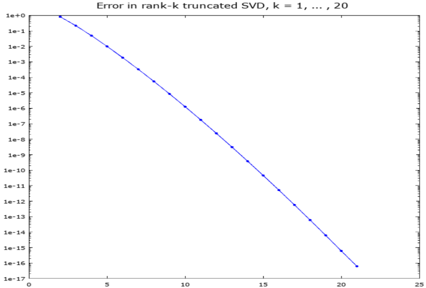
$$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \in \mathbb{R}^{n\times n} \text{ diagonal, } d_1 \geq d_2 \cdots \geq d_n > 0.$$

- $H = U\, D\, U^T$ is SVD of H.

- SVD of Hilbert Matrix



- TRUNCATED SVD of Hilbert Matrix



Computation of SVD

Let $A = U\,S\,V^T$ be the SVD of A, where $S = \begin{pmatrix} D \\ 0 \end{pmatrix}$, with D non-negative diagonal. Then

$$\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} 0 & 0 & D \\ 0 & 0 & 0 \\ D & 0 & 0 \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^T, \quad \text{but}$$

$$\begin{pmatrix} 0 & 0 & D \\ 0 & 0 & 0 \\ D & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \\ 0 & I & 0 \\ -\frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \\ 0 & I & 0 \\ -\frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} -D & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & D \end{pmatrix}$$

$$\text{Therefore} \quad \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} = \widehat{U}\,\widehat{D}\,\widehat{U}^T, \quad \text{where}$$

$$\widehat{U} = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \\ 0 & I & 0 \\ -\frac{I}{\sqrt{2}} & 0 & \frac{I}{\sqrt{2}} \end{pmatrix}, \quad \widehat{D} = \begin{pmatrix} -D & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & D \end{pmatrix}.$$

Solving Least Squares Problem with SVD (I)

Let the SVD of $A = U\,S\,V^T$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $s_1 \geq s_2 \cdots \geq s_n \geq 0$.

Solving Least Squares Problem with SVD (I)

Let the SVD of $A = U\,S\,V^T$, where

$U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \in \mathbb{R}^{m\times m}$ and $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n) \in \mathbb{R}^{n\times n}$

be orthogonal; and $S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n}$ be diagonal

with $s_1 \geq s_2 \cdots \geq s_n \geq 0$.

The Least Squares Problem (LS) is

$$\mathbf{min}_{\mathbf{x}\in\mathbb{R}^n} \|A\,\mathbf{x} - \mathbf{b}\|_2 \quad \text{for a given } \mathbf{b} \in \mathbb{R}^m.$$

Taking gradient,
$$\begin{aligned} \nabla\left(\|A\,\mathbf{x} - \mathbf{b}\|_2^2\right) &= \nabla\left((A\,\mathbf{x} - \mathbf{b})^T (A\,\mathbf{x} - \mathbf{b})\right) \\ &= 2A^T\,(A\,\mathbf{x} - \mathbf{b}) = \mathbf{0}. \end{aligned}$$

Thus the LS solution satisfies $A^T\,A\,\mathbf{x} = A^T\,\mathbf{b}$.

Solving Least Squares Problem with SVD (II)

$$
\begin{aligned}
A &= U\,S\,V^T \\
&= (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T .
\end{aligned}
$$

Solving Least Squares Problem with SVD (II)

$$\begin{aligned} A &= U\, S\, V^T \\ &= (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_n \\ \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)^T. \end{aligned}$$

$$A^T A\, \mathbf{x} = A^T \mathbf{b} \Longleftrightarrow V\, S^T\, S \left(V^T \mathbf{x}\right) = V\, S^T \left(U^T \mathbf{b}\right).$$

Define $s_j^\dagger = \begin{cases} s_j^{-1}, & \text{if } s_j > 0, \\ 0, & \text{otherwise.} \end{cases}$ and $S^\dagger = \begin{pmatrix} s_1^\dagger & & & \mathbf{0} \\ & \ddots & & \vdots \\ & & s_n^\dagger & \mathbf{0} \end{pmatrix}$.

