Introduction to Numerical Linear Algebra II

Petros Drineas

These slides were prepared by Ilse Ipsen for the 2015 Gene Golub SIAM Summer School on RandNLA

Overview

We will cover this material in relatively more detail, but will still skip a lot ...

1. Norms {Measuring the length/mass of mathematical quantities}

   General norms
   Vector p norms
   Matrix norms induced by vector p norms
   Frobenius norm

2. Singular Value Decomposition (SVD)

   The most important tool in Numerical Linear Algebra

3. Least Squares problems

   Linear systems that do not have a solution

Norms

General Norms

How to measure the mass of a matrix or length of a vector

Norm $\|\cdot\|$ is function $\mathbb{R}^{m\times n} \rightarrow \mathbb{R}$ with

1. Non-negativity $\|A\| \geq 0, \quad \|A\| = 0 \iff A = 0$
2. Triangle inequality $\|A + B\| \leq \|A\| + \|B\|$
3. Scalar multiplication $\|\alpha A\| = |\alpha| \, \|A\|$ for all $\alpha \in \mathbb{R}$.

Properties

- Minus signs $\|-A\| = \|A\|$
- Reverse triangle inequality $|\, \|A\| - \|B\| \,| \leq \|A - B\|$
- New norms

  For norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$, and nonsingular $M \in \mathbb{R}^{m\times m}$

  $\|A\|_M \stackrel{\text{def}}{=} \|M\,A\|$ is also a norm

Vector p Norms

For $\mathrm{x} \in \mathbb{R}^n$ and integer $p \geq 1$

$$\|\mathrm{x}\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

- One norm $\quad \|\mathrm{x}\|_1 = \sum_{j=1}^{n} |x_j|$
- Euclidean (two) norm $\quad \|\mathrm{x}\|_2 = \sqrt{\sum_{j=1}^{n} |x_j|^2} = \sqrt{\mathrm{x}^T \mathrm{x}}$
- Infinity (max) norm $\quad \|\mathrm{x}\|_\infty = \max_{1 \leq j \leq n} |x_j|$

Exercises

1. Determine $\|\mathbb{1}\|_p$ for $\mathbb{1} \in \mathbb{R}^n$ and $p = 1, 2, \infty$

2. For $\mathrm{x} \in \mathbb{R}^n$ with $x_j = j$, $1 \leq j \leq n$, determine closed-form expressions for $\|\mathrm{x}\|_p$ for $p = 1, 2, \infty$

Inner Products and Norm Relations

For $x, y \in \mathbb{R}^n$

- Cauchy-Schwartz inequality

$$|x^T y| \le \|x\|_2 \|y\|_2$$

- Hölder inequality

$$|x^T y| \le \|x\|_1 \|y\|_\infty \qquad |x^T y| \le \|x\|_\infty \|y\|_1$$

- Relations

$$\begin{aligned} \|x\|_\infty \le \quad & \|x\|_1 \quad \le n \|x\|_\infty \\ \|x\|_2 \le \quad & \|x\|_1 \quad \le \sqrt{n} \|x\|_2 \\ \|x\|_\infty \le \quad & \|x\|_2 \quad \le \sqrt{n} \|x\|_\infty \end{aligned}$$

Exercises

1. Prove the norm relations on the previous slide
2. For $x \in \mathbb{R}^n$ show $\quad \|x\|_2 \leq \sqrt{\|x\|_\infty \|x\|_1}$

Vector Two Norm

- Theorem of Pythagoras

  For $x, y \in \mathbb{R}^n$

$$x^T y = 0 \iff \|x \pm y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

- Two norm does not care about orthonormal matrices

  For $x \in \mathbb{R}^n$ and $V \in \mathbb{R}^{m \times n}$ with $V^T V = I_n$

$$\|Vx\|_2 = \|x\|_2$$

Vector Two Norm

- Theorem of Pythagoras

  For $x, y \in \mathbb{R}^n$

$$x^T y = 0 \quad \Longleftrightarrow \quad \|x \pm y\|_2^2 = \|x\|_2^2 + \|y\|_2^2$$

