STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2015
LECTURE 5

1. EXISTENCE OF SVD

Theorem 1 (Existence of SVD)**.** *Every matrix has a singular value decomposition (condensed version).*

Proof. Let $A \in \mathbb{C}^{m\times n}$. We define the matrix

$$W = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \in \mathbb{C}^{(m+n)\times(m+n)}.$$

It is easy to verify that $W = W^*$ (after Wielandt, who's the first to consider this matrix) and by the spectral theorem for Hermitian matrices, W has an EVD,

$$W = Z\Lambda Z^*$$

where $Z \in \mathbb{C}^{(m+n)\times(m+n)}$ is a unitary matrix and $\Lambda \in \mathbb{R}^{(m+n)\times(m+n)}$ is a diagonal matrix with real diagonal elements. If $\mathbf{z}$ is an eigenvector of W, then we can write

$$W\mathbf{z} = \sigma\mathbf{z}, \quad \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

and therefore

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \sigma \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

or, equivalently,

$$A\mathbf{y} = \sigma\mathbf{x}, \quad A^*\mathbf{x} = \sigma\mathbf{y}.$$

Now, suppose that we apply W to the vector obtained from $\mathbf{z}$ by negating $\mathbf{y}$. Then we have

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix} = \begin{bmatrix} -A\mathbf{y} \\ A^*\mathbf{x} \end{bmatrix} = \begin{bmatrix} -\sigma\mathbf{x} \\ \sigma\mathbf{y} \end{bmatrix} = -\sigma \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix}.$$

In other words, if $\sigma \neq 0$ is an eigenvalue, then $-\sigma$ is also an eigenvalue. So we may assume without loss of generality that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = 0 \cdots = 0$$

where $r = \operatorname{rank}(A)$. So the diagonal matrix Λ of eigenvalues of W may be written as

$$\Lambda = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, -\sigma_1, -\sigma_2, \dots, -\sigma_r, 0, \dots, 0) \in \mathbb{C}^{(m+n)\times(m+n)}.$$

Observe that there is a zero block of size $(m+n-2r)\times(m+n-2r)$ in the bottom right corner of Λ.

We scale the eigenvector $\mathbf{z}$ of W so that $\mathbf{z}^*\mathbf{z} = 2$. Since W is symmetric, eigenvectors corresponding to the distinct eigenvalues σ and $-\sigma$ are orthogonal, so it follows that

$$\begin{bmatrix} \mathbf{x}^* & \mathbf{y}^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix} = 0.$$

Date: October 21, 2015, version 1.0. Comments, bug reports: `lekheng@galton.uchicago.edu`.

These yield the system of equations

$$\mathbf{x}^*\mathbf{x} + \mathbf{y}^*\mathbf{y} = 2,$$
$$\mathbf{x}^*\mathbf{x} - \mathbf{y}^*\mathbf{y} = 0,$$

which has the unique solution

$$\mathbf{x}^*\mathbf{x} = 1, \quad \mathbf{y}^*\mathbf{y} = 1.$$

Now note that we can represent the matrix of normalized eigenvectors of W corresponding to nonzero eigenvalues (note that there are exactly $2r$ of these) as

$$\tilde{Z} = \frac{1}{\sqrt{2}} \begin{bmatrix} X & X \\ Y & -Y \end{bmatrix} \in \mathbb{C}^{(m+n)\times 2r}.$$

Note that the factor $1/\sqrt{2}$ appears because of the way we have chosen the norm of $\mathbf{z}$. We also let

$$\tilde{\Lambda} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, -\sigma_1, -\sigma_2, \ldots, -\sigma_r) \in \mathbb{C}^{2r\times 2r}.$$

