SVD, Moore-Penrose Pseudo-inverse, and LSP

Rafikul Alam
Department of Mathematics
Indian Institute of Technology Guwahati
Guwahati - 781039, INDIA

Outline

- Moore-Penrose pseudo-inverse
- SVD based method for LSP

Moore-Penrose pseudo-inverse

Definition: Let $D \in \mathbb{C}^{m\times n}$ be diagonal. Then the Moore-Penrose pseudoinverse of D denoted by D^+ is defined to be the matrix obtained by transposing D and reciprocating each nonzero diagonal entries of D.

Moore-Penrose pseudo-inverse

Definition: Let $D \in \mathbb{C}^{m\times n}$ be diagonal. Then the Moore-Penrose pseudoinverse of D denoted by D^+ is defined to be the matrix obtained by transposing D and reciprocating each nonzero diagonal entries of D.

Example: Let $D := \left[\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & 0 \\ & & d_r & \\ \hline & 0 & & 0 \end{array}\right] \in \mathbb{C}^{m\times n}$ with $d_j \neq 0$ for $j = 1 : r$. Then

Moore-Penrose pseudo-inverse

Definition: Let $D \in \mathbb{C}^{m\times n}$ be diagonal. Then the Moore-Penrose pseudoinverse of D denoted by D^+ is defined to be the matrix obtained by transposing D and reciprocating each nonzero diagonal entries of D.

Example: Let $D := \left[\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & 0 \\ & & d_r & \\ \hline & 0 & & 0 \end{array}\right] \in \mathbb{C}^{m\times n}$ with $d_j \neq 0$ for $j = 1 : r$. Then

$$D^+ = \left[\begin{array}{ccc|c} 1/d_1 & & & \\ & \ddots & & 0 \\ & & 1/d_r & \\ \hline & 0 & & 0 \end{array}\right]^\top \in \mathbb{C}^{n\times m}.$$

Moore-Penrose pseudo-inverse

Definition: Let $D \in \mathbb{C}^{m\times n}$ be diagonal. Then the Moore-Penrose pseudoinverse of D denoted by D^+ is defined to be the matrix obtained by transposing D and reciprocating each nonzero diagonal entries of D.

Example: Let $D := \left[\begin{array}{ccc|c} d_1 & & & \\ & \ddots & & 0 \\ & & d_r & \\ \hline & 0 & & 0 \end{array}\right] \in \mathbb{C}^{m\times n}$ with $d_j \neq 0$ for $j = 1 : r$. Then

$$D^+ = \left[\begin{array}{ccc|c} 1/d_1 & & & \\ & \ddots & & 0 \\ & & 1/d_r & \\ \hline & 0 & & 0 \end{array}\right]^\top \in \mathbb{C}^{n\times m}.$$

This is a dream come true - invert what you can and ignore the rest!!!

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*, \; \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \; U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \; V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A.

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*,\ \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},\ U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix},\ V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then

$$A^+ := V\Sigma^+ U^* = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}^\top U^* = V_r \Sigma_r^{-1} U_r^*$$

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*,\ \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},\ U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix},\ V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then

$$A^+ := V\Sigma^+ U^* = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}^\top U^* = V_r \Sigma_r^{-1} U_r^*$$

is called the Moore-Penrose pseudoinverse of A.

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*,\ \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},\ \ U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix},\ \ V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then

$$A^+ := V\Sigma^+ U^* = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}^\top U^* = V_r \Sigma_r^{-1} U_r^*$$

is called the Moore-Penrose pseudoinverse of A.

- Note that $A^+ = A^{-1}$ when A is nonsingular.

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*,\ \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},\ \ U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix},\ \ V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then

$$A^+ := V\Sigma^+ U^* = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}^\top U^* = V_r \Sigma_r^{-1} U_r^*$$

is called the Moore-Penrose pseudoinverse of A.

- Note that $A^+ = A^{-1}$ when A is nonsingular.
- The MATLAB command `pinv(A)` computes the pseudo-inverse A^+ of an $m \times n$ matrix A.

Moore-Penrose pseudoinverse

Definition: Let $A \in \mathbb{C}^{m\times n}$. Suppose that $\mathrm{rank}(A) = r$. Let

$$A = U\Sigma V^*,\ \Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix},\ U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix},\ V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then

$$A^+ := V\Sigma^+ U^* = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}^\top U^* = V_r \Sigma_r^{-1} U_r^*$$

is called the Moore-Penrose pseudoinverse of A.

