Outline

1. Introduction to The Singular Value Decomposition
2. The Geometry of the 2 x 2 Case
3. The General Case
4. SVD Steps
5. Singular Values and Volumes
6. The Pseudoinverse
7. Least Squares
8. Application: Image Compression
9. Principal Components Analysis
10. WYSK
The Singular Value Decomposition

**Matrix decomposition**: fundamental tool for
— understanding the action of a matrix
— establishing its suitability to solve a problem
— solving linear systems more efficiently and effectively

Symmetric matrices: *eigendecomposition*
More general matrices: the *singular value decomposition*

Image compression and the SVD
Original image → Highest compression → Less compression → Original recovered
The Geometry of the $2 \times 2$ Case

*Orthonormal* vectors $\mathbf{v}_1$ and $\mathbf{v}_2 \Rightarrow \text{orthogonal matrix } V = [\mathbf{v}_1 \ \mathbf{v}_2]$

*Orthonormal* vectors $\mathbf{u}_1$ and $\mathbf{u}_2 \Rightarrow \text{orthogonal matrix } U = [\mathbf{u}_1 \ \mathbf{u}_2]$

Want $\mathbf{v}_i$ and $\mathbf{u}_i$ such that $A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1$ and $A\mathbf{v}_2 = \sigma_2 \mathbf{u}_2$:

$$AV = U\Sigma \quad \text{where} \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

The *singular value decomposition* (SVD) of $A$:

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

$\sigma_i$ called the *singular values* of $A$
The Geometry of the $2 \times 2$ Case

Properties of symmetric positive definite matrices such as $A^T A$
— Real and positive eigenvalues
— Eigenvectors are orthogonal

$$A^T A = (U \Sigma V^T)^T (U \Sigma V^T)$$
$$= V \Sigma^T U^T U \Sigma V^T$$
$$= V \Sigma^T \Sigma V^T$$
$$= V \Lambda' V^T$$

where

$$\Lambda' = \begin{bmatrix} \lambda'_1 & 0 \\ 0 & \lambda'_2 \end{bmatrix} = \Sigma^T \Sigma = \begin{bmatrix} \sigma^2_1 & 0 \\ 0 & \sigma^2_2 \end{bmatrix}$$

This is the \textit{eigendecomposition} of $A^T A$

Columns of $V$ called the \textit{right singular vectors} of $A$
The Geometry of the $2 \times 2$ Case

Eigendecomposition of symmetric positive definite $AA^T$

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$$

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

$$= U\Sigma \Sigma^T U^T$$

$$= U\Lambda' U^T$$

$\Lambda' = \Sigma^T \Sigma = \Sigma \Sigma^T$

$\Rightarrow$ Eigenvalues are diagonal entries of $\Lambda'$

$\Rightarrow$ Eigenvectors are columns of $U$

— Called the left singular vectors of $A$
The Geometry of the $2 \times 2$ Case

Elements of the SVD of $A$:

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

— The singular values

$$\sigma_i = \sqrt{\lambda'_i}$$

where $\lambda'_i$ are the eigenvalues of $A^T A$ and $AA^T$

— The columns of $V$ are the eigenvectors of $A^T A$

— The columns of $U$ are the eigenvectors of $AA^T$

Can compute $u_i = A v_i / \| \cdot \|$ since $AV = U\Sigma$
The Geometry of the $2 \times 2$ Case

**Example:** symmetric positive definite matrix that scales in $e_1$-direction

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

$$AA^T = A^T A = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$  

eigenvalues: $\lambda_1' = 9 \quad \lambda_2' = 1$

$$\Rightarrow \quad \sigma_1 = 3 \quad \text{and} \quad \sigma_2 = 1$$

$$U = V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

**SVD** $A = U \Sigma V^T$:

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Positive definite matrix $\Rightarrow$ SVD identical to eigendecomposition
The Geometry of the $2 \times 2$ Case

