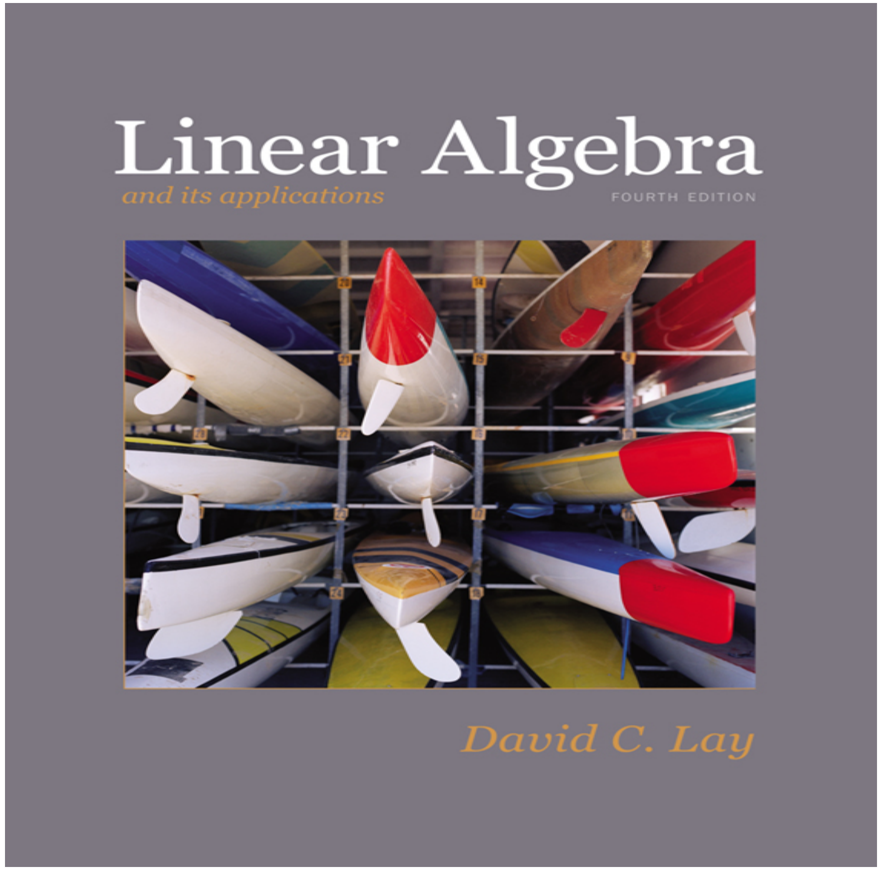
Symmetric Matrices and Singular Value Decomposition



Overview

- Example
- Symmetric Matrices and their properties
- Singular Value Decomposition

Singular Value Decomposition (SVD)

- Let A be an $m \times n$ matrix. Then singular value decomposition of A is **A = U ∑ V** ^{T}

- In the "thin/economy version" of SVD:
  - U is an mxr **orthogonal** matrix and is called the **left singular vectors.**
  - ∑ is rxr and **diagonal** and contains **singular values** of A in descending order, $\sigma_1 \geq \sigma_2 \geq \cdot\cdot\cdot \geq \sigma_r > 0$
  - V is an nxr **orthogonal** matrix and is called the **right singular vectors.**

A Conceptual Example

- An audiobook company likes to make recommendations to users based on the their ratings of other books.
- A: mxn where m rows denote the users and n columns denote the book titles.
- The first 3 columns include titles in historical drama genre and the last 2 column include title in Sci-Fi genre.
- A can be thought of as “user to book” mapping.

	Poisonwood Bible	**The Revenant**	**Ines of my Soul**	**The Martian**	**Lord of the Rings**
Susan	0	1	0	2	2
Vu	0	0	0	5	5
Joe	0	2	0	4	4
Rafa	5	5	5	0	0
Jose	4	4	4	0	0
Ana	3	3	3	0	0
Mike	1	1	1	0	0

	Poisonwood Bible	The Revenant	Ines of my Soul	The Martian	Lord of the Rings
Susan	0	1	0	2	2
Vu	0	0	0	5	5
Joe	0	2	0	4	4
Rafa	5	5	5	0	0
Jose	4	4	4	0	0
Ana	3	3	3	0	0
Mike	1	1	1	0	0

Historical Drama **Sci-Fi**

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

U: rows: Users- cols: Concepts

V: rows: Books-cols: Concepts

V$^{\mathrm{T}}$: rows: concepts-cols: Books

$$= \begin{bmatrix} 0.07 & -0.29 & 0.33 \\ 0.07 & -0.73 & -0.68 \\ 0.15 & -0.60 & 0.65 \\ 0.68 & 0.12 & -0.05 \\ 0.55 & 0.09 & -0.04 \\ 0.41 & 0.07 & -0.03 \\ 0.14 & 0.02 & -0.01 \end{bmatrix} \begin{bmatrix} 12.5 & 0 & 0 \\ 0 & 9.5 & 0 \\ 0 & 0 & 1.3 \end{bmatrix} \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.13 & -0.03 & 0.13 & -0.69 & -0.69 \\ 0.41 & -0.80 & 0.41 & 0.09 & 0.09 \end{bmatrix}$$

User to Concept Mapping **Concept Strength** **Concept to Book Mapping**

We factored the matrix, so what?!

- The goal is to make audiobook recommendations. (say to user S_1)
- e.g. Find all people who liked Poisonwood Bible.(**q**=[5 0 0 0 0]) and recommend this book to people who're similar to people in this category but haven't read this book yet.
- Should map this book to the concept space. (why?)
- e.g. if user S_1 has rated her books as follows: $\boldsymbol{S_1} =$ [0 3 5 0 0 0], she has nothing in common with q, but if use the concept space things will be more meaningful.

Example

- Map (project) q to the concept space.

- $q * v = [5\ 0\ 0\ 0\ 0] \begin{bmatrix} 0.56 & 0.13 \\ 0.59 & -0.03 \\ 0.56 & 0.13 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix}$=[-2.8 -0.63]

- $s_1 * v = [0\ 3\ 5\ 0\ 0] \begin{bmatrix} 0.56 & 0.13 \\ 0.59 & -0.03 \\ 0.56 & 0.13 \\ 0.09 & -0.69 \\ 0.09 & -0.69 \end{bmatrix}$=[-4.9 -0.55]

SYMMETRIC MATRIX

- A **symmetric matrix** is a matrix A such that $A^T = A$.
- Such a matrix is necessarily square.
- Its main diagonal entries are arbitrary, but its other entries occur in pairs—on opposite sides of the main diagonal.

Symmetric: $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$, $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & 8 \\ 0 & 8 & -7 \end{bmatrix}$, $\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$

Nonsymmetric: $\begin{bmatrix} 1 & -3 \\ 3 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & -4 & 0 \\ -6 & 1 & -4 \\ 0 & -6 & 1 \end{bmatrix}$, $\begin{bmatrix} 5 & 4 & 3 & 2 \\ 4 & 3 & 2 & 1 \\ 3 & 2 & 1 & 0 \end{bmatrix}$

Example

If possible, diagonalize the matrix $A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$.

```
> A = matrix(c(6,-2,-1,-2,6,-1,-1,-1,5), nrow = 3)
> A
     [,1] [,2] [,3]
[1,]    6   -2   -1
[2,]   -2    6   -1
[3,]   -1   -1    5
```

```
> e = eigen(A)
> e$values
[1] 8 6 3
> round(e$vectors, digits=3)
       [,1]   [,2]   [,3]
[1,]  0.707 -0.408 -0.577
[2,] -0.707 -0.408 -0.577
[3,]  0.000  0.816 -0.577
```

```
> library(MASS)
> e$vectors%*%diag(e$values)%*%ginv(e$vectors)
     [,1] [,2] [,3]
[1,]    6   -2   -1
[2,]   -2    6   -1
[3,]   -1   -1    5
```

Since A has distinct eigenvalues, there is sufficient condition for A to be diagonalizable: $A = PDP^{-1}$

Example – cont.

- Find the dot products of eigenvectors of A.

```
> round(dot(e$vectors[,1],e$vectors[,2]), 3)
[1] 0
> round(dot(e$vectors[,1],e$vectors[,3]),3)
[1] 0
> round(dot(e$vectors[,2],e$vectors[,3]),3)
[1] 0
```

```
> round(t(e$vectors)%*%e$vectors,3)
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

The Eigenvectors are orthonormal

- Compare transpose of P with its inverse.