- Note that $A^+ = A^{-1}$ when A is nonsingular.
- The MATLAB command `pinv(A)` computes the pseudo-inverse A^+ of an $m \times n$ matrix A.
- If $U_r := \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix},\ V_r := \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}$ and $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$ then

$$A = U_r\Sigma_r V_r^* = \sigma_1 u_1 v_1^* + \cdots + \sigma_r u_r v_r^* \text{ and } A^+ = V_r \Sigma_r^{-1} U_r^* = \frac{v_1 u_1^*}{\sigma_1} + \cdots + \frac{v_r u_r^*}{\sigma_r}.$$

Moore-Penrose pseudoinverse

The Moore-Penrose pseudoinverse A^+ defines a bijective mapping from $R(A^*)$ to $R(A)$ and annihilates the vectors in $N(A^*)$. Indeed, consider the SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$. Let U, V, Σ_r be given by $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}, V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$.

Moore-Penrose pseudoinverse

The Moore-Penrose pseudoinverse A^+ defines a bijective mapping from $R(A^*)$ to $R(A)$ and annihilates the vectors in $N(A^*)$. Indeed, consider the SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$. Let U, V, Σ_r be given by $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}, V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$.

Then $AV = U\Sigma$ and $A^+U = V\Sigma^+$ show that

$$\begin{array}{rcl} A : v_1 & \longmapsto & \sigma_1 u_1 \\ & \vdots & \\ A : v_r & \longmapsto & \sigma_r u_r \\ A : v_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A : v_n & \longmapsto & 0 \end{array} \quad \text{and} \quad \begin{array}{rcl} A^+ : u_1 & \longmapsto & v_1/\sigma_1 \\ & \vdots & \\ A^+ : u_r & \longmapsto & v_r/\sigma_r \\ A^+ : u_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A^+ : u_m & \longmapsto & 0 \end{array}$$

Moore-Penrose pseudoinverse

The Moore-Penrose pseudoinverse A^+ defines a bijective mapping from $R(A^*)$ to $R(A)$ and annihilates the vectors in $N(A^*)$. Indeed, consider the SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$. Let U, V, Σ_r be given by $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}, V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$.

Then $AV = U\Sigma$ and $A^+U = V\Sigma^+$ show that

$$\begin{array}{rcl} A : v_1 & \longmapsto & \sigma_1 u_1 \\ & \vdots & \\ A : v_r & \longmapsto & \sigma_r u_r \\ A : v_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A : v_n & \longmapsto & 0 \end{array} \quad \text{and} \quad \begin{array}{rcl} A^+ : u_1 & \longmapsto & v_1/\sigma_1 \\ & \vdots & \\ A^+ : u_r & \longmapsto & v_r/\sigma_r \\ A^+ : u_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A^+ : u_m & \longmapsto & 0 \end{array}$$

Note that $A^+Ax = x$ for $x \in R(A^*)$ and $AA^+y = y$ for $y \in R(A)$.

Moore-Penrose pseudoinverse

? How

The Moore-Penrose pseudoinverse A^+ defines a bijective mapping from $R(A^*)$ to $R(A)$ and annihilates the vectors in $N(A^*)$. Indeed, consider the SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$. Let U, V, Σ_r be given by $U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix}, V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$.

Then $AV = U\Sigma$ and $A^+U = V\Sigma^+$ show that

$$
\begin{array}{lcl} A : v_1 & \longmapsto & \sigma_1 u_1 \\ & \vdots & \\ A : v_r & \longmapsto & \sigma_r u_r \\ A : v_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A : v_n & \longmapsto & 0 \end{array}
\quad \text{and} \quad
\begin{array}{lcl} A^+ : u_1 & \longmapsto & v_1/\sigma_1 \\ & \vdots & \\ A^+ : u_r & \longmapsto & v_r/\sigma_r \\ A^+ : u_{r+1} & \longmapsto & 0 \\ & \vdots & \\ A^+ : u_m & \longmapsto & 0 \end{array}
$$

Note that $A^+Ax = x$ for $x \in R(A^*)$ and $AA^+y = y$ for $y \in R(A)$. Conversely, A^+ can be defined as the bijective mapping from $R(A^*)$ to $R(A)$ that annihilates vectors in $N(A^*)$.