Action: unit circle $\Rightarrow$ action ellipse
— Semi-major axis length $\sigma_1$ — Semi-minor axis length $\sigma_2$

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \text{ circle } \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

Thick point: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, Thin point: $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(Left: previous example; Right: next example)
The Geometry of the $2 \times 2$ Case

**Example:** a shear

\[
A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow \quad A^T A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \quad AA^T = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}
\]

Eigenvalues: $\lambda_1' = 5.82$ and $\lambda_2' = 0.17 \quad \Rightarrow \quad \sigma_1 = 2.41$ and $\sigma_2 = 0.41$

Eigenvectors of $A^T A \Rightarrow$ orthonormal column vectors of

\[
V = \begin{bmatrix} 0.38 & -0.92 \\ 0.92 & 0.38 \end{bmatrix}
\]

Eigenvectors of $AA^T \Rightarrow$ orthonormal column vectors of

\[
U = \begin{bmatrix} 0.92 & -0.38 \\ 0.38 & 0.92 \end{bmatrix}
\]

SVD of $A$:

\[
\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.92 & -0.38 \\ 0.38 & 0.92 \end{bmatrix} \begin{bmatrix} 2.41 & 0 \\ 0 & 0.41 \end{bmatrix} \begin{bmatrix} 0.38 & -0.92 \\ 0.92 & 0.38 \end{bmatrix}
\]
The Geometry of the $2 \times 2$ Case

Break down the action of $A$ in terms of the SVD

$$
\begin{bmatrix}
1 & 2 \\
0 & 1
\end{bmatrix} = \begin{bmatrix} 0.92 & -0.38 \\
0.38 & 0.92 \end{bmatrix} \begin{bmatrix} 2.41 & 0 \\
0 & 0.41 \end{bmatrix} \begin{bmatrix} 0.38 & -0.92 \\
0.92 & 0.38 \end{bmatrix}
$$

Clockwise from top left:
- Initial point set forming a circle with two reference points
- $V^T x$ rotates clockwise $67.5^\circ$
- $\Sigma V^T x$ stretches in $e_1$ and shrinks in $e_2$
- $U \Sigma V^T x$ rotates counterclockwise $22.5^\circ$
The General Case

Now: $m \times n$ matrix $A$ — not necessarily square nor invertible

$A = U \Sigma V^T$

$A^T A = V \Lambda' V^T \Rightarrow \Lambda'$ is $n \times n$

$AA^T = U \Lambda' U^T \Rightarrow \Lambda'$ is $m \times m$

Both $\Lambda'$ hold the same non-zero eigenvalues $\Rightarrow$ rank $\leq \min\{m, n\}$
The General Case

Want $v_i$ and $u_i$ such that $A v_i = \sigma_i u_i$

$$A V = U \Sigma$$

Rank $r$ of $A$ plays a role in the SVD

Main properties:
- $\Sigma$ has non-zero singular values $\sigma_1, \ldots, \sigma_r$ and all other entries zero
- First $r$ columns of $U$ form an orthonormal basis for column space of $A$
- Last $m - r$ columns of $U$ form an orthonormal basis for null space of $A^T$
- First $r$ columns of $V$ form an orthonormal basis for row space of $A$
- Last $n - r$ columns of $V$ form an orthonormal basis for null space of $A$
**The General Case**

**Example:** Rank 2 matrix

\[
A = \begin{bmatrix}
1 & 0 \\
0 & 2 \\
0 & 1 \\
\end{bmatrix}
\]

\[
A^T A = \begin{bmatrix}
1 & 0 \\
0 & 5 \\
\end{bmatrix} \\
\lambda'_1 = 5 \\
\lambda'_2 = 1 \\
V = \begin{bmatrix}
0 & 1 \\
1 & 0 \\
\end{bmatrix}
\]