- Two norm does not care about orthonormal matrices

  For $x \in \mathbb{R}^n$ and $V \in \mathbb{R}^{m \times n}$ with $V^T V = I_n$

$$\|V x\|_2 = \|x\|_2$$

$$\|V x\|_2^2 = (V x)^T (V x) = x^T V^T V x = x^T x = \|x\|_2^2$$

Exercises

For $x, y \in \mathbb{R}^n$ show

1. Parallelogram equality

$$\|x + y\|_2^2 + \|x - y\|_2^2 = 2\left(\|x\|_2^2 + \|y\|_2^2\right)$$

2. Polarization identity

$$x^T y = \frac{1}{4}\left(\|x + y\|_2^2 - \|x - y\|_2^2\right)$$

Matrix Norms (Induced by Vector p Norms)

For $A \in \mathbb{R}^{m \times n}$ and integer $p \geq 1$

$$\|A\|_p \stackrel{\text{def}}{=} \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|y\|_p = 1} \|A\,y\|_p$$

- One Norm: Maximum absolute column sum

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| = \max_{1 \leq j \leq n} \|A\,e_j\|_1$$

- Infinity Norm: Maximum absolute row sum

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| = \max_{1 \leq i \leq m} \|A^T\,e_i\|_1$$

Matrix Norm Properties

Every norm realized by some vector $y \neq 0$

$$\|A\|_p = \frac{\|Ay\|_p}{\|y\|_p} = \|Az\|_p \quad \text{where} \quad z \equiv \frac{y}{\|y\|_p} \quad \|z\|_p = 1$$

$\{$Vector y is different for every A and every $p\}$

Submultiplicativity

- Matrix vector product: For $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^n$

$$\|Ay\|_p \leq \|A\|_p \|y\|_p$$

- Matrix product: For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times l}$

$$\|AB\|_p \leq \|A\|_p \|B\|_p$$

More Matrix Norm Properties

For $A \in \mathbb{R}^{m \times n}$ and permutation matrices $P \in \mathbb{R}^{m \times m}$, $Q \in \mathbb{R}^{n \times n}$

- Permutation matrices do not matter

$$\|P\,A\,Q\|_p = \|A\|_p$$

- Submatrices have smaller norms than parent matrix

If $P\,A\,Q = \begin{pmatrix} B & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ then $\|B\|_p \leq \|A\|_p$

Exercises

1. Different norms realized by different vectors

   Find y and z so that $\|A\|_1 = \|Ay\|_1$ and $\|A\|_\infty = \|Az\|_\infty$ when

   $$A = \begin{pmatrix} 1 & 4 \\ 0 & 2 \end{pmatrix}$$

2. If $Q \in \mathbb{R}^{n \times n}$ is permutation then $\|Q\|_p = 1$

3. Prove the two types of submultiplicativity

4. If $D = \operatorname{diag}\begin{pmatrix} d_{11} & \cdots & d_{nn} \end{pmatrix}$ then

   $$\|D\|_p = \max_{1 \le j \le n} |d_{jj}|$$

5. If $A \in \mathbb{R}^{n \times n}$ nonsingular then $\|A\|_p \|A^{-1}\|_p \ge 1$

Norm Relations and Transposes

For $A \in \mathbb{R}^{m \times n}$

- Relations between different norms

$$\frac{1}{\sqrt{n}} \|A\|_\infty \leq \quad \|A\|_2 \quad \leq \sqrt{m}\, \|A\|_\infty$$
$$\frac{1}{\sqrt{m}} \|A\|_1 \leq \quad \|A\|_2 \quad \leq \sqrt{n}\, \|A\|_1$$

- Transposes

$$\|A^T\|_1 = \|A\|_\infty \qquad \|A^T\|_\infty = \|A\|_1 \qquad \|A^T\|_2 = \|A\|_2$$

Proof: Two Norm of Transpose

$\left\{\text{True for } A = 0, \text{ so assume } A \neq 0\right\}$

- Let $\|A\|_2 = \|Az\|_2$ for some $z \in \mathbb{R}^n$ with $\|z\|_2 = 1$
- Reduce to vector norm

$$\|A\|_2^2 = \|Az\|_2^2 = (Az)^T(Az) = z^TA^TAz = z^T \underbrace{\left(A^TAz\right)}_{\text{vector}}$$