```
> round(t(e$vectors), digits=3)
       [,1]   [,2]   [,3]
[1,]  0.707 -0.707  0.000
[2,] -0.408 -0.408  0.816
[3,] -0.577 -0.577 -0.577
```

```
> round(ginv(e$vectors), digits = 3)
       [,1]   [,2]   [,3]
[1,]  0.707 -0.707  0.000
[2,] -0.408 -0.408  0.816
[3,] -0.577 -0.577 -0.577
```

The inverse and transpose are the same.

ORTHONORMAL SETS

- **Theorem:** If a matrix U has orthonormal columns then $U^T = U^{-1}$, i.e. $U^TU = I$

- **Proof:** To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$.
- Let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ and compute

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}$$

- The entries in the matrix at the right are inner products, using transpose notation.
- The columns of U are orthogonal => $\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0$ $\quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0$ $\quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0$
- The columns of U all have unit length => $\mathbf{u}_1^T\mathbf{u}_1 = 1$ $\quad \mathbf{u}_2^T\mathbf{u}_2 = 1$ $\quad \mathbf{u}_3^T\mathbf{u}_3 = 1$
- Therefore, $U^TU = I$

Theorem

- **An nxn matrix A is orthogonally diagonalizable if and only if A is symmetric matrix.**
- First we prove if A is orthogonally diagonalizable then A is symmetric:
- A is diagonalizable => $A = PDP^{-1}$
- P is orthogonal => $A=PDP^{T}$
- $A=PDP^{T}$ => $A^{T} = (PDP^{T})^{T}$
- $A^{T} = P^{TT} D^{T} P^{T} = PDP^{T} = A$
- A is symmetric.

Theorem

- For the second part we need to prove if A is symmetric then it is **orthogonally** diagonalizable. This proof is much harder, we'll prove it for the case that A has distinct eigenvalues.
- Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2.
- $\lambda_1 v_1 . v_2 = (\lambda_1 v_1)^T v_2 = (A v_1)^T \ v_2$ Since $\mathbf{v}_1$ is an eigenvector
- $= {v_1}^T A^T v_2$
- $= {v_1}^T A v_2$ Since $A^T = A$
- $= {v_1}^T \lambda_2 v_2 = \lambda_2 {v_1}^T v_2 = \lambda_2 v_1 . v_2$ Since $\mathbf{v}_2$ is an eigenvector
- Combing the first line and last line results in $\lambda_1 v_1 . v_2 - \lambda_2 v_1 . v_2 = 0$
- $(\lambda_1 - \lambda_2) v_1 . v_2 = 0 \Rightarrow v_1 . v_2 = 0$

Spectral Decomposition

- Using the column–row expansion of the product A=PDP ^{T}, we can write

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

- This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A.
- This representation helps with having an expression for matrix (A) approximation.

Proof for Spectral Decomposition

- Recall:

Column–Row Expansion of *AB*

If A is $m \times n$ and B is $n \times p$, then

$$AB = [\,\mathrm{col}_1(A) \quad \mathrm{col}_2(A) \quad \cdots \quad \mathrm{col}_n(A)\,] \begin{bmatrix} \mathrm{row}_1(B) \\ \mathrm{row}_2(B) \\ \vdots \\ \mathrm{row}_n(B) \end{bmatrix}$$

$$= \mathrm{col}_1(A)\,\mathrm{row}_1(B) + \cdots + \mathrm{col}_n(A)\,\mathrm{row}_n(B)$$

Proof for Spectral Decomposition – cont.

EXAMPLE 4 Let $A = \begin{bmatrix} -3 & 1 & 2 \\ 1 & -4 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$. Verify that

$$AB = \mathrm{col}_1(A)\,\mathrm{row}_1(B) + \mathrm{col}_2(A)\,\mathrm{row}_2(B) + \mathrm{col}_3(A)\,\mathrm{row}_3(B)$$

$$\mathrm{col}_1(A)\,\mathrm{row}_1(B) = \begin{bmatrix} -3 \\ 1 \end{bmatrix}\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} -3a & -3b \\ a & b \end{bmatrix}$$

$$\mathrm{col}_2(A)\,\mathrm{row}_2(B) = \begin{bmatrix} 1 \\ -4 \end{bmatrix}\begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} c & d \\ -4c & -4d \end{bmatrix}$$

$$\mathrm{col}_3(A)\,\mathrm{row}_3(B) = \begin{bmatrix} 2 \\ 5 \end{bmatrix}\begin{bmatrix} e & f \end{bmatrix} = \begin{bmatrix} 2e & 2f \\ 5e & 5f \end{bmatrix}$$

Thus

$$\sum_{k=1}^{3} \mathrm{col}_k(A)\,\mathrm{row}_k(B) = \begin{bmatrix} -3a + c + 2e & -3b + d + 2f \\ a - 4c + 5e & b - 4d + 5f \end{bmatrix}$$

This matrix is obviously AB. Notice that the $(1, 1)$-entry in AB is the sum of the $(1, 1)$-entries in the three outer products, the $(1, 2)$-entry in AB is the sum of the $(1, 2)$-entries in the three outer products, and so on. ■

Proof for Spectral Decomposition – cont.

$$A = PDP^T \qquad P = \begin{bmatrix} \uparrow & \uparrow \\ u_1 & u_2 \\ \downarrow & \downarrow \end{bmatrix}_{2\times 2} \qquad D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}_{2\times 2} \qquad P^T = \begin{bmatrix} \leftarrow u_1^T \rightarrow \\ \leftarrow u_2^T \rightarrow \end{bmatrix}_{2\times 2}$$

$$PD = \begin{bmatrix} u_1 \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \end{bmatrix} + \begin{bmatrix} u_2 \updownarrow \end{bmatrix} \begin{bmatrix} \lambda_2 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 u_1 \updownarrow & \begin{matrix} 0 \\ 0 \end{matrix} \end{bmatrix} + \begin{bmatrix} \begin{matrix} 0 \\ 0 \end{matrix} & \lambda_2 u_2 \updownarrow \end{bmatrix} = \begin{bmatrix} \lambda_1 u_1 \updownarrow & \lambda_2 u_2 \updownarrow \end{bmatrix}_{2\times 2} \quad \star$$

$$PDP^T = \begin{bmatrix} \lambda_1 u_1 \updownarrow & \lambda_2 u_2 \updownarrow \end{bmatrix} \begin{bmatrix} \leftarrow u_1^T \rightarrow \\ \leftarrow u_2^T \rightarrow \end{bmatrix} = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T$$

For an $n \times n$ A $\Rightarrow$ $A = \lambda_1 u_1 u_1^T + \cdots + \lambda_n u_n u_n^T$

Example

- Construct a spectral decomposition of the matrix A that has the following orthogonal diagonalization and verify your results.

$$A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

Solution

SOLUTION Denote the columns of P by $\mathbf{u}_1$ and $\mathbf{u}_2$. Then

$$A = 8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T$$

To verify this decomposition of A, compute

$$\mathbf{u}_1\mathbf{u}_1^T = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$$

$$\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$$

and

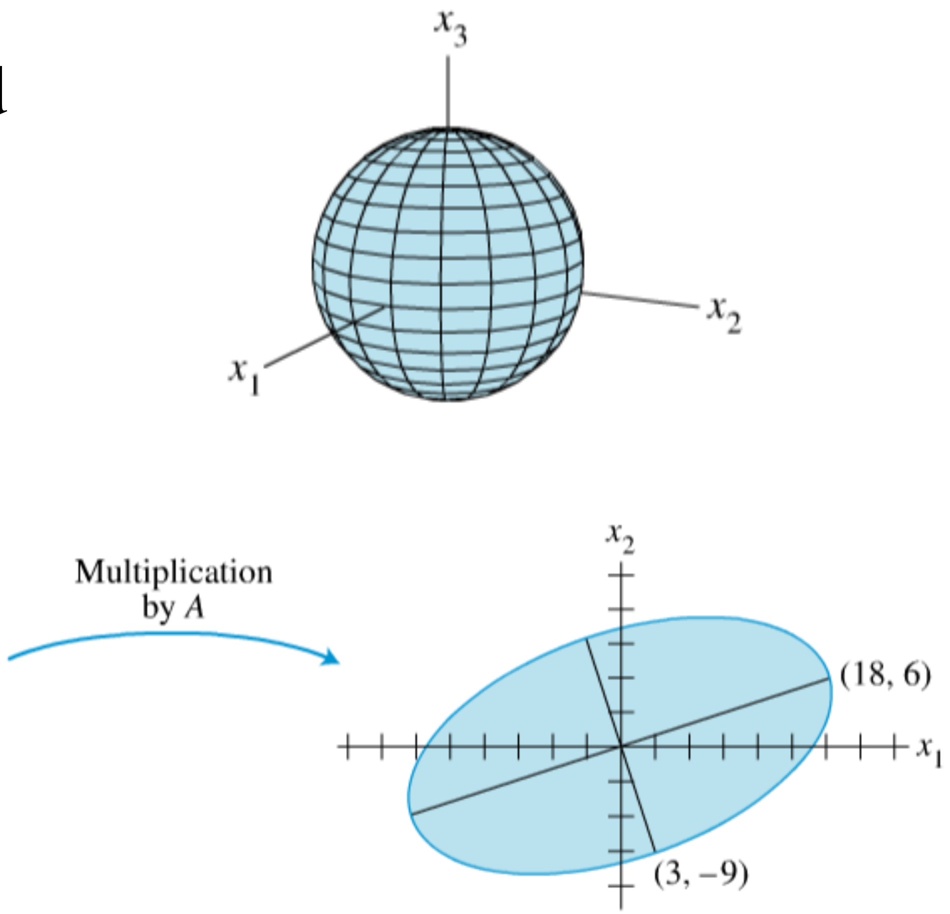
$$8\mathbf{u}_1\mathbf{u}_1^T + 3\mathbf{u}_2\mathbf{u}_2^T = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = A$$

Diagonalization of Nonsquare Matrices

- Diagonalization theorem plays a part in many interesting applications.
- Unfortunately, not all matrices can be factored as $A = P D P^{-1}$ with D diagonal.
- However, a factorization $\mathbf{A} = U \Sigma V^T$ is possible for any $m \times n$ matrix A.
- A special factorization of this type, called the singular value decomposition, is one of the most useful matrix factorizations in applied linear algebra.

Example

- Assume x is the unit sphere $\{x: \|x\| = 1\}$ in R^3 and $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.
- The linear transformation $x \mapsto Ax$ maps the unit sphere onto an ellipse in R^2.
- $Ax = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4x_1 + 11x_2 + 14x_3 \\ 8x_1 + 7x_2 - 2x_3 \end{bmatrix}$.
- Find a unit vector x at which the length $\|Ax\|$ is maximized, and compute this maximum length.



Solution

- The quantity $\| A x \|^2$ is maximized at the same x that maximizes $\| A x \|$, and $\| A x \|^2$ is easier to study.
- $||Ax||^2 = (Ax)^T(Ax) = x^T A^T (Ax) = x^T (A^T A) x$
- Note that A^TA is symmetric.
- So the problem now is to maximize the quadratic form $x^T (A^T A) x$ subject to the constraint $\| x \| = 1$
- We will apply the following theorem that states:

> Let **B** be a symmetric matrix, and define m and M as in (2). Then M is the greatest eigenvalue λ_1 of **B** and m is the least eigenvalue of **B**. The value of $\mathbf{x}^T\mathbf{B}\mathbf{x}$ is M when $\mathbf{x}$ is a unit eigenvector $\mathbf{u}_1$ corresponding to M. The value of $\mathbf{x}^T\mathbf{B}\mathbf{x}$ is m when $\mathbf{x}$ is a unit eigenvector corresponding to m.

- The B in the above theorem would be A^TA in this problem. Therefore, we need to find eigenvector and eigenvalues of A^TA.

Solution-cont.

