ECEn 671: Mathematics of Signals and Systems
Moon: Chapter 7.

Randal W. Beard

Brigham Young University

November 19, 2020

Table of Contents

Singular Value Decomposition

Pseudo Inverse and the SVD

SVD and Numerically Sensitive Problems

Rank Reducing Approximations

Applications

Section 1

Singular Value Decomposition

Singular Value Decomposition

Theorem (Moon Theorem 7.1)

Every matrix $A \in \mathbb{C}^{m\times n}$ can be factored as $A = U\Sigma V^H$ where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with diagonal elements $\sigma_1 \geq \sigma_2 \geq cdots \geq \sigma_p \geq 0$.

The diagonal elements are called the singular values of A. If A is real then U and V are real and orthogonal.

Singular Value Decomposition, Proof

Note that the A^HA is Hermitian, and positive definite because $x^HA^HAx = \|Ax\|^2 \geq 0 \qquad \forall x \in \mathbb{C}^n$.

So, from Chapter 6 we know that the eigenvalues of A^HA are real with $m_i = q_i$ for each λ_i.

Let $(\lambda_i, \mathbf{v}_i)$ be an eigenpairs of A^HA then

$$A^HAV = V\Lambda \qquad\qquad V\text{-unitary}$$

where

$$V = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}, \qquad\qquad \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

with

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

Singular Value Decomposition, Proof

Since the $rank(A^H A) \leq \min(m, n) = p$, then number of non-zero eigenvalues is $r \leq p$.
For $1 \leq i \leq r$ let $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}$.
Then

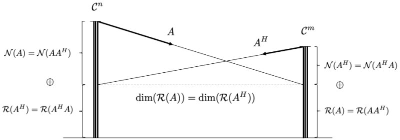
$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \left\langle \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}, \frac{A\mathbf{v}_j}{\sqrt{\lambda_j}} \right\rangle = \frac{1}{\sqrt{\lambda_i \lambda_j}} \mathbf{v}_i^H A^H A \mathbf{v}_j$$
$$= \frac{\lambda_j}{\sqrt{\lambda_i \lambda_j}} \mathbf{v}_i^H \mathbf{v}_j = \delta_{ij}$$

Use Gram-Schmidt to extend $\mathbf{u}_1, \ldots, \mathbf{u}_r$ to $[\mathbf{u}_1, \ldots, \mathbf{u}_m]$ such that $U = [\mathbf{u}_1, \ldots \mathbf{u}_m]$ is unitary.

Singular Value Decomposition, Proof

Lemma

If $(\lambda_i, \mathbf{v}_i)$ is an eigenpair of $A^H A$, then $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}$ are eigenvectors of AA^H.



Proof.

Note that since $\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i \qquad i = 1, \ldots, p$ then

$$
\begin{aligned}
& \mathbf{u}_i \in \mathcal{R}(A) \qquad i = 1, \ldots, p \\
\Longrightarrow\ & \mathbf{u}_i \in \mathcal{N}(A^H) \qquad i = p+1, \ldots, m \\
\Longrightarrow\ & \mathbf{u}_i \in \mathcal{N}(AA^H) \qquad i = p+1, \ldots, m \\
\Longrightarrow\ & AA^H \mathbf{u}_i = 0 \cdot \mathbf{u}_i = 0 \\
\Longrightarrow\ & (0, \mathbf{u}_i) \text{ is an eigenpair of } AA^H \qquad i = p+1, \ldots, m
\end{aligned}
$$

Singular Value Decomposition, Proof

Now lets look at

$$U^H AV = \begin{pmatrix} \mathbf{u}_1^H \\ \vdots \\ \mathbf{u}_m^H \end{pmatrix} A \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1^H A\mathbf{v}_1 & \cdots & \mathbf{u}_1^H A\mathbf{v}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_m^H A\mathbf{v}_1 & \cdots & \mathbf{u}_m^H \mathbf{v}_n \end{pmatrix}.$$

The $(i,j)^{th}$ element of $U^H AV$ is $\mathbf{u}_i^H A\mathbf{v}_j$.