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

$$A^+ U_r = V_r \Sigma_r^{-1} = V \begin{bmatrix} \Sigma_r^{-1} \\ 0 \end{bmatrix} \text{ and } A^+ U_{m-r} = 0 = V \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which show that $A^+ U = A^+ \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*.$

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

$$A^+ U_r = V_r \Sigma_r^{-1} = V \begin{bmatrix} \Sigma_r^{-1} \\ 0 \end{bmatrix} \text{ and } A^+ U_{m-r} = 0 = V \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which show that $A^+ U = A^+ \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*.$

Examples:

$$\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}^+ = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix},$$

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

$$A^+ U_r = V_r \Sigma_r^{-1} = V \begin{bmatrix} \Sigma_r^{-1} \\ 0 \end{bmatrix} \text{ and } A^+ U_{m-r} = 0 = V \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which show that $A^+ U = A^+ \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*.$

Examples:

$$\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}^+ = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix},$$

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

$$A^+ U_r = V_r \Sigma_r^{-1} = V \begin{bmatrix} \Sigma_r^{-1} \\ 0 \end{bmatrix} \text{ and } A^+ U_{m-r} = 0 = V \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which show that $A^+ U = A^+ \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*.$

Examples:

$$\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}^+ = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} -3 & 0 \end{bmatrix}^+ = \begin{bmatrix} -\frac{1}{3} \\ 0 \end{bmatrix},$$

Moore-Penrose pseudoinverse

Let $U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ and $V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ be column partition of U and V. Then the mapping $A^+ u_j = v_j/\sigma_j$ for $j = 1 : r$ and $A^+ u_j = 0$ for $j = r+1 : m$ can be rewritten as

$$A^+ U_r = V_r \Sigma_r^{-1} = V \begin{bmatrix} \Sigma_r^{-1} \\ 0 \end{bmatrix} \text{ and } A^+ U_{m-r} = 0 = V \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which show that $A^+ U = A^+ \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Longrightarrow A^+ = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*.$

Examples:

$$\begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}^+ = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -3 & 0 \end{bmatrix}^+ = \begin{bmatrix} -\frac{1}{3} \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \end{bmatrix}.$$

Properties of pseudoinverse

Let $P \in \mathbb{C}^{n\times n}$. Then P is called a projection if $P^2 = P$. If P is a projection then $\mathbb{C}^n = R(P) \oplus N(P)$. Indeed $x = Px + (I - P)x$.

Properties of pseudoinverse

Let $P \in \mathbb{C}^{n\times n}$. Then P is called a projection if $P^2 = P$. If P is a projection then $\mathbb{C}^n = R(P) \oplus N(P)$. Indeed $x = Px + (I - P)x$.

A projection P is said to be an orthogonal projection if $R(P) \perp N(P)$. If P is an orthogonal projection then

$$\mathbb{C}^n = R(P) \oplus N(P) \text{ and } R(P) \perp N(P).$$

Properties of pseudoinverse

Let $P \in \mathbb{C}^{n\times n}$. Then P is called a projection if $P^2 = P$. If P is a projection then $\mathbb{C}^n = R(P) \oplus N(P)$. Indeed $x = Px + (I - P)x$.

A projection P is said to be an orthogonal projection if $R(P) \perp N(P)$. If P is an orthogonal projection then

$$\mathbb{C}^n = R(P) \oplus N(P) \text{ and } \; R(P) \perp N(P).$$

Fact: Let $A \in \mathbb{C}^{m\times n}$ and A^+ be the pseudoinverse of A. Then the following hold:

- $AA^+A = A$ and $A^+AA^+ = A^+$.
- $P := AA^+$ is the orthogonal projection on $R(A)$.
- $Q := A^+A$ is the orthonoal projection on $R(A^*)$.

Properties of pseudoinverse

Let $P \in \mathbb{C}^{n\times n}$. Then P is called a projection if $P^2 = P$. If P is a projection then $\mathbb{C}^n = R(P) \oplus N(P)$. Indeed $x = Px + (I - P)x$.

A projection P is said to be an orthogonal projection if $R(P) \perp N(P)$. If P is an orthogonal projection then

$$\mathbb{C}^n = R(P) \oplus N(P) \text{ and } \; R(P) \perp N(P).$$

Fact: Let $A \in \mathbb{C}^{m\times n}$ and A^+ be the pseudoinverse of A. Then the following hold:

- $AA^+A = A$ and $A^+AA^+ = A^+$.
- $P := AA^+$ is the orthogonal projection on $R(A)$.
- $Q := A^+A$ is the orthonoal projection on $R(A^*)$.
- If A has full column rank then $A^+ = (A^*A)^{-1}A^*$.
- If A has full row rank then $A^+ = A^*(AA^*)^{-1}$.