- Cauchy-Schwartz inequality and submultiplicativity imply

$$z^T\left(A^TAz\right) \leq \|z\|_2 \, \|A^TAz\|_2 \leq \|A^T\|_2 \, \|A\|_2$$

- Thus $\|A\|_2^2 \leq \|A^T\|_2 \, \|A\|_2$ and $\|A\|_2 \leq \|A^T\|_2$
- Reversing roles of A and A^T gives $\|A^T\|_2 \leq \|A\|_2$

Matrix Two Norm

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

- Orthonormal matrices do not change anything

  If $U \in \mathbb{R}^{k \times m}$ with $U^T U = I_m$, $V \in \mathbb{R}^{l \times n}$ with $V^T V = I_n$

$$\|U\, A\, V^T\|_2 = \|A\|_2$$

- Gram matrices

  For $A \in \mathbb{R}^{m \times n}$

$$\|A^T A\|_2 \; = \; \|A\|_2^2 \; = \; \|A\, A^T\|_2$$

- Outer products

  For $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$

$$\|xy^T\|_2 = \|x\|_2 \, \|y\|_2$$

Proof: Norm of Gram Matrix

- Submultiplicativity and transpose

$$\|A^T A\|_2 \le \|A\|_2 \, \|A^T\|_2 = \|A\|_2 \, \|A\|_2 = \|A\|_2^2$$

Thus $\|A^T A\|_2 \le \|A\|_2^2$

- Let $\|A\|_2 = \|Az\|_2$ for some $z \in \mathbb{R}^n$ with $\|z\|_2 = 1$
- Cauchy-Schwartz inequality and submultiplicativity imply

$$\begin{aligned} \|A\|_2^2 = \|Az\|_2^2 &= (Az)^T (Az) = z^T \underbrace{\left(A^T A\, z\right)}_{\text{vector}} \\ &\le \|z\|_2 \, \|A^T A\, z\|_2 \le \|A^T A\|_2 \end{aligned}$$

Thus $\|A\|_2^2 \le \|A^T A\|_2$

Exercises

1. Infinity norm of outer products

   For $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, show $\|x\, y^T\|_\infty = \|x\|_\infty \, \|y\|_1$

2. If $A \in \mathbb{R}^{n \times n}$ with $A \neq 0$ is idempotent then

   (i) $\|A\|_p \geq 1$

   (ii) $\|A\|_2 = 1$ if A also symmetric

3. Given $A \in \mathbb{R}^{n \times n}$

   Among all symmetric matrices, $\frac{1}{2}(A + A^T)$ is a (the?) matrix that is closest to A in the two norm

Frobenius Norm

The true mass of a matrix

For $A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \in \mathbb{R}^{m \times n}$

$$\|A\|_F \stackrel{\text{def}}{=} \sqrt{\sum_{j=1}^{n} \|a_j\|_2^2} = \sqrt{\sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2} = \sqrt{\text{trace}(A^T A)}$$

Frobenius Norm

The true mass of a matrix

For $A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \in \mathbb{R}^{m \times n}$

$$\|A\|_F \stackrel{\text{def}}{=} \sqrt{\sum_{j=1}^{n} \|a_j\|_2^2} = \sqrt{\sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2} = \sqrt{\text{trace}(A^T A)}$$

- Vector If $x \in \mathbb{R}^n$ then $\|x\|_F = \|x\|_2$
- Transpose $\|A^T\|_F = \|A\|_F$
- Identity $\|I_n\|_F = \sqrt{n}$

More Frobenius Norm Properties

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

- Orthonormal invariance

  If $U \in \mathbb{R}^{k \times m}$ with $U^T U = I_m$, $V \in \mathbb{R}^{l \times n}$ with $V^T V = I_n$

$$\|U A V^T\|_F = \|A\|_F$$

- Relation to two norm

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{\mathrm{rank}(A)}\, \|A\|_2 \leq \sqrt{\min\{m, n\}}\, \|A\|_2$$