If $i \leq p$ then

$$\begin{aligned} \mathbf{u}_i^H A\mathbf{v}_j &= \frac{1}{\sqrt{\lambda_i}} \mathbf{v}_i^H A^H A\mathbf{v}_j \\ &= \frac{\lambda_j}{\sqrt{\lambda_i}} \mathbf{v}_i^H \mathbf{v}_j = \sqrt{\lambda_j}\delta_{ij} \end{aligned}$$

Singular Value Decomposition, Proof

If $i > p$, then

$$\begin{aligned} \mathbf{u}_i \in \mathcal{N}(A^H) &\Rightarrow A^H \mathbf{u}_i = 0 \\ &\Rightarrow \mathbf{u}_i^H A = 0 \\ &\Rightarrow \mathbf{u}_i^H A \mathbf{v}_j = 0 \end{aligned}$$

Therefore

$$U^H A V = \Sigma$$

where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$ is real and diagonal, where $\sigma_j = 0$ when $j > p$. Therefore

$$A = U \Sigma V^H$$

as required. $\square$

Singular Value Decomposition

Note that the singular values of A are the square root of the eigenvalues of $A^H A$ and AA^H.

Also note that we can write

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$$

where

$$\Sigma_1 = \underbrace{\mathrm{diag}(\sigma_1, \ldots, \sigma_p)}_{\mathbb{R}^{r \times r}}$$

$$\Sigma_2 = 0$$

Singular Value Decomposition

Then

$$
\begin{aligned}
A &= \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} \\
&= \underbrace{U_1}_{m\times p} \underbrace{\Sigma_1}_{p\times p} \underbrace{V_1^H}_{n\times p} \qquad \leftarrow \text{alternate form of SVD} \\
&= \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^H \qquad \leftarrow \text{alternate form of SVD}
\end{aligned}
$$

where $\mathbf{u}_i$'s are orthonormal and $\mathbf{v}_i$'s are orthonormal.

Singular Value Decomposition and Matrix Norm

Note that

$$
\begin{aligned}
\|A\|_2 &= \sup_{\|x\|_2=1} \|Ax\|_2 = \sup_{\|x\|_2=1} \sqrt{x^H A^H A x} \\
&= \sup_{\|x\|_2=1} \sqrt{x^H V_1 \Sigma_1 U_1^H U_1 \Sigma_1 V_1^H x} \\
&= \sup_{\|x\|_2=1} \sqrt{x^H V_1 \Sigma_1^2 V_1^H x} \\
&= \sup_{\|x\|_2=1} \sqrt{\begin{pmatrix} x^H \mathbf{v}_1 & \cdots & x^H \mathbf{v}_r \end{pmatrix} \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^H x \\ \vdots \\ \mathbf{v}_p^H x \end{pmatrix}} \\
&= \sigma_1,
\end{aligned}
$$

where the minimizer is $x = \mathbf{v}_1$.

Singular Value Decomposition and Rank

Lemma

If $A \in \mathbb{C}^{m \times n}$, then $\mathrm{rank}(A) = p$ where p is the number of non-zero singular values.

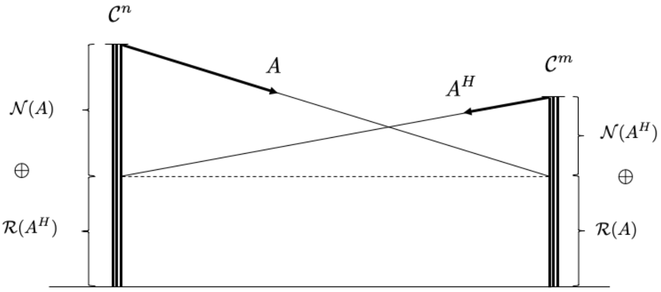
Proof.

$$\mathrm{rank}(A) = \mathrm{rank}(U\Sigma V^H) = \mathrm{rank}(\Sigma)$$

since U and V are both full rank. Clearly $\mathrm{rank}(\Sigma) = p$. $\square$

Singular Value Decomposition and Fundamental Subspaces

Fundamental subspace diagram:



$\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^H)) = \text{rank}(A)$

Question: What information does the SVD provide?

Answer: The SVD completely characterizes all of the spaces.