Properties of pseudoinverse

Let $P \in \mathbb{C}^{n\times n}$. Then P is called a projection if $P^2 = P$. If P is a projection then $\mathbb{C}^n = R(P) \oplus N(P)$. Indeed $x = Px + (I-P)x$.

A projection P is said to be an orthogonal projection if $R(P) \perp N(P)$. If P is an orthogonal projection then

$$\mathbb{C}^n = R(P) \oplus N(P) \text{ and } R(P) \perp N(P).$$

Fact: Let $A \in \mathbb{C}^{m\times n}$ and A^+ be the pseudoinverse of A. Then the following hold:

- $AA^+A = A$ and $A^+AA^+ = A^+$.
- $P := AA^+$ is the orthogonal projection on $R(A)$.
- $Q := A^+A$ is the orthonoal projection on $R(A^*)$.
- If A has full column rank then $A^+ = (A^*A)^{-1}A^*$.
- If A has full row rank then $A^+ = A^*(AA^*)^{-1}$.

Proof: Use the SVD $A = U\Sigma V^*$ and the fact that $A^+ = V\Sigma^+U^*$. ■

SVD and solution of LSP

Theorem: Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Suppose $\mathrm{rank}(A) = r$. Let

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*, \quad U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \quad V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then solutions of the LSP $Ax \approx b$ are given by

$$x = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y \quad \text{for any} \quad y \in \mathbb{C}^{n-r}.$$

SVD and solution of LSP

Theorem: Let $A \in \mathbb{C}^{m \times n}$ and $b \in \mathbb{C}^m$. Suppose $\mathrm{rank}(A) = r$. Let

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*, \quad U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \quad V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then solutions of the LSP $Ax \approx b$ are given by

$$x = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y \quad \text{for any} \quad y \in \mathbb{C}^{n-r}.$$

Proof: Set $\begin{bmatrix} z \\ y \end{bmatrix} := V^* x$ with $z \in \mathbb{C}^r$ and $\begin{bmatrix} c \\ d \end{bmatrix} := U^* b$ with $c \in \mathbb{C}^r$. Then

SVD and solution of LSP

Theorem: Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Suppose $\mathrm{rank}(A) = r$. Let

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*, \quad U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \quad V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then solutions of the LSP $Ax \approx b$ are given by

$$x = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y \ \text{ for any } \ y \in \mathbb{C}^{n-r}.$$

Proof: Set $\begin{bmatrix} z \\ y \end{bmatrix} := V^* x$ with $z \in \mathbb{C}^r$ and $\begin{bmatrix} c \\ d \end{bmatrix} := U^* b$ with $c \in \mathbb{C}^r$. Then

$$\|Ax - b\|_2 \quad = \quad \|U^*(Ax-b)\|_2 = \left\| \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^* x - U^* b \right\|_2$$

SVD and solution of LSP

Theorem: Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. Suppose $\mathrm{rank}(A) = r$. Let

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*, \quad U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}, \quad V = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$$

be an SVD of A. Then solutions of the LSP $Ax \approx b$ are given by

$$x = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y \quad \text{for any} \quad y \in \mathbb{C}^{n-r}.$$

Proof: Set $\begin{bmatrix} z \\ y \end{bmatrix} := V^* x$ with $z \in \mathbb{C}^r$ and $\begin{bmatrix} c \\ d \end{bmatrix} := U^* b$ with $c \in \mathbb{C}^r$. Then

$$\begin{aligned} \|Ax - b\|_2 &= \|U^*(Ax - b)\|_2 = \left\| \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^* x - U^* b \right\|_2 \\ &= \left\| \begin{bmatrix} \Sigma_r z - c \\ -d \end{bmatrix} \right\|_2 = \sqrt{\|\Sigma_r z - c\|_2^2 + \|d\|_2^2} = \|d\|_2 \\ &\Leftrightarrow \quad z = \Sigma_r^{-1} c. \text{ Hence } x = V \begin{bmatrix} \Sigma_r^{-1} c \\ y \end{bmatrix} \text{ is a solution for any } y. \end{aligned}$$

Proof

Thus $x = V \begin{bmatrix} \Sigma_r^{-1} c \\ y \end{bmatrix} = V_r \Sigma_r^{-1} c + V_{n-r} y = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y$ is a solution of the LSP $Ax \approx b$ for any $y \in \mathbb{C}^{n-r}$. $\blacksquare$