\[
A A^T = \begin{bmatrix}
1 & 0 & 0 \\
0 & 4 & 2 \\
0 & 2 & 1 \\
\end{bmatrix} \\
\lambda'_1 = 5 \\
\lambda'_2 = 1 \\
\lambda'_3 = 0 \\
U = \begin{bmatrix}
0.89 & 0 & -0.44 \\
0.44 & 0 & 0.89 \\
\end{bmatrix}
\]

\[
\Sigma = \begin{bmatrix}
2.23 & 0 \\
0 & 1 \\
0 & 0 \\
\end{bmatrix}
\]

\[
A = U \Sigma V^T : \\
\begin{bmatrix}
1 & 0 \\
0 & 2 \\
0 & 1 \\
\end{bmatrix} = \begin{bmatrix}
0 & 1 & 0 \\
0.89 & 0 & -0.44 \\
0.44 & 0 & 0.89 \\
\end{bmatrix} \begin{bmatrix}
2.23 & 0 \\
0 & 1 \\
0 & 0 \\
\end{bmatrix} \begin{bmatrix}
0 & 1 \\
1 & 0 \\
\end{bmatrix}
\]

\[m > n \implies u_3 \text{ is in the null space of } A^T \implies A^T u_3 = 0\]
The General Case

SVD and action of a matrix

Clockwise from top left:
1) Initial circle point set
2) $V^T x$ reflects
3) $\Sigma V^T x$ stretches in $e_1$
4) $U \Sigma V^T x$ rotates

$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 1 \end{bmatrix}$
The General Case

Example:

\[ A = \begin{bmatrix} -0.8 & 0 & 0.8 \\ 1 & 1.5 & -0.3 \end{bmatrix} \]

\[ A^T A = \begin{bmatrix} 1.64 & 1.5 & -0.94 \\ 1.5 & 2.25 & -0.45 \\ -0.94 & -0.45 & 0.73 \end{bmatrix} \]

\[ \lambda'_1 = 3.77 \quad \lambda'_2 = 0.84 \quad \lambda'_3 = 0 \]

\[ V = \begin{bmatrix} -0.63 & 0.38 & 0.67 \\ -0.71 & -0.62 & -0.31 \\ 0.30 & -0.68 & 0.67 \end{bmatrix} \]

\[ AA^T = \begin{bmatrix} 1.28 & -1.04 \\ -1.04 & 3.34 \end{bmatrix} \]

\[ \lambda'_1 = 3.77 \quad \lambda'_2 = 0.84 \]

\[ U = \begin{bmatrix} 0.39 & -0.92 \\ -0.92 & -0.39 \end{bmatrix} \]

\[ \Sigma = \begin{bmatrix} 1.94 & 0 & 0 \\ 0 & 0.92 & 0 \end{bmatrix} \]

SVD: \[ A = U \Sigma V^T \]

\[ \begin{bmatrix} -0.8 & 0 & 0.8 \\ 1 & 1.5 & -0.3 \end{bmatrix} = \begin{bmatrix} 0.39 & -0.92 \\ -0.92 & -0.39 \end{bmatrix} \begin{bmatrix} 1.94 & 0 & 0 \\ 0 & 0.92 & 0 \end{bmatrix} \begin{bmatrix} -0.63 & -0.71 & 0.3 \\ 0.38 & -0.62 & -0.68 \\ 0.67 & -0.31 & 0.67 \end{bmatrix} \]

\[ m < n \implies v_3 \text{ in null space of } A \implies Av_3 = 0 \]
The General Case

SVD and action of a matrix

\[
A = \begin{bmatrix}
-0.8 & 0 & 0.8 \\
1 & 1.5 & -0.3
\end{bmatrix}
\]

Clockwise from top left:
1) Initial circle point set
2) \( V^T x \)
3) \( \Sigma V^T x \)
4) \( U \Sigma V^T x \)
The General Case

Example: a projection into the \([e_1, e_2]\)-plane — a rank deficient matrix

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

\(A\) is symmetric and idempotent \(\Rightarrow A = A^T A = AA^T\)

\(A = U\Sigma V^T:\)