Singular Value Decomposition and Fundamental Subspaces

Given that

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 \Sigma_1 V_1^H.$$

Let $y \in \mathcal{R}(A)$, then $\exists x \in \mathbb{C}^n$ such that $y = Ax$. Which implies that

$$\begin{aligned} y &= U_1 \Sigma_1 V_1^H x \\ &= U_1 z \text{ where } z = \Sigma_1 V_1^H x \\ &= [\mathbf{u}_1 \cdots \mathbf{u}_p] \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix} = \mathbf{u}_1 z_1 + \cdots + \mathbf{u}_p z_p \end{aligned}$$

$$\implies y \in span\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$$

$$\implies \boxed{\mathcal{R}(A) = span(U_1)}$$

Singular Value Decomposition and Fundamental Subspaces

Since the columns of U_2 are orthonormal to U_1 and $\text{span}(U) = \mathbb{C}^m$ and $\mathcal{R}(A) \oplus \mathcal{N}(A^H) = \mathbb{C}^m$ we must have that

$$\boxed{\mathcal{N}(A^H) = \text{span}(U_2)}$$

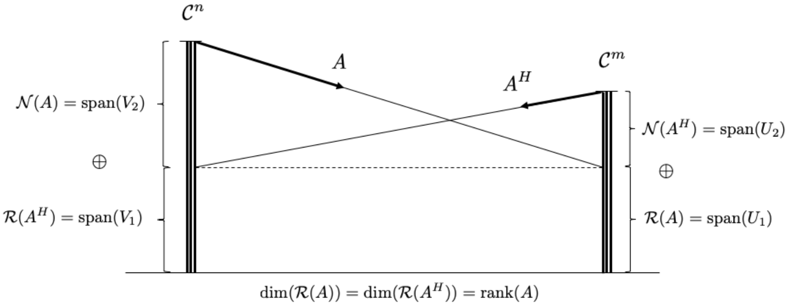
A similar argument shows that

$$\boxed{\mathcal{R}(A^H) = \text{span}(V_1)}$$

$$\boxed{\mathcal{N}(A) = \text{span}(V_2)}$$

Singular Value Decomposition and Fundamental Subspaces

Therefore, the fundamental subspace diagram becomes



Section 2

Pseudo Inverse and the SVD

Pseudo Inverses of A

Least squares solution to $Ax = b$ (i.e. $\min \|Ax - b\|_2$) where A-tall is

$$\hat{x} = (A^H A)^{-1} A^H b \triangleq A^\dagger b.$$

Minimum norm solution to $Ax = b$ (i.e. $\min \|x\|$ for $Ax = b$) where A-fat is

$$x = A^H (AA^H)^{-1} b \triangleq A^\dagger b.$$

How does the SVD help compute the pseudo inverse. We will consider both when A is full rank, and when A is not full rank.

SVD and Least Squared: Full Rank A

Assume $A \in \mathbb{C}^{m \times n}$ is tall, i.e., $m > n$, and that $\text{rank}(A) = n$. Then

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^H = U_1 \Sigma V^H$$

where $U_1 \in \mathbb{C}^{m \times n}$, $\Sigma \in \mathbb{R}^{n \times n}$, and $V \in \mathbb{C}^{n \times n}$.

Then

$$\begin{aligned} (A^H A)^{-1} A^H &= (V \Sigma U_1^H U_1 \Sigma V^H)^{-1} V \Sigma U_1^H \\ &= (V \Sigma^2 V^H)^{-1} V \Sigma U_1^H \\ &= V \Sigma^{-2} V^H V \Sigma U_1^H \\ &= V \Sigma^{-1} U_1^H \end{aligned}$$

where $\Sigma^{-1} = \text{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n})$.

SVD and min Norm: Full Rank A

Assume $A \in \mathbb{C}^{m\times n}$ is fat, i.e., $m < n$, and that $\mathrm{rank}(A) = m$. Then

$$\begin{aligned} A &= U \begin{pmatrix} \Sigma & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} \\ &= U \Sigma V_1^H \end{aligned}$$

where $U \in \mathbb{C}^{m\times m}$, $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_m)$, $V_1 \in \mathbb{C}^{n\times m}$.