It is easy to see that

$$Z\Lambda Z^* = \tilde{Z}\tilde{\Lambda}\tilde{Z}^*$$

just by multiplying out the zero block in Λ. So we have

$$\begin{aligned} \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} &= W = Z\Lambda Z^* = \tilde{Z}\Lambda\tilde{Z}^* \\ &= \frac{1}{2}\begin{bmatrix} X & X \\ Y & -Y \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & -\Sigma_r \end{bmatrix} \begin{bmatrix} X^* & Y^* \\ X^* & -Y^* \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} X\Sigma_r & -X\Sigma_r \\ Y\Sigma_r & Y\Sigma_r \end{bmatrix} \begin{bmatrix} X^* & Y^* \\ X^* & -Y^* \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 0 & 2X\Sigma_r Y^* \\ 2Y\Sigma_r X^* & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & X\Sigma_r Y^* \\ Y\Sigma_r X^* & 0 \end{bmatrix} \end{aligned}$$

and therefore

$$A = X\Sigma_r Y^*, \quad A^* = Y\Sigma_r X^*$$

where X is an $m \times r$ matrix, Σ is $r \times r$, and Y is $n \times r$, and r is the rank of A. We have obtained the condensed SVD of A.

The last missing bit is the orthonormality of the columns of X and Y. This follows from the fact that distinct columns of

$$\begin{bmatrix} X & X \\ Y & -Y \end{bmatrix}$$

are mutually orthogonal and so if we pick $\left[\begin{smallmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} \mathbf{x}_j \\ \mathbf{y}_j \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} \mathbf{x}_j \\ -\mathbf{y}_j \end{smallmatrix}\right]$ for $i \neq j$, and take their inner products, we get

$$\mathbf{x}_i^*\mathbf{x}_j + \mathbf{y}_i^*\mathbf{y}_j = 0,$$
$$\mathbf{x}_i^*\mathbf{x}_j - \mathbf{y}_i^*\mathbf{y}_j = 0.$$

Adding them gives $\mathbf{x}_i^*\mathbf{x}_j = 0$ and substracting them gives $\mathbf{y}_i^*\mathbf{y}_j = 0$ for all $i \neq j$, as required. $\square$

- see Theorem 4.1 in Trefethen and Bau for an alternative non-constructive proof that does not require the use of spectral theorem

2. OTHER CHARACTERIZATIONS OF SVD

- the proof of the above theorem gives us two more characterizations of singular values and singular vectors:
  (i) in terms of eigenvalues and eigenvectors of an $(m+n)\times(m+n)$ Hermitian matrix:
$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} U & U \\ V & -V \end{bmatrix}\begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}\begin{bmatrix} U^* & V^* \\ U^* & -V^* \end{bmatrix}$$
  (ii) in terms of a coupled system of equations
$$\begin{cases} A\mathbf{v} = \sigma\mathbf{u}, \\ A^*\mathbf{u} = \sigma\mathbf{v} \end{cases}$$
- the following is yet another way to characterize them in terms of eigenvalues/eigenvectors of an $m \times m$ Hermitian matrix and an $n \times n$ Hermitian matrix

Lemma 1. *The suqare of the singular values of a matrix A are eigenvalues of AA^* and A^*A. The left singular vectors of A are the eigenvectors of AA^* and the right singular vectors of A are the eigenvectors of A^*A.*

Proof. From the relationships $A\mathbf{y} = \sigma\mathbf{x}$, $A^*\mathbf{x} = \sigma\mathbf{y}$, we obtain
$$A^*A\mathbf{y} = \sigma^2\mathbf{y}, \quad AA^*\mathbf{x} = \sigma^2\mathbf{x}.$$
Therefore, if $\pm\sigma$ are eigenvalues of W, then σ^2 is an eigenvalue of both AA^* and A^*A. Also
$$AA^* = (U\Sigma V^*)(V\Sigma^{\mathsf{T}}U^*) = U\Sigma\Sigma^{\mathsf{T}}U^*,$$
$$A^*A = (V\Sigma^{\mathsf{T}}U^*)(U\Sigma V^*) = V\Sigma^{\mathsf{T}}\Sigma V^*.$$
Note that $\Sigma^* = \Sigma^{\mathsf{T}}$ since singular values are real. The matrices $\Sigma^{\mathsf{T}}\Sigma$ and $\Sigma\Sigma^{\mathsf{T}}$ are respectively $n \times n$ and $m \times m$ diagonal matrices with diagonal elements σ_i^2 and 0. □

