FINM 331: DATA ANALYSIS FOR FINANCE AND STATISTICS
FALL 2015
SINGULAR VALUE DECOMPOSITION

1. DEFINITION

- let $A \in \mathbb{R}^{m \times n}$, we can always write
$$A = U\Sigma V^{\mathsf{T}} \tag{1.1}$$
  - $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are both orthogonal matrices
$$U^{\mathsf{T}}U = I_m = UU^{\mathsf{T}}, \quad V^{\mathsf{T}}V = I_n = VV^{\mathsf{T}}$$
  - $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix in the sense that $\sigma_{ij} = 0$ if $i \neq j$
  - if $m > n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$
  - if $m < n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & \sigma_m & 0 & \cdots & 0 \end{bmatrix}$$
  - if $m = n$, then Σ looks like
$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{bmatrix}$$
  - the diagonal elements of Σ, denoted σ_i, $i = 1, \dots, n$, are all nonnegative, and can be ordered such that
$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \quad \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$$
  - r is the rank of A
- this decomposition of A is called the *singular value decomposition*, or SVD
  - the values σ_i, for $i = 1, 2, \dots, n$, are the *singular values* of A
  - the columns of U are the *left singular vectors*
  - the columns of V are the *right singular vectors*
- an alternative decomposition of A omits the singular values that are equal to zero:
$$A = \tilde{U}\tilde{\Sigma}\tilde{V}^{\mathsf{T}}$$
  - $\tilde{U} \in \mathbb{R}^{m \times r}$ is a matrix with orthonormal columns, i.e. satisfying $\tilde{U}^{\mathsf{T}}\tilde{U} = I_r$ (but not $\tilde{U}\tilde{U}^{\mathsf{T}} = I_m$!)

Date: November 7, 2015, version 1.2. Comments, bug reports: `lekheng@galton.uchicago.edu`.

- $\tilde{V} \in \mathbb{R}^{n\times r}$ is also a matrix with orthonormal columns, i.e. satisfying $\tilde{V}^\mathsf{T}\tilde{V} = I_r$ (but again not $\tilde{V}\tilde{V}^\mathsf{T} = I_n$!)
- $\tilde{\Sigma}$ is an $r \times r$ diagonal matrix with diagonal elements $\sigma_1, \dots, \sigma_r$
- again $r = \operatorname{rank}(A)$
- the columns of $\tilde{U}$ are the left singular vectors corresponding to the nonzero singular values of A, and form an orthonormal basis for the range of A
- the columns of $\tilde{V}$ are the right singular vectors corresponding to the nonzero singular values of A, and form an orthonormal basis for the cokernel of A
- this is called the *condensed* or *compact* or *reduced* SVD
- note that in this case, $\tilde{\Sigma}$ is a square matrix

• the form in (**??**) is sometimes called the *full* SVD

• we may also write the reduced SVD as a sum of rank-1 matrices

$$A = \sigma_1\mathbf{u}_1\mathbf{v}_1^\mathsf{T} + \sigma_2\mathbf{u}_2\mathbf{v}_2^\mathsf{T} + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^\mathsf{T}$$

- $\tilde{U} = [\mathbf{u}_1, \dots, \mathbf{u}_r]$, i.e. $\mathbf{u}_1, \dots, \mathbf{u}_r \in \mathbb{R}^m$ are the left singular vectors of A
- $\tilde{V} = [\mathbf{v}_1, \dots, \mathbf{v}_r]$, i.e. $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{R}^n$ are the right singular vectors of A
- note that for $\mathbf{x} = [x_1, \dots, x_m]^\mathsf{T} \in \mathbb{R}^m$ and $\mathbf{y} = [y_1, \dots, y_n]^\mathsf{T} \in \mathbb{R}^n$,

$$\mathbf{x}\mathbf{y}^\mathsf{T} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}, \quad \mathbf{x}\mathbf{y}^\mathsf{T} = \begin{bmatrix} x_1\bar{y}_1 & x_1\bar{y}_2 & \cdots & x_1\bar{y}_n \\ x_2\bar{y}_1 & x_2\bar{y}_2 & \cdots & x_2\bar{y}_n \\ \vdots & \vdots & & \vdots \\ x_m\bar{y}_1 & x_m\bar{y}_2 & \cdots & x_m\bar{y}_n \end{bmatrix}$$