```
> A = matrix(c(4, 8, 11, 7 , 14, -2), nrow = 2)
> A
     [,1] [,2] [,3]
[1,]    4   11   14
[2,]    8    7   -2
```

```
> Cx=t(A)%*%A
> Cx
     [,1] [,2] [,3]
[1,]   80  100   40
[2,]  100  170  140
[3,]   40  140  200
```

```
> e=eigen(Cx)
> e
$values
[1]  3.600000e+02  9.000000e+01 -6.064117e-15

$vectors
           [,1]       [,2]       [,3]
[1,] -0.3333333 -0.6666667  0.6666667
[2,] -0.6666667 -0.3333333 -0.6666667
[3,] -0.6666667  0.6666667  0.3333333
```

```
> Ax = A%*%e$vectors[,1]
> Ax
     [,1]
[1,]  -18
[2,]   -6
> norm(Ax,'2')
[1] 18.97367
```

- So vector [-1/3 -2/3 -2/3] gets mapped to [-18, -6]

Singular Values of an mxn Matrix

- We start by finding the maximum value of ‖A v‖ where v is the eigenvector of A^TA.
- $||Av||^2 = (Av)^T(Av)$
- $= v^TA^T(Av) = v^T(A^TAv)$
- $= v^T\ (\lambda v) = \lambda v^Tv$
- v is a unit vector => $||Av||^2 = \lambda$
- So the eigenvalues of A^TA are all nonnegative.
- The singular values of A are the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \ldots, \sigma_n$.
- They are lengths of the vectors $Av_1, \ldots, Av_n$
- They are arranged in decreasing order.

Example

In previous example where $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, find the singular values of A and the length of the major and minor semiaxes of the ellipse.

```
> e
$values
[1]  3.600000e+02  9.000000e+01 -6.064117e-15

$vectors
           [,1]       [,2]       [,3]
[1,] -0.3333333 -0.6666667  0.6666667
[2,] -0.6666667 -0.3333333 -0.6666667
[3,] -0.6666667  0.6666667  0.3333333
```

```
> sigma = sqrt(round(e$values,3))
> sigma
[1] 18.973666  9.486833  0.000000

> majorAxes = norm(A%*%e$vector[,1],'2')
> majorAxes
[1] 18.97367

> minorAxes = norm(A%*%e$vector[,2],'2')
> minorAxes
[1] 9.486833
```

Theorem: The Singular Value Decomposition

- Let A be an m × n matrix. Then A can be written as **A = U ∑ V** ^{T}
- V is an nxn orthogonal matrix that contains the eigenvectors A^TA.

$$V = \begin{bmatrix} \uparrow & & \uparrow \\ v_1 & \cdots & v_n \\ \downarrow & & \downarrow \end{bmatrix}$$

- ∑ is mxn and diagonal (for which the diagonal entries in D are the first r singular values of A, σ 1 ≥ σ 2 ≥ · · · ≥ σ r > 0)

$$\Sigma = \begin{array}{c} \begin{array}{cc} \overbrace{\quad}^{r} & \overbrace{\quad}^{n-r} \end{array} \\ \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \}r \\ \}m-r \end{array} \end{array} \quad \text{where} \quad D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}$$

Ex

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

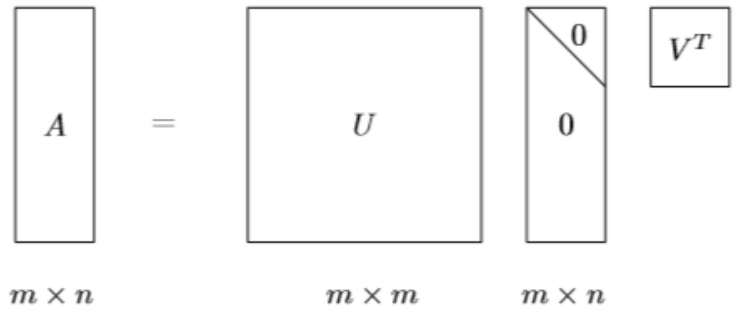
- U is an mxm orthogonal matrix where

$$U = \begin{bmatrix} \uparrow & & \uparrow \\ u_1 & \cdots & u_m \\ \downarrow & & \downarrow \end{bmatrix}$$

$$u_i = \frac{1}{\sigma_i} A v_i$$

Proof

- To prove that $A = U\Sigma V^T$ we'll prove that $AV = U\Sigma$
- We first simplify the left and the right side and show that the two expressions are equivalent.



$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|} A\mathbf{v}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$$

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \qquad (1 \le i \le r)$$

$$AV = [\, A\mathbf{v}_1 \quad \cdots \quad A\mathbf{v}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \,] = [\, \sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \,]$$

$$U\Sigma = [\, \mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m \,] \left[\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right]$$

$$= [\, \sigma_1\mathbf{u}_1 \quad \cdots \quad \sigma_r\mathbf{u}_r \quad \mathbf{0} \quad \cdots \quad \mathbf{0} \,]$$

$$= AV$$

Example

Construct a singular value decomposition of $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$ and verify your result.

Solution