- the SVD is something like a swiss army knife of linear algebra, matrix theory, and numerical linear algebra, you can do a lot with it
- over the next few sections we will see that the singular value decomposition is a singularly powerful tool — once we have it, we could solve just about any problem involving matrices
  - given $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \operatorname{im}(A) \subseteq \mathbb{C}^m$, find all solutions of $A\mathbf{x} = \mathbf{b}$
  - given $A \in \mathrm{GL}(n)$, find A^{-1}
  - given $A \in \mathbb{C}^{m\times n}$, find $\|A\|_2$ and $\|A\|_F$
  - given $A \in \mathbb{C}^{m\times n}$, find $\|A\|_{\sigma,p,k}$ for $p \in [0,\infty]$ and $k \in \mathbb{N}$
  - given $A \in \mathbb{C}^{n\times n}$, find $|\det(A)|$
  - given $A \in \mathbb{C}^{m\times n}$, find $A^\dagger$
  - given $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$, find all solutions of the $\min_{\mathbf{x}\in\mathbb{C}^n} \|A\mathbf{x} - \mathbf{b}\|_2$
- the good news is that unlike the Jordan canonical form, the SVD is actually computable
- there are two main methods to compute it: Golub-Reinsch and Golub-Kahan, we will look at these briefly later, right now all you need to know is that you can call MATLAB to give you the SVD, both the full and compact versions
- in all of the following we shall assume that we have the full SVD of $A = U\Sigma V^*$
- furthermore $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A and $r = \operatorname{rank}(A)$

3. SOLVING LINEAR SYSTEMS

- assuming that we want to solve a linear system $A\mathbf{x} = \mathbf{b}$ that is known to be consistent, i.e. $\mathbf{b} \in \operatorname{im}(A)$
- first we form the matrix-vector product $\mathbf{c} = U^*\mathbf{b}$
- let $\mathbf{y} = V^*\mathbf{x}$

- then $A\mathbf{x} = \mathbf{b}$ becomes $\Sigma\mathbf{y} = \mathbf{c}$
- since $A\mathbf{x} = \mathbf{b}$ is consistent, so

$$\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_n \end{bmatrix}$$

  is consistent, which means that we must have $c_{r+1} = \cdots = c_m = 0$, i.e.

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- now let

$$\mathbf{y} = \begin{bmatrix} c_1/\sigma_1 \\ \vdots \\ c_r/\sigma_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix}$$

  where $y_{r+1}, \dots, y_n$ are free parameters
- all solutions of $A\mathbf{x} = \mathbf{b}$ are given by

$$\mathbf{x} = V\mathbf{y}$$

  for some choices of $y_{r+1}, \dots, y_n$
- to see this, observe that

$$A\mathbf{x} = U\Sigma V^* V\mathbf{y} = U\Sigma\mathbf{y} = U \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \begin{bmatrix} c_1/\sigma_1 \\ \vdots \\ c_r/\sigma_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix} = U \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = UU^*\mathbf{b} = \mathbf{b}$$

4. Inverting nonsingular matrices

- note that only square matrices could have inverses, when we say A is nonsinular or invertible, the fact that A is square is implied
- by the way, the set of all nonsingular matrices in $\mathbb{C}^{n\times n}$ forms group under matrix multiplication
- it is called the general linear group and denoted

$$\mathrm{GL}(n) := \{A \in \mathbb{C}^{n\times n} : \det(A) \neq 0\}$$

- if $A \in \mathbb{C}^{n\times n}$ is nonsingular, then $\mathrm{rank}(A) = n$ and so $\sigma_1, \dots, \sigma_n$ are all non-zero