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix} \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

Rank = 2

\(\Rightarrow\) first 2 columns of \(U\) form orthonormal basis for column space of \(A\)

\(\Rightarrow\) first 2 columns of \(V\) form orthonormal basis for row space of \(A\)

\(e_3\) vector projected to the zero vector \(\Rightarrow\) spans the null space of \(A\) and \(A^T\)
SVD Steps

\[ A = U \Sigma V^T \]

Here: review steps — for a robust algorithm ⇒ advanced numerical methods

**Input:** an \( m \times n \) matrix \( A \)

**Output:** \( U, V, \Sigma \) such that \( A = U \Sigma V^T \)

1. Find the *eigenvalues* \( \lambda'_1, \ldots, \lambda'_n \) of \( A^T A \)
   - Order the \( \lambda'_i \) so that \( \lambda'_1 \geq \lambda'_2 \geq \ldots \geq \lambda'_n \)
   - Suppose \( \lambda'_1, \ldots, \lambda'_r > 0 \), then the *rank* of \( A \) is \( r \)

2. Create an \( m \times n \) diagonal matrix \( \Sigma \) with \( \sigma_{i,i} = \sqrt{\lambda'_i}, i = 1, \ldots, r \)

3. Find the corresponding (normalized) eigenvectors \( \mathbf{v}_i \) of \( A^T A \)

4. Create an \( n \times n \) matrix \( V \) with column vectors \( \mathbf{v}_i \)

5. Find the (normalized) eigenvectors \( \mathbf{u}_i \) of \( AA^T \)

6. Create an \( m \times m \) matrix \( U \) with column vectors \( \mathbf{u}_i \)
SVD Steps

Notes on steps:

- Can compute $u_i$, $i = 1, r$ as $u_i = A v_i / \| \cdot \|$
  - If $m > n$ then the remaining $u_i$ are found from the null space of $A^T$

- The only “hard” task is finding the $\lambda_i'$
  - Since $A^T A$ is symmetric $\Rightarrow$ Can choose a highly efficient algorithm

- Forming $A^T A$ can result in an ill-posed problem
  - $\kappa(A^T A) = \kappa(A)^2$
  - Avoid direct computation of this matrix
    — employ the Householder method
Singular Values and Volumes

Application: compute the determinant

\[ \det U = \pm 1 \quad \text{and} \quad \det V = \pm 1 \quad \Rightarrow \quad |\det A| = \det \Sigma = \sigma_1 \cdots \sigma_n \]

Example: given a 2D triangle \( T \) with area \( \varphi \)
Transform \( T \to T' \) with 2D linear map with singular values \( \sigma_1, \sigma_2 \)
Area of \( T' = \pm \sigma_1 \sigma_2 \varphi \)

Example: given a 3D object \( O \) with volume \( \varphi \)
Transform \( O \to O' \) with 3D linear map with singular values \( \sigma_1, \sigma_2, \sigma_3 \)
Volume of \( O' = \pm \sigma_1 \sigma_2 \sigma_3 \varphi \)

Recall determinants without using singular values

\[ \det A = \lambda_1 \cdots \lambda_n \]
The Pseudoinverse

The inverse of a matrix:
— Limited to square, nonsingular matrices
— Mainly a theoretical tool for analyzing the solution to a linear system

The generalized inverse or pseudoinverse \( A^\dagger \)
— For general matrices
— Suited for practical use
— Can be computed with the SVD

Given an \( m \times n \) diagonal matrix \( \Sigma \) with diagonal elements \( \sigma_i \)
The pseudoinverse: the \( n \times m \) matrix \( \Sigma^\dagger \) with

\[
\sigma_i^\dagger = \begin{cases} 
\frac{1}{\sigma_i} & \text{if } \sigma_i > 0 \\
0 & \text{else}
\end{cases}
\]

If \( \text{rank}(\Sigma) = r \) then
— \( \Sigma^\dagger \Sigma \) holds the \( r \times r \) identity matrix
— All other elements are zero
The Pseudoinverse

