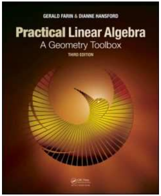
Practical Linear Algebra: A Geometry Toolbox

Third edition

Chapter 16: The Singular Value Decomposition

Gerald Farin & Dianne Hansford

CRC Press, Taylor & Francis Group, An A K Peters Book
www.farinhansford.com/books/pla



Outline

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*

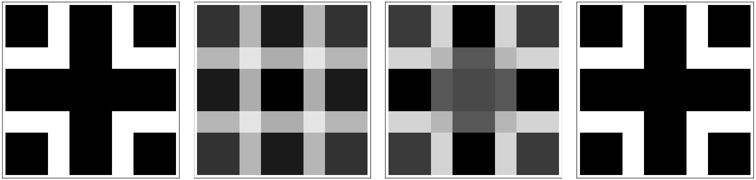


Image compression and the SVD
Original image→Highest compression→Less compression→Original recovered

The Geometry of the 2×2 Case

Orthonormal vectors $\mathbf{v}_1$ and $\mathbf{v}_2 \Rightarrow$ *orthogonal* matrix $V = [\mathbf{v}_1 \ \mathbf{v}_2]$
Orthonormal vectors $\mathbf{u}_1$ and $\mathbf{u}_2 \Rightarrow$ *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 $A\mathbf{v}_2 = \sigma_2\mathbf{u}_2$:

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

The singular value decomposition (SVD) of A:

$$A = U\Sigma V^{\mathrm{T}}$$

σ_i called the singular values of A

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

$$\begin{aligned} A^{\mathrm{T}}A &= (U\Sigma V^{\mathrm{T}})^{\mathrm{T}}(U\Sigma V^{\mathrm{T}}) \\ &= V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} \\ &= V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}} \\ &= V\Lambda' V^{\mathrm{T}} \end{aligned}$$

where

$$\Lambda' = \begin{bmatrix} \lambda'_1 & 0 \\ 0 & \lambda'_2 \end{bmatrix} = \Sigma^{\mathrm{T}}\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

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

Columns of V called the right singular vectors of A

The Geometry of the 2 × 2 Case

Eigendecomposition of symmetric positive definite AA^{T}

$$\begin{aligned} AA^{\mathrm{T}} &= (U\Sigma V^{\mathrm{T}})(U\Sigma V^{\mathrm{T}})^{\mathrm{T}} \\ &= U\Sigma V^{\mathrm{T}} V\Sigma^{\mathrm{T}} U^{\mathrm{T}} \\ &= U\Sigma\Sigma^{\mathrm{T}} U^{\mathrm{T}} \\ &= U\Lambda' U^{\mathrm{T}} \end{aligned}$$

$\Lambda' = \Sigma^{\mathrm{T}}\Sigma = \Sigma\Sigma^{\mathrm{T}}$
$\Rightarrow$ Eigenvalues are diagonal entries of Λ'
$\Rightarrow$ Eigenvectors are columns of U
— Called the left singular vectors of A

Elements of the SVD of A:

$$A = U\Sigma V^{\mathrm{T}}$$

— The singular values

$$\sigma_i = \sqrt{\lambda_i'}$$

where λ_i' are the eigenvalues of $A^{\mathrm{T}}A$ and AA^{T}

— The columns of V are the eigenvectors of $A^{\mathrm{T}}A$

— The columns of U are the eigenvectors of AA^{T}

Can compute $\mathbf{u}_i = A\mathbf{v}_i/\|\cdot\|$ since $AV = U\Sigma$

The Geometry of the 2 × 2 Case

Example: symmetric positive definite matrix that scales in $\mathbf{e}_1$-direction

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

$$AA^{\mathrm{T}} = A^{\mathrm{T}}A = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{eigenvalues: } \lambda_1' = 9 \quad \lambda_2' = 1$$

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

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

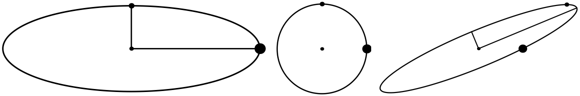
SVD $A = U\Sigma V^{\mathrm{T}}$:

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

Positive definite matrix ⇒ SVD identical to eigendecomposition

The Geometry of the 2 × 2 Case

Action: unit circle ⇒ *action ellipse*
— Semi-major axis length σ_1 — Semi-minor axis length σ_2