- so
$$A^{-1} = (U\Sigma V^*)^{-1} = (V^*)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^*$$
where
$$\Sigma^{-1} = \operatorname{diag}(\sigma_1^{-1}, \dots, \sigma_n^{-1})$$

5. Computing matrix 2-norm and F-norm

- recall the definition of the matrix 2-norm,
$$\|A\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$
which is also called the spectral norm
- we examine the expression
$$\|A\mathbf{x}\|_2^2 = x^*A^*Ax$$
- the matrix A^*A is Hermitian and positive semidefinite, i.e. $\mathbf{x}^*(A^*A)\mathbf{x} \geq 0$ for all nonzero $\mathbf{x} \in \mathbb{C}^n$
- exercise: show that if a matrix M is Hermitian positive semidefinite, then its EVD and SVD coincide
- as such, A^*A has SVD given by
$$A^*A = V\Sigma V^*$$
where V is a unitary matrix whose columns are the eigenvectors of A^*A, and Σ is a diagonal matrix of the form
$$\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix}$$
where each σ_i^2 is nonnegative and an eigenvalue of A^*A
- these eigenvalues can be ordered such that
$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 > 0, \quad \sigma_{r+1}^2 = \cdots = \sigma_n^2 = 0,$$
where $r = \operatorname{rank}(A)$
- let $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ and let $\mathbf{w} = V^*\mathbf{x}$, then we obtain
$$\begin{aligned} \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} &= \frac{\mathbf{x}^*A^*A\mathbf{x}}{\mathbf{x}^*\mathbf{x}} \\ &= \frac{\mathbf{x}^*V\Sigma V^*\mathbf{x}}{\mathbf{x}^*VV^*\mathbf{x}} \\ &= \frac{\mathbf{w}^*\Sigma\mathbf{w}}{\mathbf{w}^*\mathbf{w}} \\ &= \frac{\sum_{i=1}^n \sigma_i^2|w_i|^2}{\sum_{i=1}^n |w_i|^2} \\ &\leq \sigma_1^2 \end{aligned}$$
- exercise: show that if $a_1, \dots, a_n$ are nonnegative numbers, then
$$\max_{x_i \geq 0} \frac{a_1x_1 + \cdots + a_nx_n}{x_1 + \cdots + x_n} = \max(a_1, \dots, a_n)$$
where the maximum is taken over $x_1, \dots, x_n \in [0, \infty)$ not all 0.
- it follows that
$$\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \sigma_1$$
for all nonzero $\mathbf{x}$

- since V is a unitary matrix, it follows that there exists an $\mathbf{x}$ such that

$$\mathbf{w} = V^*\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1$$

- in which case

$$\mathbf{x}^* A^* A \mathbf{x} = \mathbf{e}_1^* \Sigma \mathbf{e}_1 = \sigma_1^2$$

- in fact, this vector $\mathbf{x}$ is the eigenvector of A^*A corresponding to the eigenvalue σ_1^2
- we conclude that

$$\|A\|_2 = \sigma_1$$

- note that we have also shown that

$$\|A\|_2 = \sqrt{\rho(A^*A)}$$

  since the eigenvalues of A^*A are simply the squares of the singular values of A by Lemma 1
- another way to arrive at this same conclusion is to use the fact in Homework 1 that the 2-norm of a vector is invariant under multiplication by a unitary matrix, i.e. if $Q^*Q = I$, then $\|\mathbf{x}\|_2 = \|Q\mathbf{x}\|_2$, from which it follows that

$$\|A\|_2 = \|U\Sigma V^*\|_2 = \|\Sigma\|_2 = \sigma_1$$

- by the same problem in Homework 1, the Frobenius norm is also unitarily invarint
- this yields an expression in terms of singular values

$$\|A\|_F = \|U\Sigma V^*\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

  where $r = \operatorname{rank}(A)$

6. Computing other matrix norms

- the Schatten and Ky Fan norms can be expressed in terms of singular values
- this is cheating a bit, because we will in fact *define* them in terms of singular values
- for any $p \in [1, \infty]$, the *Schatten p-norm* of $A \in \mathbb{C}^{m\times n}$ is