Proof

Thus $x = V \begin{bmatrix} \Sigma_r^{-1} c \\ y \end{bmatrix} = V_r \Sigma_r^{-1} c + V_{n-r} y = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y$ is a solution of the LSP $Ax \approx b$ for any $y \in \mathbb{C}^{n-r}$. ■

Remark: We have $x = A^+ b + V_{n-r} y$ and $A^+ b \perp V_{n-r} y$. Hence $\|x\|_2 = \sqrt{\|A^+ b\|_2^2 + \|y\|_2^2}$
$\Longrightarrow x = A^+ b$ is a unique solution of the LSP $Ax \approx b$ having the smallest norm.

Proof

Thus $x = V \begin{bmatrix} \Sigma_r^{-1} c \\ y \end{bmatrix} = V_r \Sigma_r^{-1} c + V_{n-r} y = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y$ is a solution of the LSP $Ax \approx b$ for any $y \in \mathbb{C}^{n-r}$. $\blacksquare$

Remark: We have $x = A^+ b + V_{n-r} y$ and $A^+ b \perp V_{n-r} y$. Hence $\|x\|_2 = \sqrt{\|A^+ b\|_2^2 + \|y\|_2^2}$ $\implies x = A^+ b$ is a unique solution of the LSP $Ax \approx b$ having the smallest norm.

Algorithm: Solution of LSP $Ax \approx b$ when $\mathrm{rank}(A) = r$.

1. Compute SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$
2. Set $\begin{bmatrix} c \\ d \end{bmatrix} := U^* b$, where $c \in \mathbb{C}^r$ and $d \in \mathbb{C}^{m-r}$.
3. Set $x = \mathtt{V(:,1:r)} \Sigma_r^{-1} c$. Then x is a unique least norm solution of the LSP.
4. Compute the residual $\|d\|_2$.

Proof

Thus $x = V \begin{bmatrix} \Sigma_r^{-1} c \\ y \end{bmatrix} = V_r \Sigma_r^{-1} c + V_{n-r} y = V_r \Sigma_r^{-1} U_r^* b + V_{n-r} y = A^+ b + V_{n-r} y$ is a solution of the LSP $Ax \approx b$ for any $y \in \mathbb{C}^{n-r}$. ■

As range(A+) = span(v1, v2, ..., vr) & V(n-r)*y = L.C of v(r+1), v(r+2)....., v(n). Since V is unitary, vi perpendicular to vj, for i != j

Remark: We have $x = A^+ b + V_{n-r} y$ and $\underline{A^+ b \perp V_{n-r} y}$. Hence $\|x\|_2 = \sqrt{\|A^+ b\|_2^2 + \|y\|_2^2}$
$\implies x = A^+ b$ is a unique solution of the LSP $Ax \approx b$ having the smallest norm.

Algorithm: Solution of LSP $Ax \approx b$ when $\mathrm{rank}(A) = r$.

1. Compute SVD $A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^*$
2. Set $\begin{bmatrix} c \\ d \end{bmatrix} := U^* b$, where $c \in \mathbb{C}^r$ and $d \in \mathbb{C}^{m-r}$.
3. Set $x = \mathtt{V}(:, \mathtt{1} : \mathtt{r}) \Sigma_r^{-1} c$. Then x is a unique least norm solution of the LSP.
4. Compute the residual $\|d\|_2$.

Remark: Now $x = A^+ b = V_r \Sigma_r^{-1} U_r^* b = \dfrac{v_1 u_1^* b}{\sigma_1} + \cdots + \dfrac{v_r u_r^* b}{\sigma_r}$ shows the effect of small singular values on the solution of the LSP $Ax \approx b$..

Examples

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}^{+} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}^{\top} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Examples

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}^+ = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}^\top = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 2 \end{bmatrix}^+ = \begin{bmatrix} 1 & 2 \end{bmatrix}^\top \frac{1}{5} = \begin{bmatrix} \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Examples

$$\begin{bmatrix}1\\1\end{bmatrix}^{+} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}^{\top} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix} \text{ and } \begin{bmatrix}1 & 2\end{bmatrix}^{+} = \begin{bmatrix}1 & 2\end{bmatrix}^{\top}\frac{1}{5} = \begin{bmatrix}\frac{1}{5}\\ \frac{2}{5}\end{bmatrix}.$$

Let's apply our results to simple LSPs . First consider an overdetermined system:

$$\text{LSP} \quad \begin{bmatrix}1\\1\end{bmatrix} x = \begin{bmatrix}0\\1\end{bmatrix}$$