- Submultiplicativity

$$\|A B\|_F \leq \|A\|_2 \, \|B\|_F \leq \|A\|_F \, \|B\|_F$$

Exercises

1. Show the orthonormal invariance of the Frobenius norm

2. Show the submultiplicativity of the Frobenius norm

3. Frobenius norm of outer products

   For $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ show

$$\|x\, y^T\|_F = \|x\|_2 \, \|y\|_2$$

Singular Value Decomposition (SVD)

Full SVD

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

- Tall and skinny: $m \geq n$

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \qquad \text{where} \quad \Sigma \equiv \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \geq 0$$

- Short and fat: $m \leq n$

$$A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V^T \qquad \text{where} \quad \Sigma \equiv \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{pmatrix} \geq 0$$

$U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices

Names and Conventions

$A \in \mathbb{R}^{m\times n}$ with SVD

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \qquad \text{or} \qquad A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V^T$$

- Singular values (svalues): Diagonal elements σ_j of Σ
- Left singular vector matrix: U
- Right singular vector matrix: V
- Svalue ordering $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min\{m,n\}} \geq 0$
- Svalues of matrices B, C: $\sigma_j(B)$, $\sigma_j(C)$

SVD Properties

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

- Number of svalues equal to small dimension

  $A \in \mathbb{R}^{m \times n}$ has $\min\{m, n\}$ singular values $\sigma_j \geq 0$

- Orthogonal invariance

  If $P \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{n \times n}$ are orthogonal matrices

  then $P\,A\,Q$ has same svalues as A

- Gram Product

  Nonzero svalues ($=$ eigenvalues) of $A^T A$

  are squares of svalues of A

- Inverse

  $A \in \mathbb{R}^{n \times n}$ nonsingular $\iff \sigma_j > 0$ for $1 \leq j \leq n$

  If $A = U\,\Sigma\,V^T$ then $A^{-1} = V\,\Sigma^{-1}\,U^T$ SVD of inverse

Exercises

1. Transpose $\quad A^T$ has same singular values as A

2. Orthogonal matrices

   All singular values of $A \in \mathbb{R}^{n \times n}$ are equal to 1

   $\iff$ A is orthogonal matrix

3. If $A \in \mathbb{R}^{n \times n}$ is symmetric and idempotent then

   all singular values are 0 and/or 1

4. For $A \in \mathbb{R}^{m \times n}$ and $\alpha > 0$ express svalues of $(A^T A + \alpha I)^{-1} A^T$ in terms of α and svalues of A

5. If $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ then singular values of $\begin{pmatrix} I_n \\ A \end{pmatrix}$ are equal to $\sqrt{1 + \sigma_j^2}$ for $1 \leq j \leq n$

Singular Values

$A \in \mathbb{R}^{m \times n}$ with svalues $\sigma_1 \geq \cdots \geq \sigma_p \quad p \equiv \min\{m, n\}$

- Two norm $\|A\|_2 = \sigma_1$
- Frobenius norm $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}$
- Well conditioned in absolute sense

$$|\sigma_j(A) - \sigma_j(B)| \leq \|A - B\|_2 \qquad 1 \leq j \leq p$$

- Product

$$\sigma_j(A\,B) \leq \sigma_1(A)\,\sigma_j(B) \qquad 1 \leq j \leq p$$

Matrix Schatten Norms

$A \in \mathbb{R}^{m \times n}$, with singular values $\sigma_1 \geq \cdots \geq \sigma_\rho > 0$, and integer $p \geq 0$, the family of the Schatten p-norms is defined as

$$\|A\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^{\rho} \sigma_i^p \right)^{1/p} .$$

Different than the vector-induced matrix p-norms[1].

- Schatten zero norm[2]: equal to the matrix rank.
- Schatten one norm: the sum of the singular values of the matrix, also called the nuclear norm.
- Schatten two norm: the Frobenius norm.
- Schatten infinity norm: the spectral (or two) norm.
- Schatten p-norms are unitarily invariant, submultiplicative, satisfy Hölder's inequality, etc.

[1] Notation is, unfortunately, confusing.

[2] Not really a norm...