Leads to the pseudoinverse for a general $m \times n$ matrix $A$

$$A^\dagger = (U\Sigma V^T)^{-1} = V\Sigma^\dagger U^T$$

If $A$ is square and invertible then $A^\dagger = A^{-1}$

Properties:

$$A^\dagger AA^\dagger = A^\dagger \quad \text{and} \quad AA^\dagger A = A$$

Often times called the Moore-Penrose generalized inverse

Primary application: least squares approximation
The Pseudoinverse

**Example:** Find the pseudoinverse of

\[ A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0.89 & 0 & -0.44 \\ 0.44 & 0 & 0.89 \end{bmatrix} \begin{bmatrix} 2.23 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \]

\[ \Sigma^\dagger = \begin{bmatrix} 1/2.23 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \]

\[ A^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/5 & 1/5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/2.23 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.89 & 0.44 \\ 1 & 0 & 0 \\ 0 & -0.44 & 0.89 \end{bmatrix} \]
The Pseudoinverse

Example: square and nonsingular $A$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}$$

The pseudoinverse is equal to the inverse:

$$A^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}$$
Least Squares

Overdetermined linear system: \( m \) equations in \( n \) unknowns where \( m \geq n \)

\[
Ax = b
\]

Linear system is inconsistent — unlikely that \( b \) lives in subspace \( \mathcal{V} \) defined by columns of \( A \)

The least squares solution finds the orthogonal projection of \( b \) into \( \mathcal{V} \) — Call this projection \( b' \)

\( \Rightarrow \) Solution to \( Ax = b' \) produces vector closest to \( b \) that lives in \( \mathcal{V} \)

Normal equations

\[
A^T Ax = A^T b \quad \text{solution minimizes } \|Ax - b\|
\]

This system can be ill-posed \( \Rightarrow \) use pseudoinverse

\[
x = A^\dagger b
\]
Least Squares

Why is $x = A^\dagger b$ the least squares solution?

Find $x$ to minimize $\|Ax - b\|$

$$Ax - b = U\Sigma V^T x - b$$
$$= U\Sigma V^T x - UU^T b$$
$$= U(\Sigma y - z)$$

This new framing of the problem exposes that

$$\|Ax - b\| = \|\Sigma y - z\|$$

⇒ an easier diagonal least squares problem to solve
Least Squares

Steps:
1. Compute the SVD $A = U\Sigma V^T$
2. Compute the $m \times 1$ vector $z = U^T b$
3. Compute the $n \times 1$ vector $y = \Sigma^\dagger z$
   — Least squares solution to $m \times n$ problem $\Sigma y = z$
   
   \[
   \begin{bmatrix}
   \sigma_1 y_1 - z_1 \\
   \sigma_2 y_2 - z_2 \\
   \vdots \\
   \sigma_r y_r - z_r \\
   -z_{r+1} \\
   \vdots \\
   -z_m
   \end{bmatrix}
   \]

   requires minimizing $v = \Sigma y - z$

   $\text{rank}(\Sigma) = r$

   $y$ minimizing $v$: $y_i = z_i / \sigma_i \quad i = 1, \ldots, r \quad \Rightarrow \quad y = \Sigma^\dagger z$
4. Compute the $n \times 1$ solution vector $x = V y$
Least Squares

Summarize — The calculations in reverse order include

\[ x = Vy \]
\[ x = V(\Sigma^\dagger z) \]
\[ x = V\Sigma^\dagger(U^Tb) \]

**Example:** Revisit temperature-time data: find the best fit line coefficients — Chapter 12 (normal equations) and Chapter 13 (Householder)

\[
\begin{bmatrix}
0 & 1 \\
10 & 1 \\
20 & 1 \\
30 & 1 \\
40 & 1 \\
50 & 1 \\
60 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x \\
x \\
x \\
x \\
x \\
x \\
x \\
\end{bmatrix}
= 
\begin{bmatrix}
30 \\
25 \\
40 \\
40 \\
30 \\
5 \\
25 \\
\end{bmatrix}
Least Squares