Examples

$$\begin{bmatrix}1\\1\end{bmatrix}^{+} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}^{\top} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix} \text{ and } \begin{bmatrix}1 & 2\end{bmatrix}^{+} = \begin{bmatrix}1 & 2\end{bmatrix}^{\top}\frac{1}{5} = \begin{bmatrix}\frac{1}{5}\\ \frac{2}{5}\end{bmatrix}.$$

Let's apply our results to simple LSPs . First consider an overdetermined system:

$$\mathrm{LSP}\ \begin{bmatrix}1\\1\end{bmatrix}x = \begin{bmatrix}0\\1\end{bmatrix} \Rightarrow x = \begin{bmatrix}1\\1\end{bmatrix}^{+}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}.$$

Examples

$$\begin{bmatrix}1\\1\end{bmatrix}^{+} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}^{\top} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix} \text{ and } \begin{bmatrix}1 & 2\end{bmatrix}^{+} = \begin{bmatrix}1 & 2\end{bmatrix}^{\top}\frac{1}{5} = \begin{bmatrix}\frac{1}{5}\\ \frac{2}{5}\end{bmatrix}.$$

Let's apply our results to simple LSPs . First consider an overdetermined system:

$$\mathrm{LSP}\ \begin{bmatrix}1\\1\end{bmatrix} x = \begin{bmatrix}0\\1\end{bmatrix} \Rightarrow x = \begin{bmatrix}1\\1\end{bmatrix}^{+}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}.$$

Now consider an underdetermined system

$$\mathrm{LSP}\ \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = 5$$

Examples

Using the result, for full column matrix, A+ = (A*A)^(-1) A*

$$\begin{bmatrix}1\\1\end{bmatrix}^{+} = \frac{1}{2}\begin{bmatrix}1\\1\end{bmatrix}^{\top} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix} \text{ and } \begin{bmatrix}1 & 2\end{bmatrix}^{+} = \begin{bmatrix}1 & 2\end{bmatrix}^{\top}\frac{1}{5} = \begin{bmatrix}\frac{1}{5}\\ \frac{2}{5}\end{bmatrix}.$$

Let's apply our results to simple LSPs . First consider an overdetermined system:

$$\mathrm{LSP}\ \begin{bmatrix}1\\1\end{bmatrix}x = \begin{bmatrix}0\\1\end{bmatrix} \Rightarrow x = \begin{bmatrix}1\\1\end{bmatrix}^{+}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{2}.$$

Now consider an underdetermined system

$$\mathrm{LSP}\ \begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = 5 \Rightarrow \begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}1 & 2\end{bmatrix}^{+}5 = \begin{bmatrix}\frac{1}{5}\\ \frac{2}{5}\end{bmatrix}5 = \begin{bmatrix}1\\2\end{bmatrix}.$$

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Remark: Note that $\mathrm{cond}(A) = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and the smallest (nonzero) singular values of A.

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Remark: Note that $\mathrm{cond}(A) = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and the smallest (nonzero) singular values of A.

Perturbation: Now, consider the LSP $Ax \approx b$ and the LSP $A\hat{x} \approx b + \Delta b$.

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Remark: Note that $\mathrm{cond}(A) = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and the smallest (nonzero) singular values of A.

Perturbation: Now, consider the LSP $Ax \approx b$ and the LSP $A\hat{x} \approx b + \Delta b$. Then we have $x = A^+ b$ and $\hat{x} = A^+(b + \Delta b)$ which yield

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Remark: Note that $\mathrm{cond}(A) = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and the smallest (nonzero) singular values of A.

Perturbation: Now, consider the LSP $Ax \approx b$ and the LSP $A\hat{x} \approx b + \Delta b$. Then we have $x = A^+ b$ and $\hat{x} = A^+(b + \Delta b)$ which yield

$$\|x - \hat{x}\|_2 \leq \|A^+\|_2 \|\Delta b\|_2.$$

Perturbation of LSP

Definition: The condition number of an $m \times n$ matrix A is defined by $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$.

Remark: Note that $\mathrm{cond}(A) = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and the smallest (nonzero) singular values of A.