Exercises

1. Norm of inverse

   If $\mathsf{A} \in \mathbb{R}^{n \times n}$ nonsingular with svalues $\sigma_1 \geq \cdots \geq \sigma_n$ then $\|\mathsf{A}^{-1}\|_2 = 1/\sigma_n$

2. Appending a column to a tall and skinny matrix

   If $\mathsf{A} \in \mathbb{R}^{m \times n}$ with $m > n$, $\mathsf{z} \in \mathbb{R}^m$, $\mathsf{B} = \begin{pmatrix} \mathsf{A} & \mathsf{z} \end{pmatrix}$ then

   $$\sigma_{n+1}(\mathsf{B}) \leq \sigma_n(\mathsf{A}) \qquad \sigma_1(\mathsf{B}) \geq \sigma_1(\mathsf{A})$$

3. Appending a row to a tall and skinny matrix

   If $\mathsf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$, $\mathsf{z} \in \mathbb{R}^n$, $\mathsf{B}^T = \begin{pmatrix} \mathsf{A}^T & \mathsf{z} \end{pmatrix}$ then

   $$\sigma_n(\mathsf{B}) \geq \sigma_n(\mathsf{A}) \qquad \sigma_1(\mathsf{A}) \leq \sigma_1(\mathsf{B}) \leq \sqrt{\sigma_1(\mathsf{A})^2 + \|\mathsf{z}\|_2^2}$$

Rank

rank(A) = number of nonzero (positive) svalues of A

- Zero matrix $\mathrm{rank}(0) = 0$
- Rank bounded by small dimension
  If $A \in \mathbb{R}^{m \times n}$ then $\mathrm{rank}(A) \leq \min\{m, n\}$
- Transpose $\mathrm{rank}(A^T) = \mathrm{rank}(A)$
- Gram product $\mathrm{rank}(A^T A) = \mathrm{rank}(A) = \mathrm{rank}(A A^T)$
- General product $\mathrm{rank}(A\,B) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$
- If A nonsingular then $\mathrm{rank}(A\,B) = \mathrm{rank}(B)$

SVDs of Full Rank Matrices

All svalues of $A \in \mathbb{R}^{m \times n}$ are nonzero

- Full column-rank $\text{rank}(A) = n$ {Linearly independent columns}

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \qquad \Sigma \in \mathbb{R}^{n \times n} \text{ nonsingular}$$

- Full row-rank $\text{rank}(A) = m$ {Linearly independent rows}

$$A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V^T \qquad \Sigma \in \mathbb{R}^{m \times m} \text{ nonsingular}$$

- Nonsingular $\text{rank}(A) = n = m$ {Lin. indep. rows & columns}

$$A = U \Sigma V^T \qquad \Sigma \in \mathbb{R}^{n \times n} \text{ nonsingular}$$

Exercises

1. Rank of outer product
   If $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ then $\text{rank}(xy^T) \leq 1$

2. If $A \in \mathbb{R}^{n \times n}$ nonsingular then $\begin{pmatrix} A & B \end{pmatrix}$ has full row-rank for any $B \in \mathbb{R}^{n \times k}$

3. Orthonormal matrices
   If $A \in \mathbb{R}^{m \times n}$ has orthonormal columns then $\text{rank}(A) = n$ and all svalues of A are equal to 1

4. Gram products For $A \in \mathbb{R}^{m \times n}$

   (i) $\text{rank}(A) = n \iff A^T A$ nonsingular
   (ii) $\text{rank}(A) = m \iff AA^T$ nonsingular

Thin SVD

$$A \in \mathbb{R}^{m\times n} \quad \text{with} \quad A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T \quad \text{or} \quad A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V^T$$

Singular values $\sigma_1 \geq \cdots \geq \sigma_p \geq 0, \;\; p \equiv \min\{m, n\}$

Singular vectors $U = \begin{pmatrix} u_1 & \cdots & u_m \end{pmatrix} \qquad V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$

If $\text{rank}(A) = r$ then thin (reduced) SVD $\quad\{\text{only non zero svalues}\}$

$$A = \begin{pmatrix} u_1 & \cdots & u_r \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T \end{pmatrix} = \sum_{j=1}^{r} \sigma_j \, u_j \, v_j^T$$