Step 1) Compute the SVD \( A = U\Sigma V^T \)

\[
\Sigma = \begin{bmatrix}
95.42 & 0 \\
0 & 1.47 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\end{bmatrix}
\]

\[
\Sigma^\dagger = \begin{bmatrix}
0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.68 & 0 & 0 & 0 & 0 & 0 \\
U & 7 \times 7 \\
\end{bmatrix}
\]

Step 2) \( z = U^Tb = \begin{bmatrix}
54.5 \\
51.1 \\
3.2 \\
-15.6 \\
9.6 \\
15.2 \\
10.8 \\
\end{bmatrix} \)

Step 3) \( y = \Sigma^\dagger z = \begin{bmatrix}
0.57 \\
34.8 \\
\end{bmatrix} \)

Step 4) \( x = Vy = \begin{bmatrix}
-0.23 \\
34.8 \\
\end{bmatrix} \) \( \Rightarrow \) best fit line: \( x_2 = -0.23x_1 + 34.8 \)
The normal equations give a best approximation

\[ x = (A^T A)^{-1} A^T b \]

to the original problem \( Ax = b \)

by considering \( b' \) in the subspace of \( A \) called \( V \)
Substitute this expression for \( x \) into \( Ax = b' \):

\[ b' = A(A^T A)^{-1} A^T b = AA^\dagger b = \text{proj}_V b \]

— Goal is to project \( b \) into \( V \) ⇒ \( AA^\dagger \) is a projection
— Property \( A^\dagger AA^\dagger = A^\dagger \) ensures necessary idempotent property
Application: Image Compression

Given $m \times n$ matrix $A$ with $k = \min(m, n)$ singular values $\sigma_i$

$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$

Using the SVD write $A$ as a sum of $k$ rank one matrices:

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_k u_k v_k^T$$

Use this for image compression

— An image is comprised of a grid of colored pixels — grayscales here

— Figure (left): input image with $4 \times 4$ pixels

— Each grayscale associated with a number ⇒ grid is a matrix
Application: Image Compression

Singular values for this matrix are $\sigma_i = 7.1, 3.8, 1.3, 0.3$
Images from left to right $I_0, I_1, I_2, I_3$ — Original image is $I_0$

\[
A_1 = \sigma_1 u_1 v_1^T \implies \text{image } I_1
\]
\[
A_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T \implies \text{image } I_2
\]

Original image nearly replicated incorporating only half the singular values
$\Rightarrow$ $\sigma_1$ and $\sigma_2$ large in comparison to $\sigma_3$ and $\sigma_4$

Image $I_3$ created from $A_3 = A_2 + \sigma_3 u_3 v_3^T$
Image $I_4$ is not displayed — identical to $I_0$
Application: Image Compression

The change in an image by adding the smallest $\sigma_i$ can be hardly noticeable
⇒ Omitting images $l_k$ corresponding to small $\sigma_k$ amounts to compressing the original image

Chapter introduction Figure: $8 \times 8$ matrix

$\sigma_i = 6.2, 1.7, 1.49, 0, \ldots, 0$
— Figure illustrates images corresponding to each non-zero $\sigma_i$
— Last image is identical to the input
⇒ the five remaining $\sigma_i = 0$ are unimportant to image quality
**Principal Components Analysis**

**Scatter plot:** data pairs recorded in Cartesian coordinates

Each circle represents a coordinate pair (point) in the \([\mathbf{e}_1, \mathbf{e}_2]\)-system

Example: Gross Domestic Product and poverty rate pairs

How might we reveal trends in this data set?
Principal Components Analysis

Given: 2D data set \( x_1, \ldots, x_n \) such that \( x_1 + \ldots + x_n = 0 \)

Let \( d \) be a unit vector

Project \( x_i \) onto line containing \( d \)