Perturbation: Now, consider the LSP $Ax \approx b$ and the LSP $A\hat{x} \approx b + \Delta b$. Then we have $x = A^+ b$ and $\hat{x} = A^+(b + \Delta b)$ which yield

$$\|x - \hat{x}\|_2 \leq \|A^+\|_2 \|\Delta b\|_2.$$

Since $\|Ax\|_2 = \|b\|_2 \cos\theta$, we have

$$\frac{\|x - \hat{x}\|_2}{\|x\|_2} \leq \frac{\mathrm{cond}(A)}{\cos\theta} \frac{\|\Delta b\|_2}{\|b\|_2}.$$

Perturbation of LSP

Let $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$. Suppose $\|\Delta A\|_2 / \|A\|_2 < 1/\mathrm{cond}(A)$ and $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A)$.

Perturbation of LSP

Let $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$. Suppose $\|\Delta A\|_2 / \|A\|_2 < 1/\mathrm{cond}(A)$ and $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A)$.

- The requirement $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A)$ is difficult to ensure for rank deficient LSP because the map $A \mapsto \mathrm{rank}(A)$ is not continuous.

Perturbation of LSP

Let $\mathrm{cond}(A) := \|A\|_2 \|A^+\|_2$. Suppose $\|\Delta A\|_2 / \|A\|_2 < 1/\mathrm{cond}(A)$ and $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A)$.

- The requirement $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A)$ is difficult to ensure for rank deficient LSP because the map $A \mapsto \mathrm{rank}(A)$ is not continuous.

Consider the LSP $Ax = b$ and the LSP $(A + \Delta A)\hat{x} = b$. Then

Perturbation of LSP

Let $\mathrm{cond}(A) := \|A\|_2\|A^+\|_2$. Suppose $\|\Delta A\|_2/\|A\|_2 < 1/\mathrm{cond}(A)$ and $\mathrm{rank}(A+\Delta A) = \mathrm{rank}(A)$.

- The requirement $\mathrm{rank}(A+\Delta A) = \mathrm{rank}(A)$ is difficult to ensure for rank deficient LSP because the map $A \mapsto \mathrm{rank}(A)$ is not continuous.

Consider the LSP $Ax = b$ and the LSP $(A+\Delta A)\hat{x} = b$. Then

$$\begin{aligned}\frac{\|x-\hat{x}\|_2}{\|x\|_2} &\lesssim \mathrm{cond}(A)^2 \frac{\|\Delta A^*\|_2}{\|A\|_2}\frac{\|r\|_2}{\|A\|_2\|x\|_2} + \mathrm{cond}(A)\frac{\|\Delta A\|_2}{\|A\|_2} \\ &\lesssim \left(\mathrm{cond}(A)^2\tan(\theta) + \mathrm{cond}(A)\right)\frac{\|\Delta A\|_2}{\|A\|_2},\end{aligned}$$

where $r := b - Ax$.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+U^* = \sum_{j=1}^{r} v_j u_j^*/\sigma_j$, where $v_j := V(:, j)$ and $u_j := U(:, j)$.

Pitfalls

Suppose that $\operatorname{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^*/\sigma_j$, where $v_j := V(:,j)$ and $u_j := U(:,j)$. Hence

$$\mathrm{LSP}\ Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^*/\sigma_j$, where $v_j := V(:,j)$ and $u_j := U(:,j)$. Hence

$$\mathrm{LSP}\; Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^*/\sigma_j$, where $v_j := V(:,j)$ and $u_j := U(:,j)$. Hence

$$\mathrm{LSP}\ Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing $A^+ b$ requires knowing the smallest (nonzero) singular value of A and hence the rank of A.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^*/\sigma_j$, where $v_j := V(:, j)$ and $u_j := U(:, j)$. Hence

$$\mathrm{LSP}\ Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing $A^+ b$ requires knowing the smallest (nonzero) singular value of A and hence the rank of A.
- Since $A \mapsto \mathrm{rank}(A)$ is discontinuous, computing $A^+ b$ in the presence of roundoff errors is a difficult task.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^* / \sigma_j$, where $v_j := V(:, j)$ and $u_j := U(:, j)$. Hence

$$\text{LSP } Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing $A^+ b$ requires knowing the smallest (nonzero) singular value of A and hence the rank of A.
- Since $A \mapsto \mathrm{rank}(A)$ is discontinuous, computing $A^+ b$ in the presence of roundoff errors is a difficult task.
- For example, $A = \mathrm{diag}(1, 0)$ is rank deficient.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^* / \sigma_j$, where $v_j := V(:, j)$ and $u_j := U(:, j)$. Hence

$$\mathrm{LSP}\ Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^r \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing $A^+ b$ requires knowing the smallest (nonzero) singular value of A and hence the rank of A.
- Since $A \mapsto \mathrm{rank}(A)$ is discontinuous, computing $A^+ b$ in the presence of roundoff errors is a difficult task.
- For example, $A = \mathrm{diag}(1, 0)$ is rank deficient. For $b = [1, 1]^\top$, LSP $Ax = b \Rightarrow x = [1, 0]^\top$.

