Matrices of Data and Singular Values

Gilbert Strang
MIT
gilstrang@gmail.com

Problem: Find the important part of a matrix

Best approximation to A by a matrix B of rank k

$$\min ||A - B|| = \sigma_{k+1} \quad \text{for} \quad B = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

$$
\begin{aligned}
S &= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{one independent column} \quad \text{rank } = 1 \\
&= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} \quad \text{column times row } (n \times 1)(1 \times n) = n \times n \\
&= 5 \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \quad \text{unit column vector } \mathbf{q} \\
&= \lambda \mathbf{q}\mathbf{q}^T \quad \mathbf{q} = \text{eigenvector of } S,\ \lambda = 5 = \text{eigenvalue of } S
\end{aligned}
$$

Flag of France (**rank 1**)

$$
= \begin{bmatrix} B & B & W & W & R & R \\ B & B & W & W & R & R \\ B & B & W & W & R & R \\ B & B & W & W & R & R \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} B & B & W & W & R & R \end{bmatrix}
$$

Symmetric matrices $S^T = S$

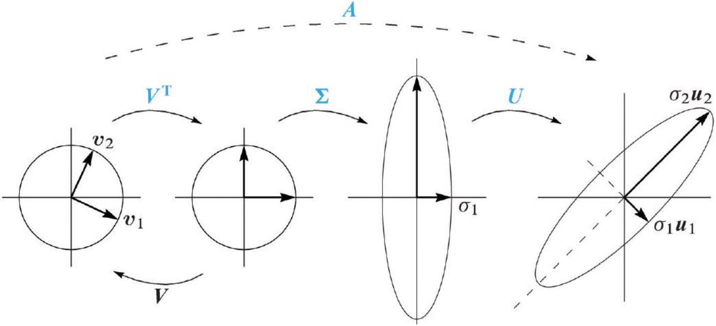
$S\mathbf{q} = \lambda\mathbf{q} \qquad S\left[\mathbf{q}_1 \ldots \mathbf{q}_n\right] = \left[\lambda_1\mathbf{q}_1 \ldots \lambda_n\mathbf{q}_n\right] \qquad SQ = Q\Lambda$

Orthogonal eigenvectors $Q^TQ = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 \ldots \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}$

$Q^T = Q^{-1}$ and $\boxed{S = Q\Lambda Q^T = \lambda_1\mathbf{q}_1\mathbf{q}_1^T + \cdots + \lambda_n\mathbf{q}_n\mathbf{q}_n^T}$

Rank one pieces $\rightarrow$ columns $\mathbf{q}$ times rows $\mathbf{q}^T$

$$\mathbf{A} = \text{(rotation)} \quad \text{(stretching)} \quad \text{(rotation)}$$
$$= (\text{orthogonal } U) \quad (\text{diagonal } \Sigma) \quad (\text{orthogonal } V^T)$$



Any matrix $A = U\Sigma V^T$	Symmetric matrix $S = Q\Lambda Q^T$
Singular Value Decomposition	**Spectral Theorem**
$AV = U\Sigma$ means $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$	$SQ = Q\Lambda$ means $S\mathbf{q}_i = \lambda_i \mathbf{q}_i$

$A^T A\mathbf{v} = \lambda\mathbf{v}$ Multiply by A $(AA^T)A\mathbf{v} = \lambda A\mathbf{v}$

AA^T has same eigenvalues $\lambda > 0$ as $A^T A$

AA^T has eigenvectors $\mathbf{u} = \frac{A\mathbf{v}}{\sqrt{\lambda}}$

$\boxed{A^T A = V\Lambda V^T}$ $\boxed{AA^T = U\Lambda U^T}$

The goal is the SVD $\boxed{A = U\Sigma V^T \text{ with } \sigma_i = \sqrt{\lambda_i}}$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

$\mathbf{v}$'s are right singular vectors, $\mathbf{u}$'s are left singular vectors

Crazy proof of the SVD

$$A(A^TA) = (AA^T)A$$
$$AV\Lambda V^T = U\Lambda U^TA$$
$$U^TAV\,\Lambda = \Lambda\,U^TAV$$