Result:

vector with (signed) length \( x_i \cdot d \)

\[
I(d) = [x_1 \cdot d]^2 + \ldots + [x_n \cdot d]^2
\]

Rotate \( d \) around the origin

For each position compute \( I(d) \)

Directions corresponding to largest and smallest \( I(d) \) are orthogonal

⇒ indicates variation in data
Principal Components Analysis

Arrange data $x_i$ in a matrix

$$X = \begin{bmatrix}
    x_1^T \\
    x_2^T \\
    \vdots \\
    x_n^T
\end{bmatrix}$$

then

$$l(d) = \|Xd\|^2 = (Xd) \cdot (Xd) = d^T X^T X d \quad (*)$$

Let $C = X^T X$  

$C$ is a symmetric positive definite $2 \times 2$ matrix

$\Rightarrow \ (*) \text{ is a } \textit{quadratic form} \quad \text{— See Figure}$
For which \( d \) is \( l(d) \) maximal?

Answer: \( d \) that corresponds to \( C \)'s dominant eigenvector
And: \( l(d) \) is minimal for \( d \) being the eigenvector corresponding to \( C \)'s smallest eigenvalue

These eigenvectors form the major and minor axis of the action ellipse of \( C \) (Thick lines in Figure)
— Eigenvectors orthogonal because \( C \) is symmetric
Principal Components Analysis

Look more closely at $C$

$$
c_{1,1} = x_{1,1}^2 + x_{2,1}^2 + \ldots + x_{n,1}^2
$$

$$
c_{1,2} = c_{2,1} = x_{1,1}x_{1,2} + x_{2,1}x_{2,2} + \ldots + x_{n,1}x_{n,2}
$$

$$
c_{2,2} = x_{1,2}^2 + x_{2,2}^2 + \ldots + x_{n,2}^2.
$$

If each element of $C$ is divided by $n$ it is called the covariance matrix
— Summary of the variation in each coordinate and between coordinates
— Dividing by $n$ will result in scaled eigenvalues
eigenvectors will not change
Principal Components Analysis

Eigenvectors provide a convenient *local coordinate frame* for the data set
— Idea behind the principle of the *eigendecomposition*
— This frame is commonly called the *principal axes*

Let \( V = [v_1 \ v_2] \) hold the normalized eigenvectors as column vectors
— \( v_1 \) is the dominant eigenvector

Orthogonal transformation of the data \( X \)
— aligns \( v_1 \) with \( e_1 \) and \( v_2 \) with \( e_2 \)

\[
\hat{X} = XV \quad \Rightarrow \quad \hat{x}_i = \begin{bmatrix} x_i \cdot v_1 \\ x_i \cdot v_2 \end{bmatrix}
\]
Summary:

- Established a **principal components coordinate system**
  - Defined by the eigenvectors of the covariance matrix
  - Greatest variance corresponds to the first coordinate

- Data coordinates are now in terms of the trend lines
  - Coordinates directly measure the distance from each trend line

⇒ Name of this method: **Principal Components Analysis (PCA)**
Principal Components Analysis

PCA can also be used for *data compression* by reducing dimensionality.

Let $V$ hold only some eigenvectors.
- Example: most significant then $V = v_1$ (left Figure).
- Example: $V = v_2$ (right Figure).

Greater spread of the data corresponds to higher variance.

Here 2D data but the real power of PCA comes with higher dimensional data.
- Difficult to visualize and understand relationships between dimensions.
Singular Value Decomposition (SVD)
- singular values
- right singular vector
- left singular vector
- SVD matrix dimensions
- SVD column, row, and null spaces
- SVD steps
- volume in terms of singular values
- eigendecomposition
- matrix decomposition

- action ellipse axes length
- pseudoinverse
- generalized inverse
- least squares solution via the pseudoinverse
- quadratic form
- contour ellipse
- Principal Components Analysis (PCA)
- covariance matrix