Optimality of SVD

- Given $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$, thin SVD $\sum_{j=1}^{r} \sigma_j u_j v_j^T$
- Approximation from k dominant svalues

$$A_k \equiv \sum_{j=1}^{k} \sigma_j u_j v_j^T \qquad 1 \leq k < r$$

Optimality of SVD

- Given $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = r$, thin SVD $\sum_{j=1}^{r} \sigma_j u_j v_j^T$
- Approximation from k dominant svalues

$$A_k \equiv \sum_{j=1}^{k} \sigma_j u_j v_j^T \qquad 1 \le k < r$$

- Absolute distance of A to set of rank-k matrices

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

$$\min_{\text{rank}(B)=k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{j=k+1}^{r} \sigma_j^2}$$

Exercises

1. Find an example to illustrate that a closest matrix of rank k is not unique in the two norm

Moore Penrose Inverse

- Given $\quad A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r \geq 1$ and SVD

$$A = U \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V^T = \sum_{j=1}^{r} \sigma_j \, u_j \, v_j^T$$

- Moore Penrose inverse

$$A^\dagger \stackrel{\text{def}}{=} V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T = \sum_{j=1}^{r} \frac{1}{\sigma_j} \, v_j \, u_j^T$$

- Zero matrix $\quad 0_{m\times n}^\dagger = 0_{n\times m}$

Special Cases of Moore Penrose Inverse

- Nonsingular

  If $A \in \mathbb{R}^{n \times n}$ with $\text{rank}(A) = n$ then $A^{\dagger} = A^{-1}$

  Inverse $A^{-1} A = I_n = A A^{-1}$

- Full column rank

  If $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = n$ then $A^{\dagger} = (A^T A)^{-1} A^T$

  Left inverse $A^{\dagger} A = I_n$

- Full row rank

  If $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = m$ then $A^{\dagger} = A^T (A A^T)^{-1}$

  Right inverse $A A^{\dagger} = I_m$

Necessary and Sufficient Conditions

$A^{\dagger}$ is Moore Penrose inverse of $A \iff A^{\dagger}$ satisfies

1. $A\,A^{\dagger}\,A = A$
2. $A^{\dagger}\,A\,A^{\dagger} = A^{\dagger}$
3. $(A\,A^{\dagger})^{T} = A\,A^{\dagger}$ $\quad$ {symmetric}
4. $(A^{\dagger}\,A)^{T} = A^{\dagger}\,A$ $\quad$ {symmetric}

Exercises

1. If $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ with $y \neq 0$ then
$$\|x\, y^\dagger\|_2 = \|x\|_2 / \|y\|_2$$
2. For $A \in \mathbb{R}^{m\times n}$ the following matrices are idempotent:
$$A\,A^\dagger \qquad A^\dagger A \qquad I_m - A\,A^\dagger \qquad I_n - A^\dagger A$$
3. If $A \in \mathbb{R}^{m\times n}$ and $A \neq 0$ then $\|A\,A^\dagger\|_2 = \|A^\dagger A\|_2 = 1$
4. If $A \in \mathbb{R}^{m\times n}$ then
$$(I_m - A\,A^\dagger)\,A = 0_{m\times n} \qquad A\,(I_n - A^\dagger A) = 0_{m\times n}$$
5. If $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = n$ then $\|(A^T A)^{-1}\|_2 = \|A^\dagger\|_2^2$
6. If $A = B\,C$ where $B \in \mathbb{R}^{m\times n}$ has $\text{rank}(B) = n$ and $C \in \mathbb{R}^{n\times n}$ is nonsingular then $A^\dagger = C^{-1} B^\dagger$

Matrix Spaces and Singular Vectors: A

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

- Column space

$$\mathrm{range}(A) = \{b : \; b = A\,x \;\; \text{for some } x \in \mathbb{R}^n\} \;\; \subset \mathbb{R}^m$$