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

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

(Left: previous example; Right: next example)

Example: a shear

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

Eigenvalues: $\lambda'_1 = 5.82$ and $\lambda'_2 = 0.17 \quad \Rightarrow \quad \sigma_1 = 2.41$ and $\sigma_2 = 0.41$

Eigenvectors of $A^{\mathrm{T}}A \Rightarrow$ orthonormal column vectors of

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

Eigenvectors of $AA^{\mathrm{T}} \Rightarrow$ orthonormal column vectors of

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

SVD of A:

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

Break down the action of A in terms of the SVD

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

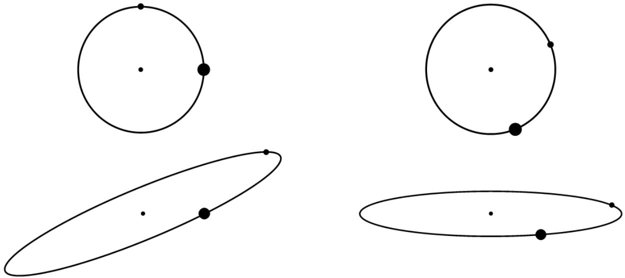
Clockwise from top left:

— Initial point set forming a circle with two reference points

— $V^{\mathrm{T}}\mathbf{x}$ rotates clockwise 67.5°

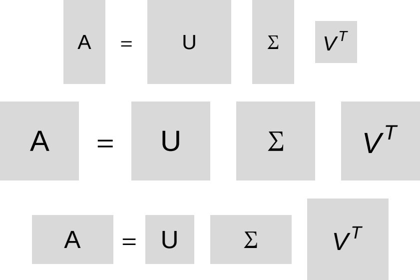
— $\Sigma V^{\mathrm{T}}\mathbf{x}$ stretches in $\mathbf{e}_1$ and shrinks in $\mathbf{e}_2$

— $U\Sigma V^{\mathrm{T}}\mathbf{x}$ rotates counterclockwise 22.5°



The General Case

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



Top: $m > n$ Middle: $m = n$ Bottom: $m < n$

U is $m \times m$ Σ is $m \times n$ V^{T} is $n \times n$

$A^{\mathrm{T}}A = V\Lambda'V^{\mathrm{T}} \Rightarrow \Lambda'$ is $n \times n$ $AA^{\mathrm{T}} = U\Lambda'U^{\mathrm{T}} \Rightarrow \Lambda'$ is $m \times m$

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

The General Case

Want $\mathbf{v}_i$ and $\mathbf{u}_i$ such that $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$

$$AV = U\Sigma$$

Rank r of A plays a role in the SVD

Main properties:

- Σ 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

Example: Rank 2 matrix

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

$$A^{\mathrm{T}}A = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \quad \begin{matrix} \lambda'_1 = 5 \\ \lambda'_2 = 1 \end{matrix} \quad V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

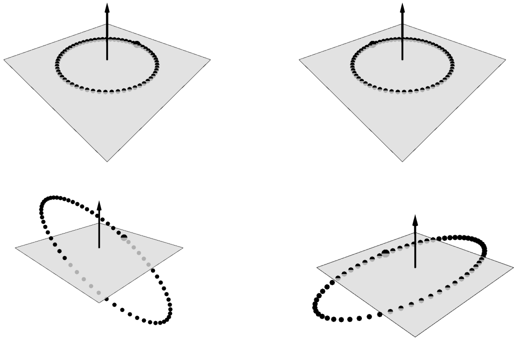
$$AA^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix} \quad \begin{matrix} \lambda'_1 = 5 \\ \lambda'_2 = 1 \\ \lambda'_3 = 0 \end{matrix} \quad U = \begin{bmatrix} 0 & 1 & 0 \\ 0.89 & 0 & -0.44 \\ 0.44 & 0 & 0.89 \end{bmatrix}$$

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

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

$m > n \Rightarrow \mathbf{u}_3$ is in the null space of $A^{\mathrm{T}} \Rightarrow A^{\mathrm{T}}\mathbf{u}_3 = \mathbf{0}$

SVD and action of a matrix



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

Clockwise from top left:

1) Initial circle point set 2) $V^{\mathrm{T}}\mathbf{x}$ reflects

3) $\Sigma V^{\mathrm{T}}\mathbf{x}$ stretches in $\mathbf{e}_1$ 4) $U\Sigma V^{\mathrm{T}}\mathbf{x}$ rotates

The General Case

Example:

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

$$A^{\mathrm{T}}A = \begin{bmatrix} 1.64 & 1.5 & -0.94 \\ 1.5 & 2.25 & -0.45 \\ -0.94 & -0.45 & 0.73 \end{bmatrix} \quad \begin{matrix} \lambda_1' = 3.77 \\ \lambda_2' = 0.84 \\ \lambda_3' = 0 \end{matrix} \quad V = \begin{bmatrix} -0.63 & 0.38 & 0.67 \\ -0.71 & -0.62 & -0.31 \\ 0.30 & -0.68 & 0.67 \end{bmatrix}$$

$$AA^{\mathrm{T}} = \begin{bmatrix} 1.28 & -1.04 \\ -1.04 & 3.34 \end{bmatrix} \quad \begin{matrix} \lambda_1' = 3.77 \\ \lambda_2' = 0.84 \end{matrix} \quad U = \begin{bmatrix} 0.39 & -0.92 \\ -0.92 & -0.39 \end{bmatrix}$$

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

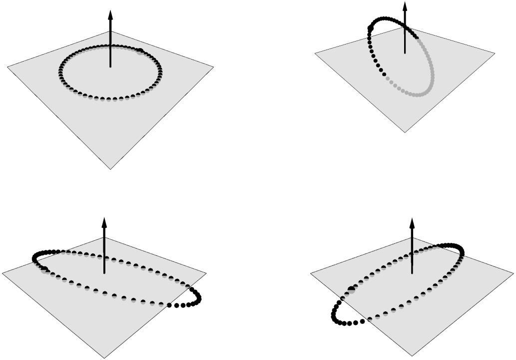
SVD: $A = U\Sigma V^{\mathrm{T}}$

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

$m < n \;\Rightarrow\; \mathbf{v}_3$ in null space of $A \;\Rightarrow\; A\mathbf{v}_3 = \mathbf{0}$

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)$V^{\mathrm{T}}\mathbf{x}$ 3)$\Sigma V^{\mathrm{T}}\mathbf{x}$ 4)$U\Sigma V^{\mathrm{T}}\mathbf{x}$

Example: a projection into the $[\mathbf{e}_1, \mathbf{e}_2]$-plane — a rank deficient matrix

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

A is symmetric and idempotent $\Rightarrow A = A^{\mathrm{T}}A = AA^{\mathrm{T}}$
$A = U\Sigma V^{\mathrm{T}}$:

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

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
$\mathbf{e}_3$ vector projected to the zero vector $\Rightarrow$ spans the null space of A and A^{T}

SVD Steps

$A = U\Sigma V^{\mathrm{T}}$

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

Input: an $m \times n$ matrix A

Output: U, V, Σ such that $A = U\Sigma V^{\mathrm{T}}$

1. Find the *eigenvalues* $\lambda'_1, \ldots, \lambda'_n$ of $A^{\mathrm{T}}A$
   - Order the λ'_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 Σ with $\sigma_{i,i} = \sqrt{\lambda'_i}, i = 1, \ldots, r$
3. Find the corresponding (normalized) eigenvectors $\mathbf{v}_i$ of $A^{\mathrm{T}}A$
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 $\mathbf{u}_i, i = 1, r$ as $\mathbf{u}_i = A\mathbf{v}_i / \| \cdot \|$
  If $m > n$ then the remaining $\mathbf{u}_i$ are found from the null space of A^{T}

- The only "hard" task is finding the λ_i'
  Since $A^{\mathrm{T}}A$ is symmetric $\Rightarrow$ Can choose a highly efficient algorithm

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

Singular Values and Volumes

Application: compute the *determinant*

$$\det U = \pm 1 \quad \text{and} \quad \det V = \pm 1 \quad \Rightarrow \quad |\det A| = \det \Sigma = \sigma_1 \cdot \ldots \cdot \sigma_n$$

Example: given a 2D triangle T with area φ
Transform $T \to T'$ with 2D linear map with singular values σ_1, σ_2
Area of $T' = \pm \sigma_1 \sigma_2 \varphi$