- if neither $\mathbf{x}$ nor $\mathbf{y}$ is the zero vector, then

$$\operatorname{rank}(\mathbf{x}\mathbf{y}^\mathsf{T}) = \operatorname{rank}(\mathbf{x}\mathbf{y}^\mathsf{T}) = 1$$

- furthermore if $\operatorname{rank}(A) = 1$, then there exists $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ so that $A = \mathbf{x}\mathbf{y}^\mathsf{T}$

2. 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

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

It is easy to verify that $W = W^\mathsf{T}$ (after Wielandt, who's the first to consider this matrix) and by the spectral theorem for Hermitian matrices, W has an EVD,

$$W = Z\Lambda Z^\mathsf{T}$$

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

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

or, equivalently,

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

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

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

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

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

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

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

These yield the system of equations

$$\begin{aligned} \mathbf{x}^\mathsf{T}\mathbf{x} + \mathbf{y}^\mathsf{T}\mathbf{y} &= 2, \\ \mathbf{x}^\mathsf{T}\mathbf{x} - \mathbf{y}^\mathsf{T}\mathbf{y} &= 0, \end{aligned}$$

which has the unique solution

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

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

$$Z\Lambda Z^\mathsf{T} = \tilde{Z}\tilde{\Lambda}\tilde{Z}^\mathsf{T}$$

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

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

and therefore

$$A = X\Sigma_r Y^\mathsf{T}, \quad A^\mathsf{T} = Y\Sigma_r X^\mathsf{T}$$

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

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

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

3. 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 $(m+n)\times(m+n)$ symmetric matrix:
  $$\begin{bmatrix} 0 & A \\ A^\mathsf{T} & 0 \end{bmatrix} = \begin{bmatrix} U & U \\ V & -V \end{bmatrix}\begin{bmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{bmatrix}\begin{bmatrix} U^\mathsf{T} & V^\mathsf{T} \\ U^\mathsf{T} & -V^\mathsf{T} \end{bmatrix}$$
  (ii) in terms of a coupled system of equations
  $$\begin{cases} A\mathbf{v} = \sigma\mathbf{u}, \\ A^\mathsf{T}\mathbf{u} = \sigma\mathbf{v} \end{cases}$$
- the following is yet another way to characterize them in terms of eigenvalues/eigenvectors of an $m \times m$ symmetric matrix and an $n \times n$ symmetric matrix

Lemma 1. *The sugare of the singular values of a matrix A are eigenvalues of AA^T and $A^\mathsf{T}A$. The left singular vectors of A are the eigenvectors of AA^T and the right singular vectors of A are the eigenvectors of $A^\mathsf{T}A$.*

Proof. From the relationships $A\mathbf{y} = \sigma\mathbf{x}$, $A^\mathsf{T}\mathbf{x} = \sigma\mathbf{y}$, we obtain

$$A^\mathsf{T}A\mathbf{y} = \sigma^2\mathbf{y}, \quad AA^\mathsf{T}\mathbf{x} = \sigma^2\mathbf{x}.$$

Therefore, if $\pm\sigma$ are eigenvalues of W, then σ^2 is an eigenvalue of both AA^T and $A^\mathsf{T}A$. Also

$$AA^\mathsf{T} = (U\Sigma V^\mathsf{T})(V\Sigma^\mathsf{T}U^\mathsf{T}) = U\Sigma\Sigma^\mathsf{T}U^\mathsf{T},$$
$$A^\mathsf{T}A = (V\Sigma^\mathsf{T}U^\mathsf{T})(U\Sigma V^\mathsf{T}) = V\Sigma^\mathsf{T}\Sigma V^\mathsf{T}.$$

Note that $\Sigma^\mathsf{T} = \Sigma^\mathsf{T}$ 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 $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \operatorname{im}(A) \subseteq \mathbb{R}^m$, find all solutions of $A\mathbf{x} = \mathbf{b}$
  - given $A \in \mathbb{R}^{n\times n}$ nonsingular, find A^{-1}
  - given $A \in \mathbb{R}^{m\times n}$, find $\|A\|_2$ and $\|A\|_F$
  - given $A \in \mathbb{R}^{n\times n}$, find $|\det(A)|$
  - given $A \in \mathbb{R}^{m\times n}$, find $A^\dagger$
  - given $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, find all solutions of the $\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2$
- 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 $A = U\Sigma V^{\mathsf{T}}$
- furthermore $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the singular values of A and $r = \operatorname{rank}(A)$