- Null space (kernel)

$$\mathrm{null}(A) = \{x : \; A\,x = 0\} \;\; \subset \mathbb{R}^n$$

If $\mathrm{rank}(A) = r \geq 1$

$$A = \begin{pmatrix} U_r & U_{m-r} \end{pmatrix} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_r^T \\ V_{n-r}^T \end{pmatrix}$$

$$\mathrm{range}(A) = \mathrm{range}(U_r) \qquad \mathrm{null}(A) = \mathrm{range}(V_{n-r})$$

Matrix Spaces and Singular Vectors: A^T

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

- Row space

$$\text{range}(A^T) = \{d : \; d = A^T y \;\; \text{for some } y \in \mathbb{R}^m\} \;\; \subset \mathbb{R}^n$$

- Left null space

$$\text{null}(A^T) = \{y : \; A^T y = 0\} \;\; \subset \mathbb{R}^m$$

If $\text{rank}(A) = r \geq 1$

$$A = \begin{pmatrix} U_r & U_{m-r} \end{pmatrix} \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_r^T \\ V_{n-r}^T \end{pmatrix}$$

$\text{range}(A^T) = \text{range}(V_r)$ $\qquad$ $\text{null}(A^T) = \text{range}(U_{m-r})$

Fundamental Theorem of Linear Algebra

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

$$\text{range}(A) \oplus \text{null}(A^T) = \mathbb{R}^m$$

implies

- $m = \text{rank}(A) + \dim \text{null}(A^T)$
- $\text{range}(A) \perp \text{null}(A^T)$

$$\text{range}(A^T) \oplus \text{null}(A) = \mathbb{R}^n$$

implies

- $n = \text{rank}(A) + \dim \text{null}(A)$
- $\text{range}(A^T) \perp \text{null}(A)$

Spaces of the Moore Penrose Inverse

{Need this for least squares}

- Column space

$$\begin{aligned} \text{range}(A^\dagger) &= \text{range}(A^T A) = \text{range}(A^T) \\ &\perp \text{null}(A) \end{aligned}$$

- Null space

$$\begin{aligned} \text{null}(A^\dagger) &= \text{null}(A\,A^T) = \text{null}(A^T) \\ &\perp \text{range}(A) \end{aligned}$$

Least Squares (LS) Problems

General LS Problems

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

$$\min_x \|Ax - b\|_2$$

- General LS solution $y = A^\dagger b + q$ for any $q \in \text{null}(A)$
- All solutions have same LS residual $r \equiv b - A\,y$
  $\left\{A\,y = A\,A^\dagger\,b \text{ since } q \in \text{null}(A)\right\}$
- LS residual orthogonal to column space $A^T r = 0$
- LS solution of minimal two norm $y = A^\dagger b$
- Computation: SVD

Exercises

1. Use properties of Moore Penrose inverses to show that the LS residual is orthogonal to the column space of A
2. Determine the minimal norm solution for $\min_x \|A x - b\|_2$ if $A = 0_{m \times n}$
3. If y is the minimal norm solution to $\min_x \|A x - b\|_2$ and $A^T b = 0$, then what can you say about y?
4. Determine the minimal norm solution for $\min_x \|A x - b\|_2$ if $A = c\, d^T$ where $c \in \mathbb{R}^m$ and $d \in \mathbb{R}^n$?

Full Column Rank LS Problems

Given $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = n$ and $b \in \mathbb{R}^m$

$$\min_x \|Ax - b\|_2$$

- Unique LS solution $y = A^{\dagger} b$
- Computation: QR decomposition
  1. Factor $A = QR$ where $Q^T Q = I_n$ and R is ◹
  2. Multiply $c = Q^T b$
  3. Solve $Ry = c$
- Do NOT solve $A^T A y = A^T b$

Exercises

1. If $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = n$ has thin QR factorization $A = QR$ where $Q^T Q = I_n$ and R is ◹ then

$$A^\dagger = R^{-1} Q^T$$

2. If $A \in \mathbb{R}^{m \times n}$ has orthonormal columns then $A^\dagger = A^T$

3. If $A \in \mathbb{R}^{m \times n}$ has $\text{rank}(A) = n$ then

$$\|I_m - A A^\dagger\|_2 = \min\{1, m - n\}$$