```
> A = matrix(c(4, 8, 11, 7 , 14, -2), nrow = 2)
> Cx=t(A)%*%A
> e=eigen(Cx)
> V = e$vectors
> V
           [,1]       [,2]       [,3]
[1,] -0.3333333 -0.6666667  0.6666667
[2,] -0.6666667 -0.3333333 -0.6666667
[3,] -0.6666667  0.6666667  0.3333333
```

$$V = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

The square roots of the eigenvalues are the singular values:

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

The nonzero singular values are the diagonal entries of D. The matrix Σ is the same size as A, with D in its upper left corner and with 0's elsewhere.

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \qquad \Sigma = [\,D \quad 0\,] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

```
> S = matrix(c(sqrt(e$value[1]),0,0,sqrt(e$value[2]),0,0), nrow = 2)
> S
         [,1]     [,2] [,3]
[1,] 18.97367 0.000000    0
[2,]  0.00000 9.486833    0
```

Solution – cont.

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}}\begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

```
> U=A%*%V[,1:2]%*%diag(1/sqrt(e$values[1:2]))
> U
           [,1]       [,2]
[1,] -0.9486833  0.3162278
[2,] -0.3162278 -0.9486833
```

$$A = \underbrace{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}}_{\uparrow \; U} \underbrace{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}}_{\uparrow \; \Sigma} \underbrace{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}_{\uparrow \; V^T}$$

```
> A2 = U%*%S%*%t(V)
> A2
     [,1] [,2] [,3]
[1,]    4   11   14
[2,]    8    7   -2
```

Applications of SVD

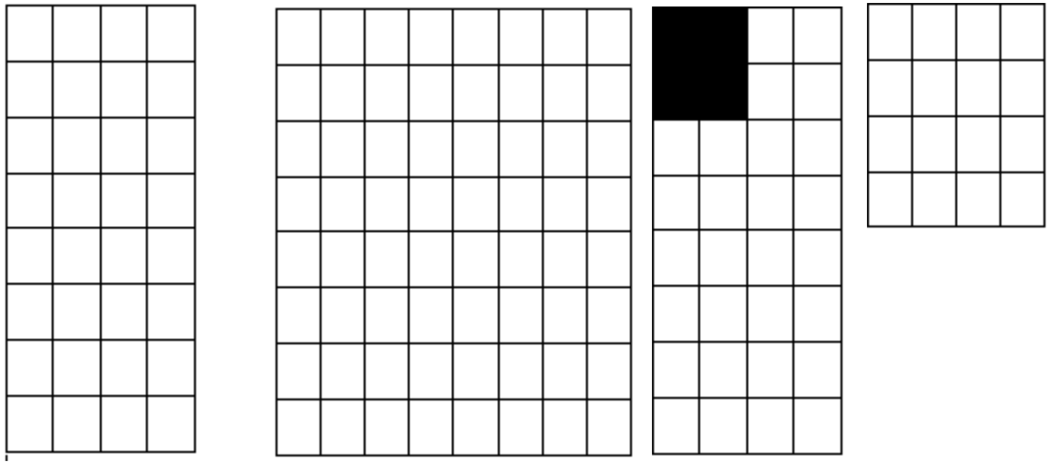
Economy Version of SVD

When Σ contains rows or columns of zeros, a more compact decomposition of A is possible. Partition the matrices as follows:

$$U = [\, U_r \quad U_{m-r} \,], \quad \text{where } U_r = [\, \mathbf{u}_1 \quad \cdots \quad \mathbf{u}_r \,]$$
$$V = [\, V_r \quad V_{n-r} \,], \quad \text{where } V_r = [\, \mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r \,]$$

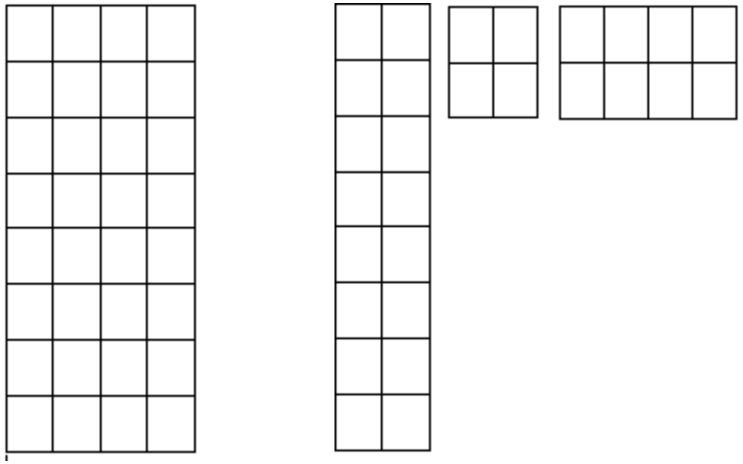
$$A = [\, U_r \quad U_{m-r} \,] \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T$$

Regular SVD



A mxn 8x4 | U mxm 8x8 | Σ mxn 8x4 | V^T nxn 4x4

Economy SVD



A mxn 8x4 | U mxr 8x2 | Σ rxr 2x2 | V^T rxn 2x4

Example

- Find the SVD of the following matrix and verify $U_r D V_r^T$

$$A = \begin{bmatrix} -18 & 13 & -4 & 4 \\ 2 & 19 & -4 & 12 \\ -14 & 11 & -12 & 8 \\ -2 & 21 & 4 & 8 \end{bmatrix}$$

- Note that svd function in R, by default gives you min(m,n) number of left and right singular vectors.

Solution

```
> A = matrix(c(-18,2,-14,-2,13,19,11,21,-4,-4,-12,4,4,12,8,8), nrow = 4)
> A
     [,1] [,2] [,3] [,4]
[1,]  -18   13   -4    4
[2,]    2   19   -4   12
[3,]  -14   11  -12    8
[4,]   -2   21    4    8
 s = svd(A)
> s
$d
[1] 4.000000e+01 2.000000e+01 1.000000e+01 2.162541e-15

$u
     [,1] [,2] [,3] [,4]
[1,] -0.5  0.5  0.5 -0.5
[2,] -0.5 -0.5 -0.5 -0.5
[3,] -0.5  0.5 -0.5  0.5
[4,] -0.5 -0.5  0.5  0.5

$v
     [,1] [,2] [,3] [,4]
[1,]  0.4 -0.8 -0.4  0.2
[2,] -0.8 -0.4  0.2  0.4
[3,]  0.2 -0.4  0.8 -0.4
[4,] -0.4 -0.2 -0.4 -0.8
```