U^TAV commutes with diagonal matrix Λ

Suppose none of the λ's is repeated

U^TAV is also diagonal!!! Call it Σ

Then $\Sigma^T\Sigma = \Lambda \qquad \sigma_i = \sqrt{\lambda_i}$

A^TA	AA^T	A
eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r > 0$		$\sigma_i = \sqrt{\lambda_i}$
$A^TA\mathbf{v} = \lambda\mathbf{v}$	$AA^T\mathbf{u} = \lambda\mathbf{u}$	$A\mathbf{v} = \sigma\mathbf{u}$
$V^TV = I$	$U^TU = I$	orthonormal $\mathbf{u}$'s and $\mathbf{v}$'s
$A^TA = V\Lambda V^T$	$AA^T = U\Lambda U^T$	$A = U\sqrt{\Lambda}V^T$ (the SVD)
$= \lambda_1\mathbf{v}_1\mathbf{v}_1^T +$	$= \lambda_1\mathbf{u}_1\mathbf{u}_1^T +$	$= \boxed{\sigma_1\mathbf{u}_1\mathbf{v}_1^T} +$
$\lambda_2\mathbf{v}_2\mathbf{v}_2^T + \cdots$	$\lambda_2\mathbf{u}_2\mathbf{u}_2^T + \cdots$	$\sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots$

Columns of V, U multiply rows of V^T, U^T: rank one pieces

Transmitting a rank one matrix $A = \mathbf{u}\mathbf{v}^T$

Don't send $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ Send $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$

12 numbers instead of 36 numbers

$2N$ numbers instead of N^2 numbers

Flag with 3 stripes also has rank 1

Don't send $\begin{bmatrix} a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \\ a & a & c & c & e & e \end{bmatrix}$ Send $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & a & c & c & e & e \end{bmatrix}$

France, Italy, Germany, 20 more countries have 3 stripes

2 by 2 triangular matrix   Rank 2

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}$$

No saving to compress a 2 by 2 image!

The example shows rank 2 $= \mathbf{u}_1 \mathbf{v}_1^T + \mathbf{u}_2 \mathbf{v}_2^T$

Many choices for the $\mathbf{u}$'s and $\mathbf{v}$'s / this choice was not the SVD

The SVD choice: $\mathbf{u}_1^T \mathbf{u}_2 = 0 \quad \mathbf{v}_1^T \mathbf{v}_2 = 0 \quad ||\mathbf{u}_2||\ ||\mathbf{v}_2|| \to \mathbf{min}$

$\mathbf{u}_1 \mathbf{v_1}^T$ is the closest rank 1 matrix to A

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T$$

$$\text{SVD gives } \sigma_1 = \frac{\sqrt{5}+1}{2} \approx 1.6 \qquad \sigma_2 = \frac{\sqrt{5}-1}{2} \approx 0.6$$

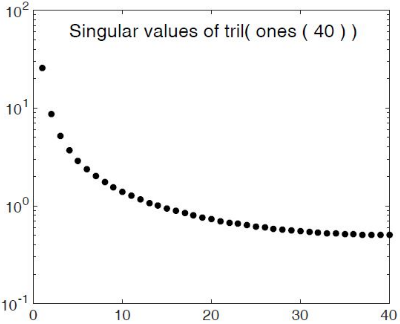
Remember $\sigma_1^2, \sigma_2^2 =$ eigenvalues of $A^T A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

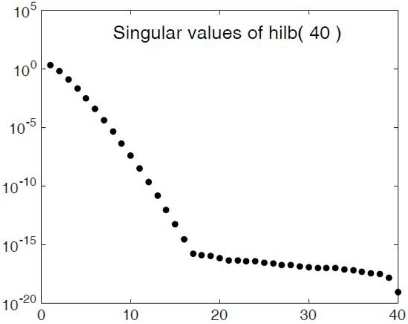
Example 2 = **Hilbert matrix** = very compressible!

$$H_{ij} = \frac{1}{i+j+1} = \text{ symmetric positive definite}$$

This Hilbert matrix is nearly singular and very ill-conditioned

Determinant of H is incredibly small





Not much decay for a lower triangular matrix of 1's

Fast decay for the Hilbert matrix (and many others)

Key to applications

The nearest rank k matrix to A: **Truncate the SVD**

$$A_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

"Pieces of A in order of importance"

Most useful when the σ's are exponentially decreasing.