4. 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
$$\|A\mathbf{x}\|_2^2 = x^{\mathsf{T}} A^{\mathsf{T}} A x$$
- the matrix $A^{\mathsf{T}}A$ is symmetric and positive semidefinite, i.e. $\mathbf{x}^{\mathsf{T}}(A^{\mathsf{T}}A)\mathbf{x} \geq 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$
- exercise: show that if a matrix M is symmetric positive semidefinite, then its EVD and SVD coincide
- as such, $A^{\mathsf{T}}A$ has SVD given by
$$A^{\mathsf{T}}A = V\Sigma V^{\mathsf{T}}$$
where V is a orthogonal matrix whose columns are the eigenvectors of $A^{\mathsf{T}}A$, 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 $A^{\mathsf{T}}A$
- 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 $\mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n$ and let $\mathbf{w} = V^{\mathsf{T}}\mathbf{x}$, then we obtain
$$\begin{aligned} \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} &= \frac{\mathbf{x}^{\mathsf{T}}A^{\mathsf{T}}A\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} \\ &= \frac{\mathbf{x}^{\mathsf{T}}V\Sigma V^{\mathsf{T}}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}VV^{\mathsf{T}}\mathbf{x}} \\ &= \frac{\mathbf{w}^{\mathsf{T}}\Sigma\mathbf{w}}{\mathbf{w}^{\mathsf{T}}\mathbf{w}} \\ &= \frac{\sum_{i=1}^n \sigma_i^2 |w_i|^2}{\sum_{i=1}^n |w_i|^2} \\ &\leq \sigma_1^2 \end{aligned}$$
- 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
$$\frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \sigma_1$$
for all nonzero $\mathbf{x}$
- since V is a orthogonal matrix, it follows that there exists an $\mathbf{x}$ such that
$$\mathbf{w} = V^{\mathsf{T}}\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{e}_1$$
- in which case
$$\mathbf{x}^{\mathsf{T}} A^{\mathsf{T}} A\mathbf{x} = \mathbf{e}_1^{\mathsf{T}} \Sigma \mathbf{e}_1 = \sigma_1^2$$
- in fact, this vector $\mathbf{x}$ is the eigenvector of $A^{\mathsf{T}}A$ corresponding to the eigenvalue σ_1^2
- we conclude that
$$\|A\|_2 = \sigma_1$$
- note that we have also shown that
$$\|A\|_2 = \sqrt{\rho(A^{\mathsf{T}}A)}$$
since the eigenvalues of $A^{\mathsf{T}}A$ are simply the squares of the singular values of A by Lemma **??**
- another way to arrive at this same conclusion is to use the that the 2-norm of a vector is invariant under multiplication by a orthogonal matrix, i.e. if $Q^{\mathsf{T}}Q = I$, then $\|\mathbf{x}\|_2 = \|Q\mathbf{x}\|_2$, from which it follows that
$$\|A\|_2 = \|U\Sigma V^{\mathsf{T}}\|_2 = \|\Sigma\|_2 = \sigma_1$$
- exercise: show that the Frobenius norm is also orthogonally invariant
- this yields an expression in terms of singular values
$$\|A\|_F = \|U\Sigma V^{\mathsf{T}}\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$
where $r = \operatorname{rank}(A)$

5. Four fundamental subspaces

- given $A \in \mathbb{R}^{m\times n}$ we may regard it as a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \mapsto A\mathbf{x}$
- there are four subspaces associated with A that we call the fundamental subspaces
  - $\ker(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$
  - $\operatorname{im}(A) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}$
  - $\ker(A^{\mathsf{T}}) = \{\mathbf{y} \in \mathbb{R}^m : A^{\mathsf{T}}\mathbf{x} = \mathbf{0}\}$
  - $\operatorname{im}(A^{\mathsf{T}}) = \{\mathbf{x} \in \mathbb{R}^n : A^{\mathsf{T}}\mathbf{y} = \mathbf{x} \text{ for some } \mathbf{y} \in \mathbb{R}^m\}$
- they are called the kernel, image, cokernel, and coimage respectively
  - the word null space is often used in place of kernel
  - the word range space is often used in place of image
- exercise: show that they are related by
$$\ker(A^{\mathsf{T}}) = \operatorname{im}(A)^{\perp} \quad \text{and} \quad \operatorname{im}(A^{\mathsf{T}}) = \ker(A)^{\perp}$$
- furthermore they decompose the domain and codomain of A into orthogonal subspaces
$$\mathbb{R}^n = \operatorname{im}(A^{\mathsf{T}}) \oplus \ker(A) \quad \text{and} \quad \mathbb{R}^m = \ker(A^{\mathsf{T}}) \oplus \operatorname{im}(A)$$
- this decomposition is sometimes called *Fredholm alternative* and is very useful for studying linear systems $A\mathbf{x} = \mathbf{b}$ and least squares problems
- the full SVD of A allows us to simply read off orthonormal bases for the four fundamental subspaces, which will be useful if we want to compute projections