```
> s$u[,1:3]%*%diag(s$d[1:3])%*%t(s$v[,1:3])
     [,1] [,2] [,3] [,4]
[1,]  -18   13   -4    4
[2,]    2   19   -4   12
[3,]  -14   11  -12    8
[4,]   -2   21    4    8
```

Example

Find a singular value decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$

SOLUTION First, compute $A^TA = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. The eigenvalues of A^TA are 18 and 0, with corresponding unit eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

These unit vectors form the columns of V:

$$V = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Example – cont.

The singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$. Since there is only one nonzero singular value, the "matrix" D may be written as a single number. That is, $D = 3\sqrt{2}$. The matrix Σ is the same size as A, with D in its upper left corner:

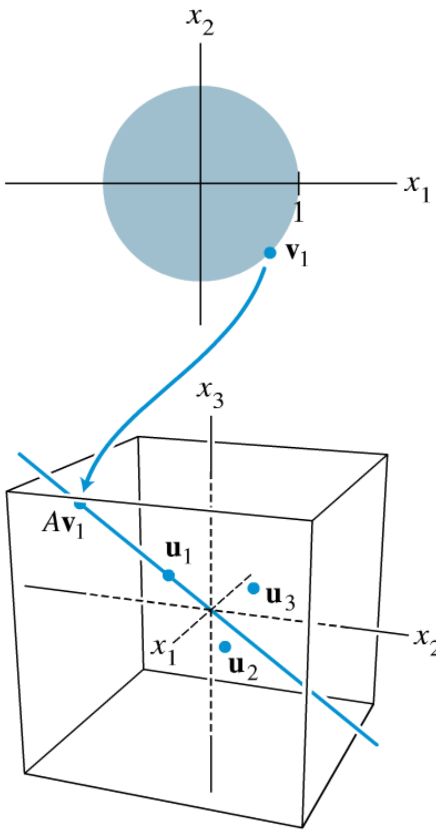
$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

To construct U, first construct $A\mathbf{v}_1$ and $A\mathbf{v}_2$:

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Example – cont.

- The other columns of U are found by extending the set $\{u_1\}$ an orthonormal basis for R^3. In this case, we need to to orthogonal unit vectors that are orthogonal to u_1.
- $u_1^T x = 0$
- $\begin{bmatrix} 2 & -4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2x_1 - 4x_2 + 4x_3 = 0$
- $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \qquad w_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \; w_3 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$



Example – cont.

- u_2 is normalized version of w_2. Note that $w_2 \perp u_1$.
- $u_2 = \begin{bmatrix} 2/sqrt(5) \\ 1/sqrt(5) \\ 0 \end{bmatrix}$
- To find u_3 we need to find a vector that's orthogonal to both u_1 and u_2.
- u_3 is normalized version of $w_3 - P_{u_2}^{w_3} = w_3 - \frac{w_3.u_2}{u_2.u_2} u_2 = \begin{bmatrix} -0.4 \\ 0.8 \\ 1 \end{bmatrix}$
- $u_3 = \begin{bmatrix} -0.4/sqrt(.4^2 + .8^2 + 1) \\ 0.8/sqrt(.4^2 + .8^2 + 1) \\ 1/sqrt(.4^2 + .8^2 + 1) \end{bmatrix}$

Example – cont.

Finally, set $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$, take Σ and V^T from above, and write

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Example – cont.

```
> A = matrix(c(1,-2,2,-1,2,-2), nrow = 3)
> A
     [,1] [,2]
[1,]    1   -1
[2,]   -2    2
[3,]    2   -2
> e = svd(A)
> e
$d
[1] 4.242641 0.000000

$u
           [,1]      [,2]
[1,] -0.3333333 0.6666667
[2,]  0.6666667 0.6666667
[3,] -0.6666667 0.3333333

$v
           [,1]      [,2]
[1,] -0.7071068 0.7071068
[2,]  0.7071068 0.7071068
```

- As you can see you do get the eigenvector corresponding to zero (second columns of V). But this isn't the full version. If you'd like to see the full version you need to set nu = 3.

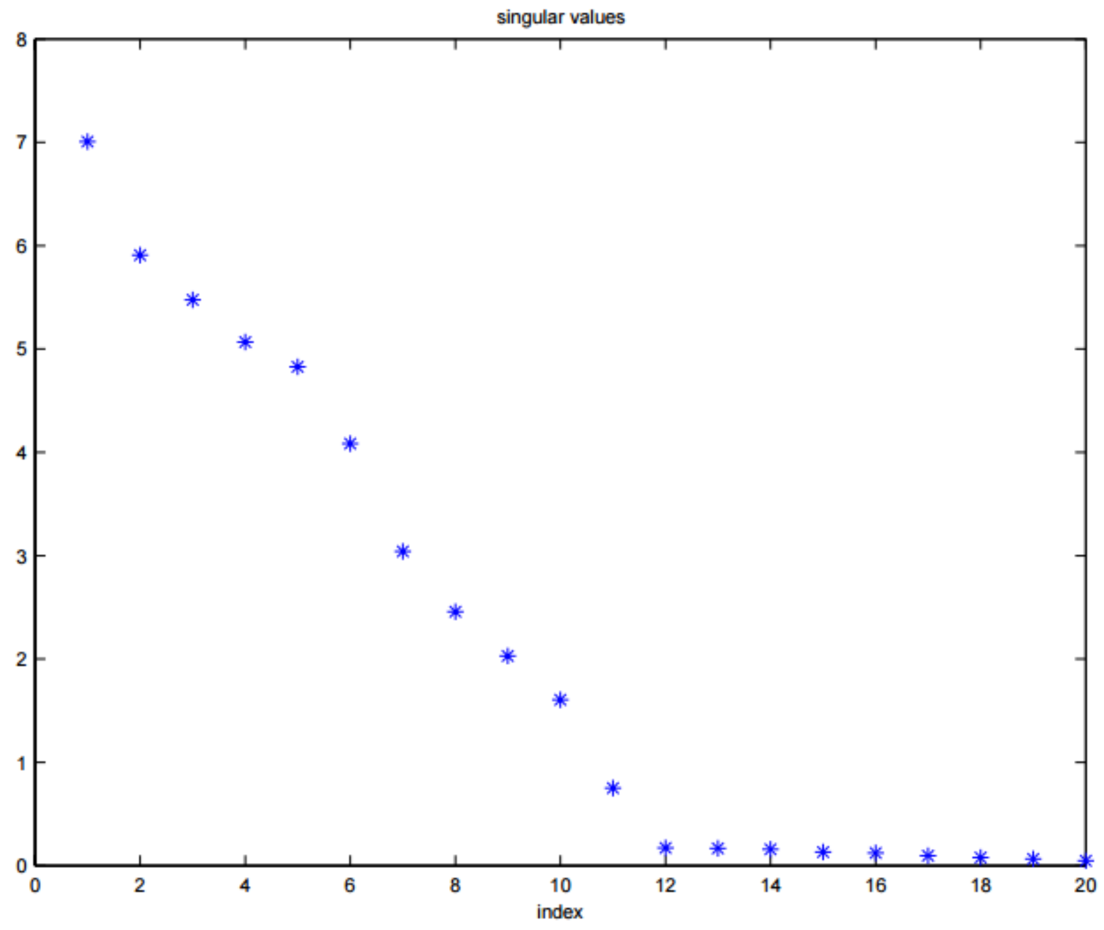
```
> e = svd(A, nu = 3)
> e
$d
[1] 4.242641 0.000000

$u
           [,1]      [,2]       [,3]
[1,] -0.3333333 0.6666667 -0.6666667
[2,]  0.6666667 0.6666667  0.3333333
[3,] -0.6666667 0.3333333  0.6666667

$v
           [,1]      [,2]
[1,] -0.7071068 0.7071068
[2,]  0.7071068 0.7071068
```

Removing Noise Using SVD

- Assume that A is contain some data matrix plus noise: A = A0 + N, where the noise N is small compared with A0.
- Then typically the singular values of A have the behavior illustrated in this figure.
- We can remove the noise by truncating the singular value expansion.



Matrix Approximation Using SVD

- Recall that a symmetric matrix A=PDP ^{T} could be expanded to its spectral decomposition as follows:

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$$

- Any m by n matrix $\mathbf{A} = \mathbf{U} \sum \mathbf{V}^{\mathbf{T}}$ can be expressed as

$$A = \sigma_1 U_1 V_1^T + \cdots + \sigma_n U_n V_n^T$$

- The truncated SVD is very important, not only for removing noise but also for compressing data and approximating a given matrix by one of lower rank.
- The difference (L2 norm) between the original and approximation is given by (k+1)th singular values.

$$\mathbf{A}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^T \qquad \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}$$

Example

- Use the first 2 singular values of matrix A to find an approximation for A. Determine the error between A and A_2.