Equation solves to

$$\mathbf{x} = V\, \mathbf{z}, \quad z_j = s_j^\dagger \left(U^T \mathbf{b}\right)_j, \qquad 1 \le j \le n. \quad \text{or}$$

$\mathbf{x} = V\, S^\dagger\, U^T\, \mathbf{b} \stackrel{def}{=} A^\dagger\, \mathbf{b}$, where $A^\dagger = V\, S^\dagger\, U^T$ = PSEUDO-INVERSE of A.

Polynomial Least Squares Model for US Population

Fitting data $\{(x_i, y_i)\,,\,, i = 1, \cdots n\}$ with order $m-1$ polynomial:

$$
\begin{aligned}
& \mathbf{min}_{\alpha_0, \alpha_1, \cdots, \alpha_{m-1}} \sum_{i=1}^{n} \left(y_i - \left(\alpha_0 + \alpha_1\, x_i + \cdots \alpha_{m-1}\, x_i^{m-1}\right)\right)^2 \\
= \quad & \mathbf{min}_{\mathbf{x}} \left\| \mathbf{b} - A\,\mathbf{x} \right\|_2^2\,, \quad \text{where}
\end{aligned}
$$

$$
\mathbf{b} \stackrel{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \stackrel{def}{=} \begin{pmatrix} 1 & x_1 & \cdots & x_1^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} \end{pmatrix}, \quad \mathbf{x} \stackrel{def}{=} \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{m-1} \end{pmatrix}.
$$

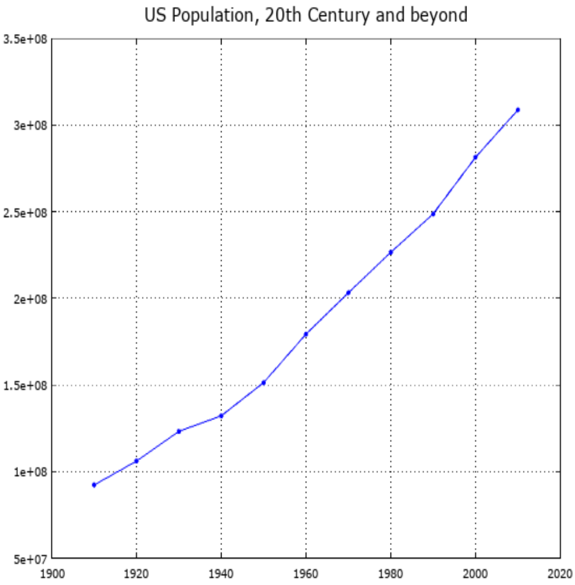
DIRECT MODEL: $y_i =$ *US population in year* x_i.

LOG MODEL: $\mathbf{exp}(y_i) =$ *US population in year* x_i.

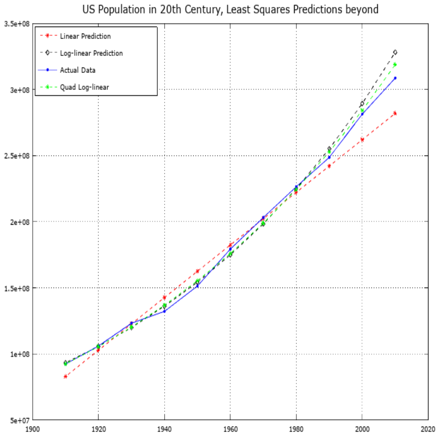
Least Squares Solution: $\min_{\mathbf{x}} \|\mathbf{b} - A\,\mathbf{x}\|_2^2$

- Optimality Condition/Normal equation:

$$\min_{\mathbf{x}} \|\mathbf{b} - A\,\mathbf{x}\|_2^2 \iff A^T\left(A\,\mathbf{x} - \mathbf{b}\right) = \mathbf{0}.$$


US Population, 20th Century and beyond
3.5e+08
3e+08
2.5e+08
2e+08
1.5e+08
1e+08
5e+07
1900
1920
1940
1960
1980
2000
2020

Predicting years 2000, 2010 with data through 1990



- Log-linear Model predicts year 2000 better than Linear Model.
- Quadratic Log Model best.