- let $\operatorname{rank}(A) = r$ and let $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the left and right singular vectors of A, indexed in the usual way in desceding magnitude of their corresponding singular values
  - $\ker(A) = \operatorname{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$
  - $\operatorname{im}(A) = \operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$
  - $\ker(A^\mathsf{T}) = \operatorname{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$
  - $\operatorname{im}(A^\mathsf{T}) = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$
- this is easy to see, let us do $\operatorname{im}(A)$ for example
  - first observe that an orthonormal basis for $\operatorname{im}(\Sigma)$ is the first r standard basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_r \in \mathbb{R}^m$ since
$$\Sigma\mathbf{x} = \sigma_1 x_1 \mathbf{e}_1 + \cdots + \sigma_r x_r \mathbf{e}_r$$
  - now just observe that
$$\operatorname{im}(A) = \operatorname{im}(U\Sigma V^\mathsf{T}) = \operatorname{im}(U\Sigma) = \operatorname{span}\{U\mathbf{e}_1, \dots, U\mathbf{e}_r\} = \operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$$
  - there is a useful fact worth noting: for nonsingular matrices $S \in \mathbb{R}^{m\times m}$ and $T \in \mathbb{R}^{n\times n}$,
$$\operatorname{im}(AT) = \operatorname{im}(A) \quad \text{and} \quad \ker(SA) = \ker(A)$$

6. PSEUDOINVERSE

- as we mentioned above, only square matrices $A \in \mathbb{R}^{n\times n}$ may have an inverse, i.e. $X \in \mathbb{R}^{n\times n}$ 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 $A \in \mathbb{R}^{m\times n}$, there exists a $X \in \mathbb{R}^{n\times m}$ satisfying*
(1) $(AX)^\mathsf{T} = AX$,
(2) $(XA)^\mathsf{T} = XA$,
(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 $\mathbb{R}$ (and also $\mathbb{R}$) we can say a lot more
- in fact we can compute $A^\dagger$ explicitly using the SVD
- easy fact: if $A \in \mathbb{R}^{n\times n}$ is invertible, then
$$A^{-1} = A^\dagger \tag{6.1}$$
- 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 $D = \operatorname{diag}(d_1, \dots, d_{\min(m,n)}) \in \mathbb{R}^{m\times n}$ is diagonal in the sense that $d_{ij} = 0$ for all $i \neq j$, then
$$D^\dagger = \operatorname{diag}(\delta_1, \dots, \delta_{\min(m,n)}) \in \mathbb{R}^{n\times m} \tag{6.2}$$

- 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 $A \in \mathbb{R}^{m\times n}$: if $A = U\Sigma V^\mathsf{T}$, then

$$A^\dagger = V\Sigma^\dagger U^\mathsf{T} \tag{6.3}$$

where

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

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

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

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

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

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

7. PROJECTIONS

- the solution $\mathbf{x}$ of the least-squares problem minimizes $\|A\mathbf{x} - \mathbf{b}\|_2$, and therefore is the vector that solves the system $A\mathbf{x} = \mathbf{b}$ as closely as possible
- we can use the SVD to show that $\mathbf{x}$ is the exact solution to a related system of equations
- write $\mathbf{b} = \mathbf{b}_0 + \mathbf{b}_1$, where

$$\mathbf{b}_1 = AA^\dagger\mathbf{b}, \quad \mathbf{b}_0 = (I - AA^\dagger)\mathbf{b}$$