```
> A
     [,1] [,2] [,3] [,4]
[1,]   16    2    3   13
[2,]    5   11   10    8
[3,]    9    7    6   12
[4,]    4   14   15    1
```

Solution

```
> A = magic(4)
> A
     [,1] [,2] [,3] [,4]
[1,]   16    2    3   13
[2,]    5   11   10    8
[3,]    9    7    6   12
[4,]    4   14   15    1

> s=svd(A)
> s
$d
[1] 3.400000e+01 1.788854e+01 4.472136e+00 4.172807e-16

$u
     [,1]       [,2] [,3]       [,4]
[1,] -0.5  0.6708204  0.5 -0.2236068
[2,] -0.5 -0.2236068 -0.5 -0.6708204
[3,] -0.5  0.2236068 -0.5  0.6708204
[4,] -0.5 -0.6708204  0.5  0.2236068

$v
     [,1] [,2]       [,3]       [,4]
[1,] -0.5  0.5  0.6708204  0.2236068
[2,] -0.5 -0.5 -0.2236068  0.6708204
[3,] -0.5 -0.5  0.2236068 -0.6708204
[4,] -0.5  0.5 -0.6708204 -0.2236068
```

```
> A2=s$u[,1:2]%*%diag(s$d[1:2])%*%t(s$v[,1:2])
> A2
     [,1] [,2] [,3] [,4]
[1,] 14.5  2.5  2.5 14.5
[2,]  6.5 10.5 10.5  6.5
[3,] 10.5  6.5  6.5 10.5
[4,]  2.5 14.5 14.5  2.5
```

```
> rootSSE=norm(A-A2,'2')
> rootSSE
[1] 4.472136
```

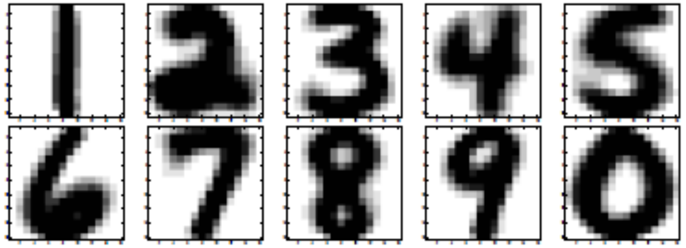
Note that error matches the 3rd singular value.

Bases for Fundamental Subspace

- Given an SVD for an $m \times n$ matrix A, let $u_1, \dots, u_m$ be the left singular vectors, $v_1, \dots, v_n$ the right singular vectors, and $\sigma_1, \dots, \sigma_n$ the singular values, and let r be the rank of A.
- $\{u_1, \dots, u_r\}$ is an orthonormal basis for Col A.
- $\{v_1, \dots, v_r\}$ is an orthonormal basis for Row A. (We will use this property in the digit recognition assignment)

Classification of Handwritten Digits

- Problem: Given a set of manually classified digits (the training set), classify a set of unknown digits (the test set) using SVD.
- We will be using the US postal Service database that contains 1707 training and 2007 test digits.
- Each image is a grayscale 16x16 image.
- Stacking all the columns of each image matrix above each other gives a vector of size 256×1.
- Transpose above so every row is



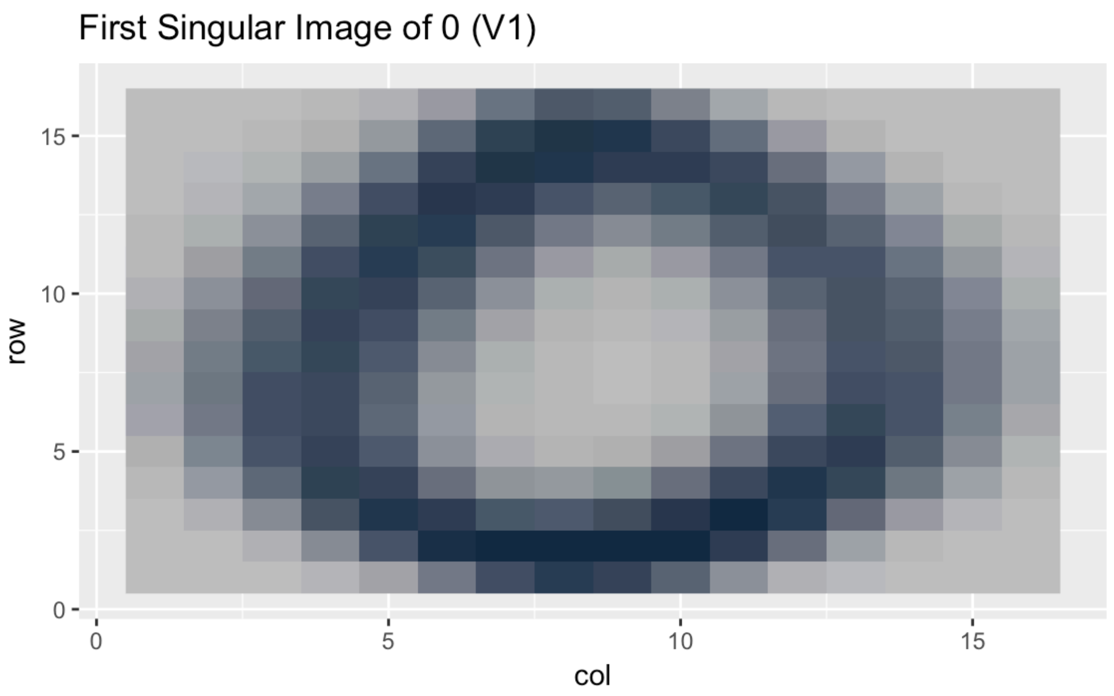
an image and each column represent pixel values of that image.

Data Files

- After transposing:
- The training images are stored in **trainInput.** (1707x256).
- The correct digit corresponding to each column of trainInput is stored in **trainOuput.** (1x1707).
- The test images are stored in **testInput.** (2007x256).
- The correct digit corresponding to each column of testInput is stored in **testOutput.** (1x2007).

Methodology

- Organize your data such that each digit is represented by a matrix. (x rows, 256 columns). You'll have 10 such matrices.
- Let each row in A represents an image of a digit (for example 0).
- Therefore the right singular vectors v_i are an orthogonal basis for row space of A.
- We will refer to the right singular vectors as "singular images." (We can interpret each v as an image patch just by reforming it into a 16x16 patch.)
- For the training set of digits, compute the SVD of each set of digits of one kind. We get 10 sets of singular "images", one for each digit.



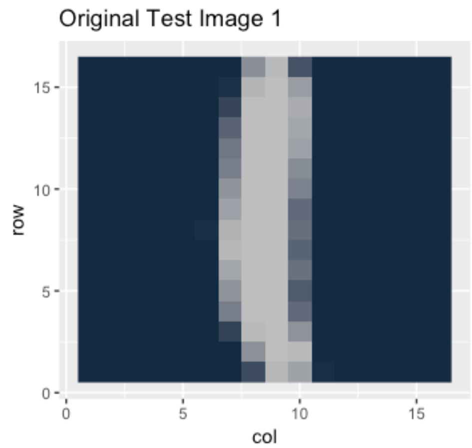
Methodology

- The idea is that each digit is well characterized by a few of the first "singular images" of its own kind; but is NOT characterized very well by a few of the first singular images of some other digit.
- We will express each unknown test image in terms of the first 20 singular images of each digit. i.e. testImage = Vx (think of least square system b = Ax)
- Classify each unknown image to be the digit corresponding to the smallest error. i.e. min|| testImage – Vx||

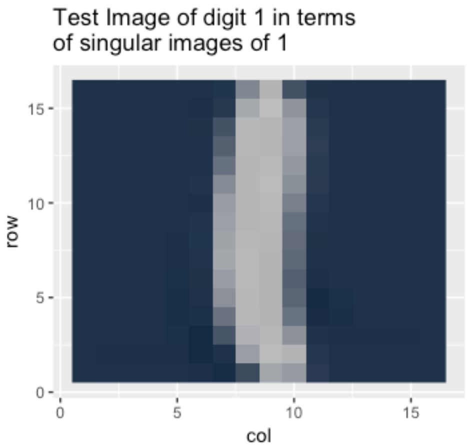
A Test Image of Digit 1 as a Linear Combination of Singular Images of Different Digits

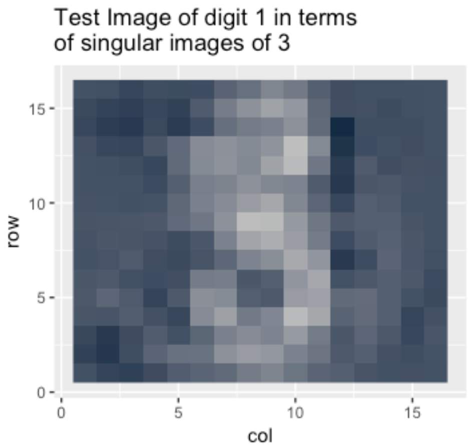
- $testImage \approx Vx = x_1V_1 + x_2V_2 + \cdots + x_{20}V_{20}$

We're describing each image as a coefficient times each singular image V1-V20

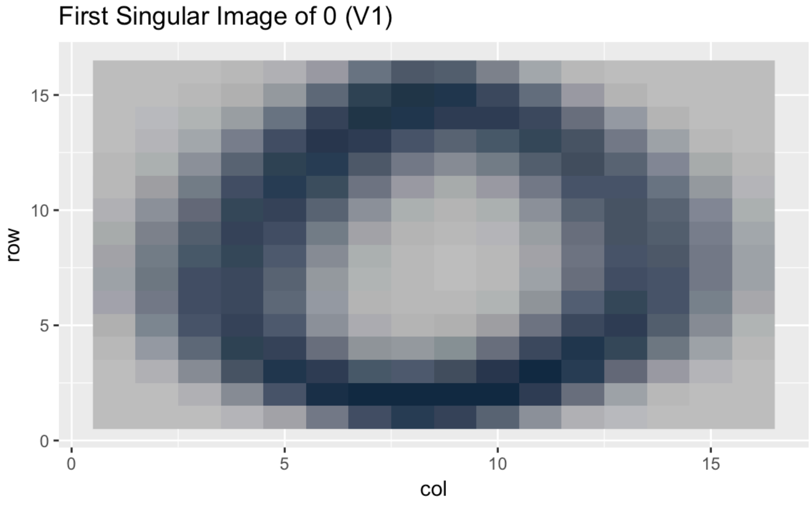


The idea is that each digit is well characterized by a few of the first "singular images" of its own kind; but is NOT characterized very well by a few of the first singular images of some other digit

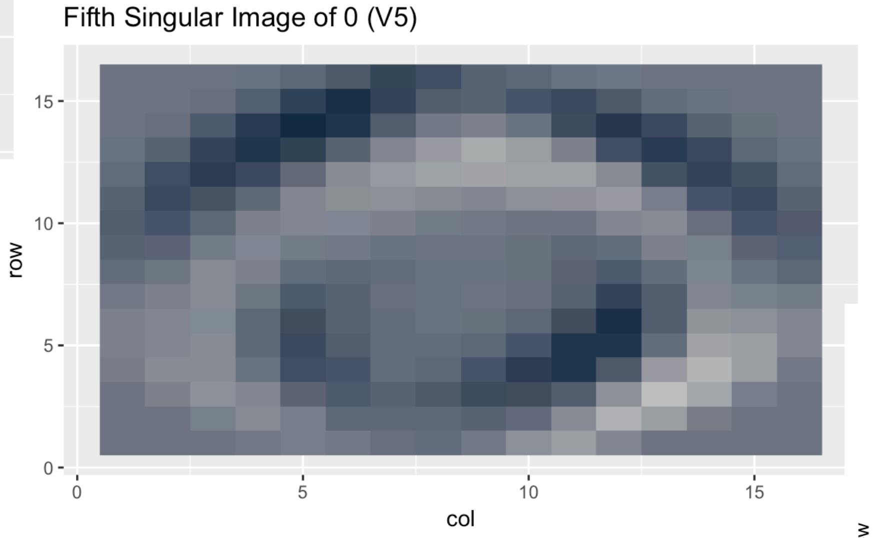


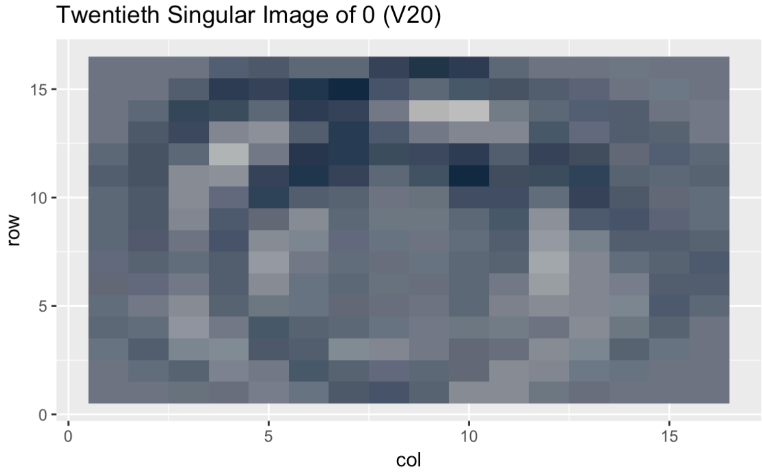


Singular Images



- Note that details of each image is more visible in first singular images.





Condition Number

- The solution to the equation Ax=b can be highly unstable depending on "characteristics" of matrix A.
- i.e. if matrix A is **ill-conditioned** small perturbation in A will result in large errors.
- Ill-conditioned matrix is an invertible matrix that can become singular if some of its entries are changed ever so slightly.
- Condition number of a matrix A is defined as:

$$\kappa = \boldsymbol{\sigma}_{\mathbf{max}} / \boldsymbol{\sigma}_{\mathbf{min}}$$

- Large condition number means that your matrix is ill conditioned.
- Any possible instabilities in numerical calculations are identified in Σ .

Interpreting Condition Number

- The larger the condition number, the closer the matrix is to being singular.
- The identity matrix has a condition number of 1.
- A singular matrix has a condition number of infinity.
- In extreme cases, the software may not be able to distinguish between a singular matrix and an ill-conditioned matrix.

Example

- This example will show how the condition number affects the sensitivity of a solution of A x = b to changes (or errors) in the entries of A.
- Suppose an experiment leads to the following system of equations:

$$
\begin{aligned}
4.5x_1 + 3.1x_2 &= 19.249 \\
1.6x_1 + 1.1x_2 &= 6.843
\end{aligned}
$$

- Find the solution for the system above.

Solution