$$\|A\|_{\sigma,p} := \left[\sum_{i=1}^{\min(m,n)} \sigma_i(A)^p\right]^{1/p}$$

  and for $p = \infty$, we have

$$\|A\|_{\sigma,\infty} = \max\{\sigma_1(A), \dots, \sigma_{\min(m,n)}(A)\}$$

- of course, the sum will stop at $r = \operatorname{rank}(A)$ since $\sigma_{r+1}(A) = \cdots = \sigma_{\min(m,n)}(A) = 0$
- as usual, we also have

$$\lim_{p\to\infty} \|A\|_{\sigma,p} = \|A\|_{\sigma,\infty} \tag{6.1}$$

  for all $A \in \mathbb{C}^{m\times n}$
- for the special values of $p = 1, 2, \infty$, we have
  - $p = 1$: nuclear norm

$$\|A\|_{\sigma,1} = \sigma_1(A) + \cdots + \sigma_{\min(m,n)}(A) = \|A\|_*$$

  - $p = 2$: Frobenius norm

$$\|A\|_{\sigma,2} = \sqrt{\sigma_1(A)^2 + \cdots + \sigma_{\min(m,n)}(A)^2} = \|A\|_F$$

  – $p = \infty$: spectral norm
$$\|A\|_{\sigma,\infty} = \max\{\sigma_1(A), \dots, \sigma_{\min(m,n)}(A)\} = \sigma_1(A) = \|A\|_2$$

- in fact one may also define 'Schatten p-norm' for values of $p \in [0,1)$ by dropping the root $1/p$ outside the sum
$$\|A\|_{\sigma,p} := \sum_{i=1}^{\min(m,n)} \sigma_i(A)^p$$
- these will not be norms in the usual sense of the word because they do not satisfy the triangle inequality but they satisfy an analogue of (6.1)
$$\lim_{p\to 0} \|A\|_{\sigma,p} = \operatorname{rank}(A)$$
- we may regard matrix rank as the 'Schatten 0-norm' since
$$\|A\|_{\sigma,0} = \sigma_1(A)^0 + \cdots + \sigma_{\min(m,n)}(A)^0 = \operatorname{rank}(A)$$
provided we define $0^0 := 0$
- 'Schatten p-norm' for $p < 1$ are useful in the same way 'ℓ^p-norms' are useful for $p < 1$, namely, as continuous surrogates of matrix rank (vector 'ℓ^p-norms' for $p < 1$ are often used as continuous surrogates as the ℓ^0-norm, i.e. a count of the number of nonzero entries), which is a discontinuous function on $\mathbb{C}^{m\times n}$
- the nuclear norm, i.e. Schatten 1-norm, being convex in addition to continuous, is the most popular surrogate for matrix rank
- for any $p \in [1,\infty)$ and $k \in \mathbb{N}$ the *Ky Fan* (p,k)*-norm* of $A \in \mathbb{C}^{m\times n}$ is
$$\|A\|_{\sigma,p,k} := \left[\sum_{i=1}^{k} \sigma_i(A)^p\right]^{1/p}$$
- clearly for any $A \in \mathbb{C}^{m\times n}$,
$$\|A\|_{\sigma,p,\infty} = \|A\|_{\sigma,p,\min(m,n)} = \|A\|_{\sigma,p}$$
and so Ky Fan norms generalize Schatten norms

7. Computing magnitude of determinant

- note that determinants are only defined for square matrices $A \in \mathbb{C}^{n\times n}$
- recall the formula
$$\det(A) = \prod_{i=1}^{n} \lambda_i(A) \tag{7.1}$$
where $\lambda_i(A) \in \mathbb{C}$ is the ith eigenvalue of A
  – recall that every $n \times n$ matrix has exactly n eigenvalues counted with multiplicity
  – by convention we usually order eigenvalues in decreasing order of magnitudes, i.e.
$$|\lambda_1(A)| \ge |\lambda_2(A)| \ge \cdots \ge |\lambda_n(A)|$$
- now if we have only the singular values of A we cannot get $\det(A)$ like we did with its eigenvalues in (7.1)
- however we can get the magnitude of the determinant
$$|\det(A)| = |\det(U\Sigma V^*)| = |\det(U)\det(\Sigma)\det(V)| = \det(\Sigma) = \det\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix} = \prod_{i=1}^{n}\sigma_i(A)$$
since $|\det(U)| = |\det(V)| = 1$