Then

$$\begin{aligned} A^H(AA^H)^{-1} &= V_1\Sigma U^H(U\Sigma V_1^H V_1 \Sigma U^H)^{-1} \\ &= V_1\Sigma U^H(U\Sigma^2 U^H)^{-1} \\ &= V_1 \Sigma U^H U \Sigma^{-2} U^H \\ &= V_1 \Sigma^{-1} U^H \end{aligned}$$

where $\Sigma^{-1} = \mathrm{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_m})$

SVD and Pseudo Inverse: Not Full Rank A

Assume $A \in \mathbb{C}^{m\times n}$ and that $\mathrm{rank}(A) = p < \min(m, n)$. Then

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 \Sigma V_1^H$$

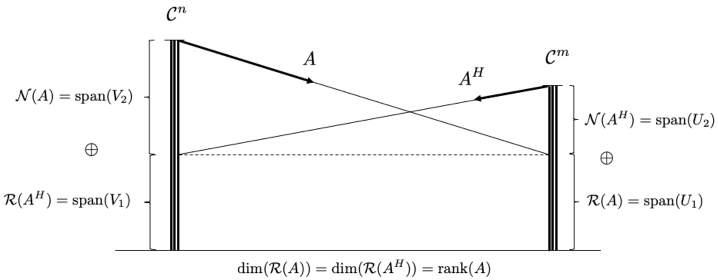
where $U_1 \in \mathbb{C}^{m\times p}$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{p\times p}$, $V_1 \in \mathbb{C}^{n\times p}$.
Consider the least squares problem

$$\begin{aligned} \hat{x} &= (A^H A)^{-1} A^H b \\ &= (V_1 \Sigma_1 U_1^H U_1 \Sigma_1 V_1^H)^{-1} V_1 \Sigma_1 U_1^H b \\ &= (V_1 \Sigma_1^2 V_1^H)^{-1} V_1 \Sigma_1 U_1^H b \\ &= V_1 \Sigma_1^{-2} V_1^H V_1 \Sigma_1 U_1^H b \\ &= V_1 \Sigma_1^{-1} U_1^H b \end{aligned}$$

where $\Sigma_1 = \mathrm{diag}(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_p})$.

SVD and Pseudo Inverse: Not Full Rank A

So we can compute it, but what did we do? How do we interpret the solution since the inverse of $A^H A$ does not exist?



Find a solution to $Ax = b$ where $b \in \mathcal{R}(A)$. But $\mathcal{N}(A) \neq \{0\}$ implies that there are more than one solution.

Therefore, find the minimum norm x that minimizes $\|Ax - b\|_2$.

SVD and Pseudo Inverse: Not Full Rank A

Note the following:

$$\underbrace{U_1}_{m \times p} : \mathbb{C}^p \to \mathcal{R}(A) \subset \mathbb{C}^m$$

so that

$$U_1^* = U_1^H : \mathbb{C}^m \to \mathbb{C}^p.$$

Also,

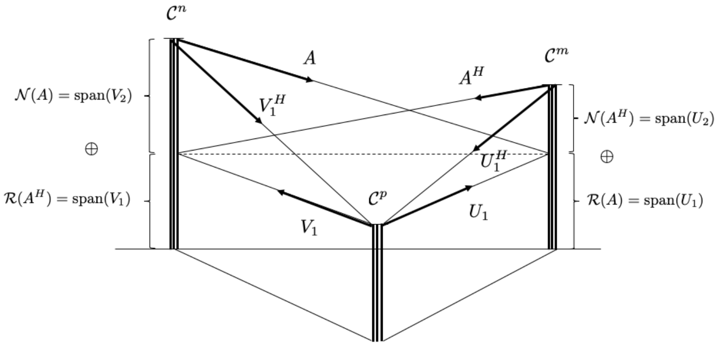
$$\underbrace{V_1}_{n \times p} : \mathbb{C}^p \to \mathcal{R}(A^H) \subset \mathbb{C}^n$$

so that

$$V_1^H : \mathbb{C}^n \to \mathbb{C}^p.$$

SVD and Pseudo Inverse: Not Full Rank A

So we have the following:

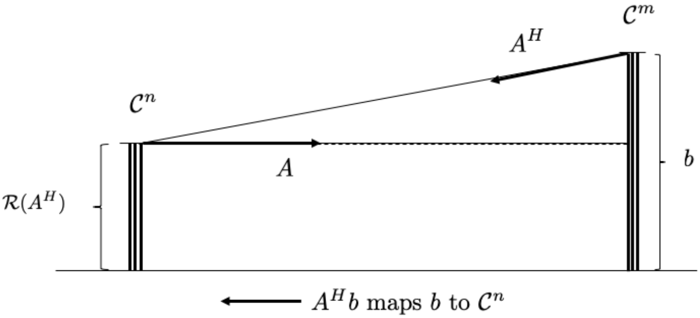


Since $rank(A) = p$ we can only take inverses in $\mathbb{C}^p$. Therefore instead of solving $Ax = b$ directly in $\mathbb{C}^n$ and $\mathbb{C}^m$ we go indirectly through $\mathbb{C}^p$.

SVD and Pseudo Inverse: Not Full Rank A

Step 1: Least Squares

Recall that to solve $\min \|Ax - b\|_2$ when A is full rank:

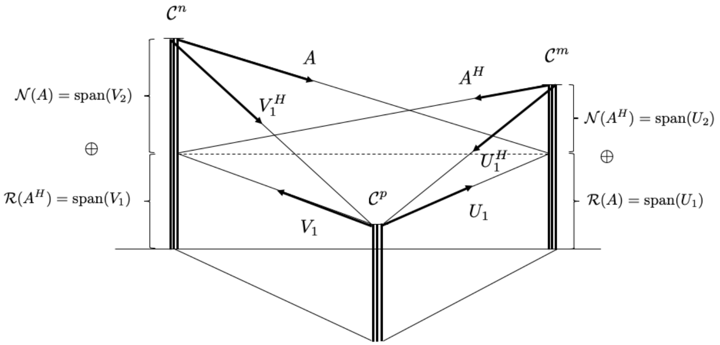


where we can invert things, i.e.

$$A^H A x = A^H b$$
$$\implies \hat{x} = (A^H A)^{-1} A^H b.$$

SVD and Pseudo Inverse: Not Full Rank A

So we have the following:



Now instead of A^H we use U_1^H to map to $\mathbb{C}^p$, i.e., given

$$Ax = b$$

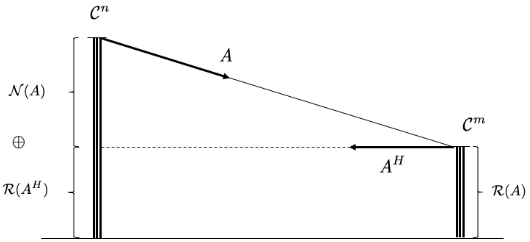
map to $\mathbb{C}^p$ using U_1^H to get:

$$U_1^H A x = U_1^H b \qquad \in \mathbb{C}^p.$$

SVD and Pseudo Inverse: Not Full Rank A

Step 2: Minimum Norm

Recall that to min $\|x\|$ such that $Ax = b$, A-full rank,



to minimize $\|x\|$ we zero out the part that is in the null space of A, i.e. let

$$x = A^H z \text{ where } z \in \mathbb{C}^m$$

then

$$AA^H z = b \qquad \Rightarrow \qquad z = (AA^H)^{-1} b$$

so that

$$\hat{x} = A^H (AA^H)^{-1} b.$$

SVD and Pseudo Inverse: Not Full Rank A

In our case, again pick x to zero the portion in the null space of A. Let

$$x = V_1 z \quad \text{where} \quad z \in \mathbb{C}^p$$

so that

$$U_1^H A x = (U_1 A V_1)\, z = U_1^H b.$$

Note that

$$U_1 A V_1 : \mathbb{C}^p \to \mathbb{C}^p.$$

In fact,

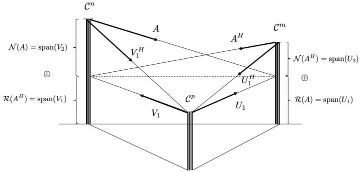
$$U_1^H A V_1 = U_1^H U_1 \Sigma_1 V_1^H V_1 = \Sigma_1.$$