- the matrix $AA^\dagger$ has the form

$$AA^\dagger = U\Sigma V^\mathsf{T} V\Sigma^\dagger U^\mathsf{T} = U\Sigma\Sigma^\dagger U^\mathsf{T} = U\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^\mathsf{T}$$

- it follows that $\mathbf{b}_1$ is a linear combination of $\mathbf{u}_1, \dots, \mathbf{u}_r$, the columns of U that form an orthogonal basis for the range of A
- from $\mathbf{x} = A^\dagger\mathbf{b}$ we obtain

$$A\mathbf{x} = AA^\dagger\mathbf{b} = P_1\mathbf{b} = \mathbf{b}_1$$

where $P_1 = AA^\dagger \in \mathbb{R}^{m\times m}$
- therefore, the solution to the least squares problem, is also the exact solution to the system

$$A\mathbf{x} = P_1\mathbf{b}$$

- it can be shown that the matrix P_1 is an *orthogonal projection*
- in general a matrix $P \in \mathbb{R}^{m\times m}$ is called a *projection* if $P^2 = P$ (this condition is also called idempotent in ring theory)
- a projection is called an orthogonal projection if it is also symmetric, i.e. an orthogonal projection is a matrix $P \in \mathbb{R}^{m\times m}$ satisfying

(i) $P = P^{\mathsf{T}}$
(ii) $P^2 = P$

- caveat: an orthogonal projection is in general *not* an orthogonal/orthogonal matrix (i.e., $P^{\mathsf{T}} \neq P^{-1}$) in fact, projections are usually non-invertible
- example: $\left[\begin{smallmatrix} 1 & \alpha \\ 0 & 0 \end{smallmatrix}\right]$ is a projection for any $\alpha \in \mathbb{R}$, it is an orthogonal projection if and only if $\alpha = 0$
- if $P \in \mathbb{R}^{m\times m}$ is a projection and $\operatorname{im}(P) = W$, we say that P is a projection onto the subspace W
- if $P \in \mathbb{R}^{m\times m}$ is a projection matrix, then $I - P$ is also a projection
- furthermore if $\operatorname{im}(P) = W$ and $\operatorname{im}(I - P) = W'$, then

$$\mathbb{R}^m = W \oplus W'$$

- if P is an orthogonal projection and $\operatorname{im}(P) = W$, then $\operatorname{im}(I - P) = W^{\perp}$
- we sometimes write P_W if we know the subspace P that projects onto
- in particular, $P_1 = AA^{\dagger}$ is a projection onto the space spanned by the columns of A, i.e., $\operatorname{im}(A)$, so $P_1 = P_{\operatorname{im}(A)}$

8. Projections onto fundamental subspaces

- we can write down the orthogonal projections onto all four fundamental subspaces in terms of the pseudoinverse

$$P_{\operatorname{im}(A)} = AA^{\dagger}, \quad P_{\ker(A^{\mathsf{T}})} = I - AA^{\dagger}, \quad P_{\operatorname{im}(A^{\mathsf{T}})} = A^{\dagger}A, \quad P_{\ker(A)} = I - A^{\dagger}A$$

- note that $P_{\operatorname{im}(A)}, P_{\ker(A^{\mathsf{T}})} \in \mathbb{R}^{m\times m}$ and $P_{\operatorname{im}(A^{\mathsf{T}})}, P_{\ker(A)} \in \mathbb{R}^{n\times n}$
- with the SVD, we can write down the projections in terms of unitary matrices

$$P_{\operatorname{im}(A)} = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^{\mathsf{T}} = U_r U_r^{\mathsf{T}}, \quad P_{\ker(A^{\mathsf{T}})} = U \begin{bmatrix} 0 & 0 \\ 0 & I_{m-r} \end{bmatrix} U^{\mathsf{T}} = U_{m-r} U_{m-r}^{\mathsf{T}},$$

$$P_{\operatorname{im}(A^{\mathsf{T}})} = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^{\mathsf{T}} = V_r V_r^{\mathsf{T}}, \quad P_{\ker(A)} = V \begin{bmatrix} 0 & 0 \\ 0 & I_{n-r} \end{bmatrix} V^{\mathsf{T}} = V_{n-r} V_{n-r}^{\mathsf{T}}$$

where $U = [U_r, U_{m-r}]$ and $V = [V_r, V_{m-r}]$