```
> A = matrix(c(4.5, 1.6, 3.1, 1.1), nrow = 2)
> A
     [,1] [,2]
[1,]  4.5  3.1
[2,]  1.6  1.1
> b = c(19.249,6.843)
> library(MASS)
> x = ginv(A)%*%b
> x
     [,1]
[1,] 3.94
[2,] 0.49
```

Example – cont.

- Now, assume that data in the right hand side is rounded to 2 decimal points.

$$
\begin{aligned}
4.5x_1 + 3.1x_2 &= 19.25 \\
1.6x_1 + 1.1x_2 &= \ \ 6.84
\end{aligned}
$$

- The entries differ less than 0.05%.
- You may expect the solution will be very close to the previous case:

```
> x
     [,1]
[1,] 3.94
[2,] 0.49
```

Solution

```
> b2 = c(19.25, 6.84)
> x2 = ginv(A)%*%b2
> x2
     [,1]
[1,]  2.9
[2,]  2.0
```

```
> x
     [,1]
[1,] 3.94
[2,] 0.49
```

- The answer varies by more than 300%!
- What is the condition number?

```
> e = svd(A)
> e
$d
[1] 5.799137611 0.001724394

$u
           [,1]       [,2]
[1,] -0.9422833 -0.3348167
[2,] -0.3348167  0.9422833

$v
           [,1]       [,2]
[1,] -0.8235675  0.5672182
[2,] -0.5672182 -0.8235675
```

```
> e$d[1]/e$d[2]
[1] 3363
```

Quantifying the Condition Number

- Given $Ax = b$, if the entries of A and b are accurate to about **r** significant digits and if the condition number of A is approximately $\mathbf{10^k}$ then the computed solution $Ax = b$ should be accurate to at least **r-k** significant digits.
- In the next examples, we will show how to use the condition number of a matrix A to estimate the accuracy of a computed solution of $A x = b$.

Example

- Find the condition number of $\begin{matrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{matrix}$
- Considering that R is accurate to about 16 significant digits, how accurate do you predict your solution to be?

```
> A = matrix(c(4,-6,7,-1,0,1,-5,2,-7,11,10,3,-7,9,19,-1), nrow = 4)
> A
     [,1] [,2] [,3] [,4]
[1,]    4    0   -7   -7
[2,]   -6    1   11    9
[3,]    7   -5   10   19
[4,]   -1    2    3   -1
```

```
> e = svd(A)
> head(e$d,1)/tail(e$d,1)
[1] 23682.93
```

- $4 < k < 5$, $r = 16 \Rightarrow 16 - 4 = 12$ and $16 - 5 = 11$
- We predict that answer should be at least 11 (to 12) digits accurate. i.e. at most (4 to) 5 significant digits are lost.

Example

- Construct a random vector x in R^4 and compute $b = A x$.
- Use R to compute the solution x_1 of $A x = b$
- To how many digits x and x_1 agree?