so we have

$$\Sigma_1 z = U_1^H b$$

$$\Longrightarrow z = \Sigma_1^{-1} U_1^H b$$

$$\Longrightarrow \boxed{\hat{x} = V_1 \Sigma_1^{-1} U_1^H b}$$



Section 3

SVD and Numerically Sensitive Problems

Numerically Sensitive Problems

Suppose that we would like to solve

$$Ax = b$$

where $A \in \mathbb{R}^{n \times n}$ and $rank(A) = n$ but the condition number $\mathcal{K}(A)$ is large. Let $A = U \Sigma V^H$, then

$$\begin{aligned} A^{-1} &= V \Sigma^{-1} U^H \\ &= \sum_{j=1}^{n} \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} \end{aligned}$$

so the solution to $Ax = b$ is

$$x = A^{-1} b = \sum_{j=1}^{n} \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j}.$$

Numerically Sensitive Problems

Recall that $\mathcal{K}(A) = \|A\| \left\|A^{-1}\right\|$ where $\|A\| = \sigma_{\max}(A)$ and $\left\|A^{-1}\right\| = \dfrac{1}{\min_{\|x\|} \|A\|} = \dfrac{1}{\sigma_{\min}(A)}$. Therefore

$$\mathcal{K}(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Therefore a large $\mathcal{K}(A)$ implies there is significant difference between the largest and smallest singular values.

Numerically Sensitive Problems

For example $\sigma_{\min}(A)$ may be very small, therefore given

$$x = \sum_{j=1}^{n} \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} b$$

x is very sensitive to small change in b due to the terms in the sum that have very small singular values.

Solution: Zero out small singular values to get the approximate solution

$$Ax = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} x \approx \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} x$$

so

$$x = V_1 \Sigma_1^{-1} U_1^H b$$

is an approximate solution that is numerically stable.

Numerically Sensitive Problems

- Moon Example 7.4.1 shows that if σ_j-small then the vector $\mathbf{u}_j \in \mathbb{R}^m$ defines a sensitive direction for b. i.e. if b is almost parallel with $\mathbf{u}_j$ then $x = \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} b$ is clearly sensitive to small changes in b. If b is perpendicular to $\mathbf{u}_j$ then $\mathbf{u}_j^H b = 0$ and we are ok.
- If A comes from noisy data (almost always) then A will usually be full rank, even if the original data that produced A would have resulted in a lower rank A if it wasn't corrupted by noise.
- But the nonzero singular values added by noise will usually be small.
- Therefore, an effective way to reduce the rank of A to get rid of the effect of noise is to zero the "small" singular values.

Section 4

Rank Reducing Approximations

Rank Reducing Approximations

Problem: Given A with $rank(A) = r$, find a matrix B that is "close" to A in some sense, but with lower rank.

Theorem (Moon Theorem 7.2)

Given $A \in \mathbb{C}^{m \times n}$ with $rank(A) = r$, then

$$A = U_1 \Sigma_1 V_1^H = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^H$$

Let $k < r$ and let

$$A_k \triangleq \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^H \qquad (rank(A_k) = k)$$

Then $\|A - A_k\|_2 = \sigma_{k+1}$ and A_k is the nearest rank k matrix to A, in the matrix 2-norm, i.e.

$$A_k = \arg \min_{rank(B)=k} \|A - B\|_2 .$$

Rank Reducing Approximations, Proof

Remark: In the previous section, we saw that we could reduce the rank by zeroing small singular values. This theorem shows that this is the best way to reduce the rank in the matrix 2-norm sense.

Proof.

$$\begin{aligned} \|A - A_k\|_2 &= \left\| \sum_{j=k+1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^H \right\|_2 \\ &= \max_{\|x\|=1} \left\| \sum_{j=k+1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^H x \right\|_2 \end{aligned}$$

Note that we maximize by letting $x^* = \mathbf{v}_{k+1}$ since any other x will be a linear combination of smaller singular values.

Rank Reducing Approximations, Proof

Therefore

$$\|A - A_k\| = \|\sigma_{k+1}\mathbf{u}_{k+1}\| = \sigma_{k+1}$$

since $\|\mathbf{u}_{k+1}\| = 1$.