- exercise: show that all eigenvalues of a unitary matrix U must have absolute value 1 and so $|\det(U)| = 1$ by (7.1)

8. Computing pseudoinverse

- as we mentioned above, only square matrices $A \in \mathbb{C}^{n\times n}$ may have an inverse, i.e. $X \in \mathbb{C}^{n\times n}$ such that
$$AX = XA = I$$
- if such an X exists, we denote it by A^{-1}
- exercise: show that A^{-1}, if exist, must be unique
- can we define something that behaves like an inverse (we will say exactly what this means in the next lecture when we discuss minimum length least squares problem) for square matrices that are not invertible in the traditional sense?
- more generally, can we define some kind of inverse for rectangular matrices?
- such considerations lead us to the notion of pseudoinverse
- the most famous one is the *Moore–Penrose pseudoinverse*

Theorem 2 (Moore–Penrose)**.** *For any $A \in \mathbb{C}^{m\times n}$, there exists a $X \in \mathbb{C}^{n\times m}$ satisfying*

(1) $(AX)^* = AX$,
(2) $(XA)^* = XA$,
(3) $XAX = X$,
(4) $AXA = A$.

Furthermore the X satisfying these four conditions must be unique and is denoted by $X = A^\dagger$.

- other types of pseudoinverse, sometimes also called *generalized inverse*, may be defined by choosing a subset of these four properties
- the Moore–Penrose theorem is actually true over any field but for $\mathbb{C}$ (and also $\mathbb{R}$) we can say a lot more
- in fact we can compute $A^\dagger$ explicitly using the SVD
- easy fact: if $A \in \mathbb{C}^{n\times n}$ is invertible, then
$$A^{-1} = A^\dagger \tag{8.1}$$
- to show this, just see that all four properties in the theorem hold true if we plug in $X = A^{-1}$
- another easy fact: if $D = \operatorname{diag}(d_1, \dots, d_{\min(m,n)}) \in \mathbb{C}^{m\times n}$ is diagonal in the sense that $d_{ij} = 0$ for all $i \neq j$, then
$$D^\dagger = \operatorname{diag}(\delta_1, \dots, \delta_{\min(m,n)}) \in \mathbb{C}^{n\times m} \tag{8.2}$$
- where
$$\delta_i = \begin{cases} 1/d_i & \text{if } d_i \neq 0 \\ 0 & \text{if } d_i = 0 \end{cases}$$
- to show this, just see that all four properties in the theorem hold true if we plug in $X = \operatorname{diag}(\delta_1, \dots, \delta_{\min(m,n)})$
- in general
$$(AB)^\dagger \neq B^\dagger A^\dagger, \quad AA^\dagger \neq I, \quad A^\dagger A \neq I$$
- we can get $A^\dagger$ in terms of the SVD of $A \in \mathbb{C}^{m\times n}$: if $A = U\Sigma V^*$, then
$$A^\dagger = V\Sigma^\dagger U^* \tag{8.3}$$

where

$$\Sigma^\dagger = \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \ddots & & & & \\ & & \sigma_r^{-1} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} \in \mathbb{C}^{n\times m}$$

- to see this, just verify that (8.3) satisfies the four properties in Theorem 2
- we may also deduce from (8.3) that

$$A^\dagger = (A^*A)^{-1}A^*$$

if A has full column rank ($r = n$) and that

$$A^\dagger = A^*(AA^*)^{-1}$$

if A has full row rank ($r = m$)