```
> set.seed(1)
> x = runif(4)
> sprintf( "%.16f", x)
[1] "0.2655086631421000" "0.3721238996367902" "0.5728533633518964" "0.9082077899947762"


> b = A%*%x


> sprintf("%.16f", ginv(A)%*%b)
[1] "0.2655086631420023" "0.3721238996345164" "0.5728533633529338" "0.9082077899937531"
```

- x and x_1 agree to 11 to 12 significant digits.

Example

- Repeat the previous example with the following matrix:

4	0	-7	4
-6	1	11	3
7	-5	10	2
-1	2	3	1

- Based on condition number, how accurate do you predict your solution to be?

```
> A = matrix(c(4,-6,7,-1,0,1,-5,2,-7,11,10,3,4,3,2,1), nrow = 4)
> A
     [,1] [,2] [,3] [,4]
[1,]    4    0   -7    4
[2,]   -6    1   11    3
[3,]    7   -5   10    2
[4,]   -1    2    3    1
> head(e$d,1)/tail(e$d,1)
[1] 10.09118
```

- We predict the solution to be about 15 significant digits accurate.

Example

- Construct a random vector x in R^4 and compute $b = A x$.
- Use R to compute the solution x_1 of $A x = b$
- To how many digits x and x_1 agree?

```
> set.seed(1)
> x = runif(4)
> sprintf( "%.16f", x)
[1] "0.2655086631421000" "0.3721238996367902" "0.5728533633518964" "0.9082077899947762"
>
>
> b = A%*%x
> sprintf("%.16f", ginv(A)%*%b)
[1] "0.2655086631421002" "0.3721238996367904" "0.5728533633518962" "0.9082077899947764"
```

- x and x_1 agree to 15 significant digits.

Relationship between SVD and PCA

- Assume A is an m by n centered matrix where rows denoted the samples and columns denote the features.
- Recall that from PCA that we can find covariance matrix by A^TA which is a symmetric matrix and can be orthogonally diagonalized.
- $\boldsymbol{A^T A = PDP^T}$ where columns of P are orthonormal eigenvectors of covariance matrix and D is a diagonal matrix of its eigenvalues.
- Also recall the new coordinates (principal components) will be calculated by finding the projection of A onto the principal directions: $\boldsymbol{Y = AP}$

Relationship between SVD and PCA

- Decompose A into its singular values/vectors: $A = U\Sigma V^T$
- $A^T A = (U\Sigma V^T)^T U\Sigma V^T$
- $= V\Sigma^T U^T U\Sigma V^T$
- $= V\Sigma^T I\Sigma V^T = VDV^T$ => $\boldsymbol{A^T A = VDV^T}$
- Compare this result with the result from pervious slide:
- $\boldsymbol{A^T A = PDP^T}$
- You can conclude that P=V and **hence columns of V are the principal directions.**
- **Also singular values are related to the eigenvalues of covariance matrix via $\lambda_i=\sigma^2_i$**

Relationship between SVD and PCA

- Right multiply both sides of $A = U\Sigma V^T$ by V:
- $\boldsymbol{AV = U\Sigma}$
- Compare this with previous PCA derivation $\boldsymbol{Y = AP}$
- We can conclude that $\boldsymbol{Y = U\Sigma}$ which is to say that **columns of $\boldsymbol{U\Sigma}$ are the coordinates in the new space. (a.k.a. scores).**
- Note that our derivation was based on the assumption that rows of A denote samples and columns denote features. If rows and columns of A are switched (like in your textbook), then the interpretation of U and V will also be switched.

Relationship between SVD and PCA

- In summary, SVD is an alternative method to calculate PCA.
- **Eigen decomposition of the covariance matrix = singular decomposition of the data matrix**
- SVD finds the same **eignvectors** (V) and also the **coefficients** that go along with them to reconstruct the data ($U\Sigma$).
- For example the ith row in the data matrix can be express as:

  $$X_i = \underbrace{U_{i,1}S_{1,1}}\, v_1 + \underbrace{U_{i,2}S_{2,2}}\, v_2 + \cdots$$

- Therefore, SVD not only finds the eignvectors but also finds the new coordinates. i.e. the coefficients that will be used to reconstruct the data.

Image Example

- SVD can be used to construct low dimensional representations of the image that does a reasonable job of reconstructing the image.
- Example: Each image in Viola Jones data set is 24x24 resulting in 576 dimensional vector upon reshaping.

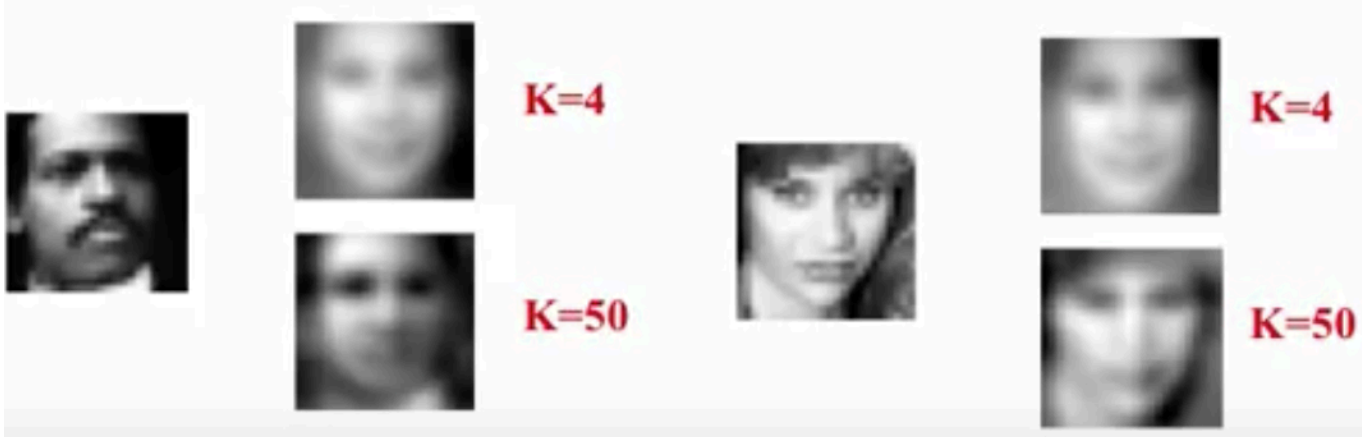
Image Example

- We can stack up all the images in a matrix A and subsequently take the SVD: $A = U\Sigma V^T$

- We can then take the first k columns of matrix V. Upon reshaping those vectors, we'll obtain the *eignfaces*.
- We can describe each image as a linear combination of the eigenfaces.

Image Example

- Consider reconstruction of the images using first 4 and first 50 principal components.
- Case 1: k=4
  - Images look very blurry and generic. It’s difficult to distinguish the images.
- Case 2: k=50
  - Images look less blurry. Face orientation and other features such as nose and neckline are clearer.
- Note that while sending the images in their original from involved sending 576 values, sending the compressed version involves 50 values plus the singular vectors that are common for all images.



Example

- Find principal directions of A= $\begin{matrix} -2 & 3.4 \\ 2 & 1.4 \\ -1 & 2.4 \\ 6 & 0.4 \\ -5 & -7.6 \end{matrix}$ by finding the eigenvalues/vectors of A^TA and SVD.

```
> A = matrix(c(-2,2,-1,6,-5,-3.4,-1.4,-2.4,-0.4,7.6), ncol = 2)
> A
     [,1] [,2]
[1,]   -2 -3.4
[2,]    2 -1.4
[3,]   -1 -2.4
[4,]    6 -0.4
[5,]   -5  7.6
> c = t(A)%*%A
> eigen(c)
$values
[1] 107.79006  39.40994

$vectors
           [,1]       [,2]
[1,] -0.6688446 -0.7434022
[2,]  0.7434022 -0.6688446
```

```
> s = svd(A)
> s
$d
[1] 10.382199  6.277734

$u
           [,1]        [,2]
[1,]  0.1146075 -0.59908180
[2,]  0.2290895  0.08767846
[3,]  0.1074263 -0.37412056
[4,]  0.4151749  0.66789635
[5,] -0.8662982  0.21762755

$v
           [,1]      [,2]
[1,]  0.6688446 0.7434022
[2,] -0.7434022 0.6688446
```

```
> sqrt(eigen(c)$values)
[1] 10.382199  6.277734
```