Because $\|A - A_k\|_2 = \sigma_{k+1}$ we know that

$$\min_{rank(B)=k} \|A - B\| \leq \sigma_{k+1}.$$

To complete the proof we need to show that

$$\sigma_{k+1} \leq \min_{rank(B)=k} \|A - B\|.$$

Rank Reducing Approximations, Proof

Let B be any rank-k matrix. Then

$$rank(B) = k \quad \Longrightarrow \quad dim(\mathcal{N}(B)) = n - k.$$

Therefore, there exists $\{x_{k+1}, \ldots, x_n\}$ such that

$$\mathcal{N}(B) = span\{x_{k+1}, \ldots, x_n\}$$

The columns of V_1 are $\{\mathbf{v}_1 \ldots \mathbf{v}_k, \mathbf{v}_{k+1} \ldots \mathbf{v}_r\}$ where $\mathbf{v}_i \in \mathbb{C}^n$. Let

$$z \in \underbrace{span\underbrace{\{x_{k+1}, \ldots, x_n\}}_{dim=n-k} \quad \cap \quad span\underbrace{\{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}}_{dim=k+1}}_{\text{dimension at least one since there are } n+1 \text{ vectors}}$$

Therefore $z \neq 0$.

Rank Reducing Approximations, Proof

Let

$$\|A-B\|_2 = \max_{\|x\|\neq 0} \frac{\|(A-B)x\|}{\|x\|} \leq \frac{\|(A-B)z\|}{\|z\|}$$
$$= \frac{\|Ax\|}{\|z\|} \text{ since } z \in \mathcal{N}(B)$$
$$= \frac{\left\|\sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^H z\right\|}{\|z\|} = \frac{\left\|\sum_{j=1}^{k+1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H z\right\|}{\|z\|}$$

Since $z \perp span\{\mathbf{v}_{k+2}, \ldots, \mathbf{v}_r\}$ the smallest we can make the numerator is σ_{k+1} by a choice of $z = \mathbf{v}_{k+1}$. So

$$\|A-B\|_2 \geq \frac{\|\sigma_{k+1}\mathbf{v}_{k+1}\|}{\|\mathbf{v}_{k+1}\|} = \sigma_{k+1}$$

for any B such that $rank(B) = k$ so that

$$\min_{rank(B)=k} \|A-B\|_2 \geq \sigma_{k+1}.$$

Section 5

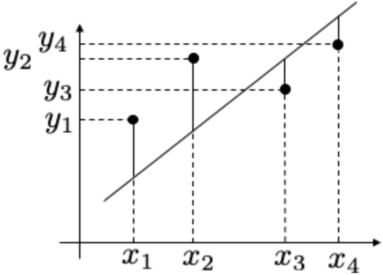
Applications

Application: Total least squares

If we are trying to fit a line to

$$y_i = ax_i$$

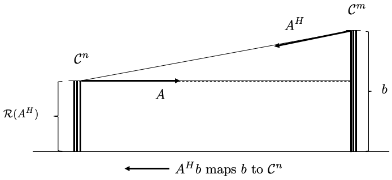
where (y_i, x_i) are measured. The least squares solution minimizes $e_i = y_i - ax_i$. Therefore $y_i - e_i = ax_i$.



In other words: fix the x_i's and play with a to minimize the error.

Application: Total least squares

For the general problem $\min \|Ax - b\|$ we assume A is perfect and that the imperfection is completely in b



Recall $A^H A x = A^H b$. When we premultiply by A^H we zero everything in b that was in the null space of A^H (i.e. we get rid of the bad parts of b).

Application: Total least squares

However A often comes from noisy data as well (like when fitting a line to data) e.g. if $\mathbf{u}_i = ax_i + b$, then

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\text{noisy}} = \begin{pmatrix} \underbrace{\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}}_{\text{noisy}} & \underbrace{\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}}_{\text{perfect}} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Application: Total least squares

Another interpretation of least squares is to find the smallest perturbation of b, i.e., δb such that

$$Ax = b + \delta b$$

where $b + \delta b \in \mathcal{R}(A)$.

The total lest squares problem is to find the smallest perturbation of b and A, denoted δb, δA such that

$$(A + \delta A)x = (b + \delta b)$$

supposing that $\begin{pmatrix} A & b \end{pmatrix}$ is full rank.