Symmetric $S = Q\Lambda Q^T = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \cdots + \lambda_N \mathbf{q}_N \mathbf{q}_N^T$

Any matrix $A = U\Sigma V^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_N \mathbf{u}_N \mathbf{v}_N^T$

Could find these pieces (column $\times$ row) one at a time

The top eigenvector $\mathbf{x} = \mathbf{q}_1 = \mathbf{u}_1$ gives λ_1 and σ_1^2

$$\lambda_1 = \max \frac{\mathbf{x}^T S \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \text{ largest eigenvalue of } S$$

$$\sigma_1^2 = \max \frac{||A\mathbf{x}||^2}{||\mathbf{x}||} = \max \frac{\mathbf{x}^T A^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = (\text{largest singular value})^2$$

This shows how S (symmetric) corresponds to $A^T A$

λ (positive) corresponds to σ^2

Principal Component Analysis (PCA)

Find the closest line (or closest subspace) to n data points

Centered matrix $A = A_0 -$ (average of each row)

Every row of A adds to zero: The mean has been subtracted.

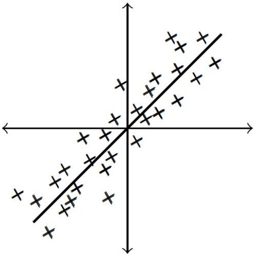
n columns in original A_0 give the data from the n individuals

One row could record heart rate for all individuals

Want to find the important information in A_0 and A

Data comes in a matrix!

Principal Components when A has $m = 2$ rows



Each column of A is a point (x, y). Each row adds to zero.

Average x and average y are zero: data is centered at $(0, 0)$

PCA finds the closest line to the data

Closest = smallest sum of (**perpendicular distances**)^2

$$S = AA^T = (2 \text{ by } n)(n \text{ by } 2) = 2 \text{ by } 2$$

Top eigenvector $\mathbf{u}_1$ gives the closest line

Finance, genetics, model reduction, many applications

$S = AA^T/(N-1)$ is the sample covariance matrix

Each σ_i^2 tells how much of S is "explained" by $\mathbf{u}_i$

Total variance $= \sigma_1^2 + \cdots + \sigma_m^2 =$ trace of S

In practice: Stop when σ_i^2 is small

This gives the "effective rank" of S and A

$$A = \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix} \text{ has } S = \frac{AA^T}{5} = \begin{bmatrix} 20 & 25 \\ 25 & 40 \end{bmatrix}$$

Research question with Alex Townsend

Start from $f(x, y)$ on the square $0 \leq x \leq 1,\ 0 \leq y \leq 1$

Create N by N matrix $A_{ij} = f\left(\frac{i}{N}, \frac{j}{N}\right)$ = "picture of f"

Is this matrix compressible like Hilbert?

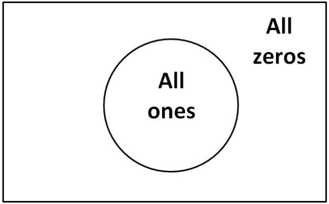
Is it incompressible like this triangular flag? Circular flag?

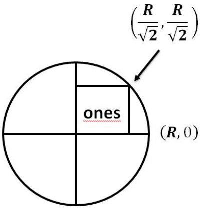
$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \text{ comes from } f(x, y) = \begin{cases} 1 & x \geq y \\ 0 & x < y \end{cases}$$

US flag and UK flag are incompressible / they have diagonals

What is the rank of the Japanese flag? Circle.

$1-0$ matrix A with a circular disk of all 1's. Radius R.





$$\text{rank} \approx 2\left(R - \frac{R}{\sqrt{2}}\right) = \left(2-\sqrt{2}\right)R$$

Remarkable fact: singular values $\sigma \geq 1$ constant

Townsend found a description of compressible matrices C

Solve Sylvester's equation $AC - CB = F$

1. F should have rapidly decreasing σ's
2. Eigenvalues of A should be separated from eigenvalues of B

Singular value decay for C depends on rational approximation

Rational approximation is often exponentially close