Pitfalls

Suppose that $\operatorname{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^*/\sigma_j$, where $v_j := V(:,j)$ and $u_j := U(:,j)$. Hence

$$\text{LSP } Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b)v_j}{\sigma_j}.$$

- The unique solution A^+b of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing A^+b requires knowing the smallest (nonzero) singular value of A and hence the rank of A.
- Since $A \mapsto \operatorname{rank}(A)$ is discontinuous, computing A^+b in the presence of roundoff errors is a difficult task.
- For example, $A = \operatorname{diag}(1, 0)$ is rank deficient. For $b = [1, 1]^\top$, LSP $Ax = b \Rightarrow x = [1, 0]^\top$. For tiny ϵ, $A + \Delta A = \operatorname{diag}(1, \epsilon)$ is nonsingular.

Pitfalls

Suppose that $\mathrm{rank}(A) = r$. Then $A^+ = V\Sigma^+ U^* = \sum_{j=1}^r v_j u_j^* / \sigma_j$, where $v_j := V(:,j)$ and $u_j := U(:,j)$. Hence

$$\text{LSP } Ax = b \Rightarrow x = A^+ b = \sum_{j=1}^{r} \frac{(u_j^* b) v_j}{\sigma_j}.$$

- The unique solution $A^+ b$ of the LSP $Ax = b$ is strongly dependent on the smallest singular value of A.
- Thus computing $A^+ b$ requires knowing the smallest (nonzero) singular value of A and hence the rank of A.
- Since $A \mapsto \mathrm{rank}(A)$ is discontinuous, computing $A^+ b$ in the presence of roundoff errors is a difficult task.
- For example, $A = \mathrm{diag}(1, 0)$ is rank deficient. For $b = [1, 1]^\top$, LSP $Ax = b \Rightarrow x = [1, 0]^\top$. For tiny ϵ, $A + \Delta A = \mathrm{diag}(1, \epsilon)$ is nonsingular. Now LSP $(A + \Delta A)\hat{x} = b \Rightarrow \hat{x} = [1, 1/\epsilon]^\top$ shows that $\|x - \hat{x}\|_2 / \|x\|_2 = 1/\epsilon$.

Pitfalls

- Roundoff will make tiny perturbations of magnitude $\leq \mathbf{u}\, \|A\|_2$ and A. But a tiny error can have dramatic effect on the solution when A is rank deficient.

Pitfalls

- Roundoff will make tiny perturbations of magnitude $\leq \mathbf{u}\, \|A\|_2$ and A. But a tiny error can have dramatic effect on the solution when A is rank deficient.
- The SVD method so outlined, is the best bet in the face of roundoff error. It has its limitations but it works well in practice.

Pitfalls

- Roundoff will make tiny perturbations of magnitude $\leq \mathbf{u} \, \|A\|_2$ and A. But a tiny error can have dramatic effect on the solution when A is rank deficient.
- The SVD method so outlined, is the best bet in the face of roundoff error. It has its limitations but it works well in practice.

Conclusion: If implemented properly, the SVD based method produces a solution of the LSP $(A + \Delta A)\hat{x} = b + \Delta b$ with

$$\max\left(\frac{\|\Delta A\|_2}{\|A\|_2}, \frac{\|\Delta b\|_2}{\|b\|_2}\right) = \mathcal{O}(\mathbf{u}).$$

Pitfalls

- Roundoff will make tiny perturbations of magnitude $\leq \mathbf{u}\, \|A\|_2$ and A. But a tiny error can have dramatic effect on the solution when A is rank deficient.
- The SVD method so outlined, is the best bet in the face of roundoff error. It has its limitations but it works well in practice.

Conclusion: If implemented properly, the SVD based method produces a solution of the LSP $(A + \Delta A)\hat{x} = b + \Delta b$ with

$$\max\left(\frac{\|\Delta A\|_2}{\|A\|_2}, \frac{\|\Delta b\|_2}{\|b\|_2}\right) = \mathcal{O}(\mathbf{u}).$$

Combining this with the perturbation bounds we get error bounds for the (smallest norm) solution of the least squares problem.