Application: Total least squares

This can be written as

$$\begin{pmatrix} A & b \end{pmatrix} \begin{pmatrix} x \\ -1 \end{pmatrix} + \begin{pmatrix} \delta A & \delta b \end{pmatrix} \begin{pmatrix} x \\ -1 \end{pmatrix} = 0$$

or

$$\left[\begin{pmatrix} A & b \end{pmatrix} + \begin{pmatrix} \delta A & \delta b \end{pmatrix} \right] \begin{pmatrix} x \\ -1 \end{pmatrix} = 0.$$

Define

$$C \stackrel{\triangle}{=} \begin{pmatrix} A & b \end{pmatrix} \text{ and } \Delta = \begin{pmatrix} \delta A & \delta b \end{pmatrix}$$

then

$$(C + \Delta) \begin{pmatrix} x \\ -1 \end{pmatrix} = 0.$$

Application: Total least squares

So $\begin{pmatrix} x \\ -1 \end{pmatrix} \in \mathcal{N}(C + \Delta)$ which implies that $C + \Delta$ is not full rank.

The problem is then to find the smallest perturbation Δ such that $C + \Delta$ looses rank.

Note that since $C = \begin{pmatrix} A & b \end{pmatrix} \in \mathbb{C}^{m \times (n+1)}$, for C to be full rank, we must have that $m > n$. Therefore we can write

$$C = \sum_{j=1}^{n+1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H.$$

Application: Total least squares

Hence, the smallest $\varDelta$ that reduces the rank of C is

$$\varDelta = -\sigma_{n+1}\mathbf{u}_{n+1}\mathbf{v}_{n+1}^H.$$

Note that $\mathbf{v}_{n+1} \in \mathcal{N}(C + \varDelta)$ since

$$(C + \varDelta)\mathbf{v}_{n+1} = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^H \mathbf{v}_{n+1} = 0$$

since $\mathbf{v}_i\mathbf{v}_j = \delta_{ij}$.

Application: Total least squares

Therefore

$$\begin{pmatrix} x \\ -1 \end{pmatrix} = \alpha \mathbf{v}_{n+1} = \alpha \begin{pmatrix} \mathbf{v}_{n+1}(n:1) \\ \mathbf{v}_{n+1}(n+1) \end{pmatrix}$$

Letting $\alpha = -\frac{1}{\mathbf{v}_{n+1}(n+1)}$ gives

$$x = \alpha \mathbf{v}_{n+1}(n:1)$$

This is valid if $\mathbf{v}_{n+1}(n+1) \neq 0$. Note that if σ_{n+1} is not a unique minimum singular value, i.e. $\sigma_{n+1} = \sigma_n = \cdots = \sigma_{k+1}$ then we want to find the smallest norm x such that

$$\begin{pmatrix} x \\ -1 \end{pmatrix} \in span\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{n+1}\}$$

Application: Homography Matrix

Application: MIMO Communication

Consider the MIMO communication system modeled by

$$\underbrace{Y(j\omega)}_{p\times 1} = \underbrace{H(j\omega)}_{1\times m}\underbrace{X(j\omega)}_{m\times 1}$$

What is the maximum gain of the system?

$$\|Y(j\omega)\| = \|H(j\omega)X(j\omega)\| \leq \|H(j\omega)\|\,\|X(j\omega)\|$$

Therefore, the maximum gain is given by

$$\begin{aligned}\gamma_{\max}(j\omega) &= \max_{X(j\omega)\neq 0} \frac{\|H(j\omega)X(j\omega)\|}{\|X(j\omega)\|} \\ &= \|H(j\omega)\| \\ &= \bar{\sigma}(H(j\omega)),\end{aligned}$$

where $\bar{\sigma}(H(j\omega))$ is the maximum singular value of $H(j\omega)$.

Application: MIMO Communication

How do you achieve this gain? Since

$$H(j\omega) = \Sigma \sigma_k(j\omega)\mathbf{u}_k(j\omega)\mathbf{v}_k^H(j\omega),$$

letting

$$X(j\omega) = \mathbf{v}_1(j\omega)$$

maximizes the gain in the system over the set $\|X(j\omega)\| = 1$.