Example: given a 3D object O with volume φ
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

$$\det A = \lambda_1 \cdot \ldots \cdot \lambda_n$$

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 psuedoinverse $A^\dagger$
— For general matrices
— Suited for practical use
— Can be computed with the SVD

Given an $m \times n$ diagonal matrix Σ with diagonal elements σ_i
The pseudoinverse: the $n \times m$ matrix $\Sigma^\dagger$ with

$$\sigma_i^\dagger = \left\{ \begin{array}{ll} 1/\sigma_i & \text{if} \quad \sigma_i > 0 \\ 0 & \text{else} \end{array} \right\}$$

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

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

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

Properties:

$$A^\dagger A A^\dagger = A^\dagger \qquad \text{and} \qquad A A^\dagger A = A$$

Often times called the Moore-Penrose generalized inverse

Primary application: *least squares approximation*

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

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

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 $m \geq n$

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

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

The least squares solution finds the orthogonal projection of $\mathbf{b}$ into $\mathcal{V}$
— Call this projection $\mathbf{b}'$
$\Rightarrow$ Solution to $A\mathbf{x} = \mathbf{b}'$ produces vector closest to $\mathbf{b}$ that lives in $\mathcal{V}$

Normal equations

$$A^{\mathrm{T}}A\mathbf{x} = A^{\mathrm{T}}\mathbf{b} \quad \text{solution minimizes} \quad \|A\mathbf{x} - \mathbf{b}\|$$

This system can be ill-posed $\Rightarrow$ use *pseudoinverse*

$$\mathbf{x} = A^{\dagger}\mathbf{b}$$

Least Squares

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

Find $\mathbf{x}$ to minimize $\|A\mathbf{x} - \mathbf{b}\|$

$$\begin{aligned} A\mathbf{x} - \mathbf{b} &= U\Sigma V^{\mathrm{T}}\mathbf{x} - \mathbf{b} \\ &= U\Sigma V^{\mathrm{T}}\mathbf{x} - UU^{\mathrm{T}}\mathbf{b} \\ &= U(\Sigma\mathbf{y} - \mathbf{z}) \end{aligned}$$

This new framing of the problem exposes that

$$\|A\mathbf{x} - \mathbf{b}\| = \|\Sigma\mathbf{y} - \mathbf{z}\|$$

$\Rightarrow$ an easier diagonal least squares problem to solve

Steps:

1. Compute the SVD $A = U\Sigma V^{\mathrm{T}}$
2. Compute the $m \times 1$ vector $\mathbf{z} = U^{\mathrm{T}}\mathbf{b}$
3. Compute the $n \times 1$ vector $\mathbf{y} = \Sigma^{\dagger}\mathbf{z}$
   — Least squares solution to $m \times n$ problem $\Sigma\mathbf{y} = \mathbf{z}$

   requires minimizing
   $\mathbf{v} = \Sigma\mathbf{y} - \mathbf{z}$

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

$$\mathbf{v} = \begin{bmatrix} \sigma_1 y_1 - z_1 \\ \sigma_2 y_2 - z_2 \\ \vdots \\ \sigma_r y_r - z_r \\ -z_{r+1} \\ \vdots \\ -z_m \end{bmatrix}$$

   $\mathbf{y}$ minimizing $\mathbf{v}$: $y_i = z_i/\sigma_i \quad i = 1, \ldots, r \quad \Rightarrow \quad \mathbf{y} = \Sigma^{\dagger}\mathbf{z}$

4. Compute the $n \times 1$ solution vector $\mathbf{x} = V\mathbf{y}$

Least Squares

Summarize — The calculations in reverse order include

$$\mathbf{x} = V\mathbf{y}$$
$$\mathbf{x} = V(\Sigma^{\dagger}\mathbf{z})$$
$$\mathbf{x} = V\Sigma^{\dagger}(U^{\mathrm{T}}\mathbf{b})$$

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

$$\begin{bmatrix} 0 & 1 \\ 10 & 1 \\ 20 & 1 \\ 30 & 1 \\ 40 & 1 \\ 50 & 1 \\ 60 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 30 \\ 25 \\ 40 \\ 40 \\ 30 \\ 5 \\ 25 \end{bmatrix}$$

Step 1) Compute the SVD $A = U\Sigma V^{\mathrm{T}}$

$$\Sigma = \begin{bmatrix} 95.42 & 0 \\ 0 & 1.47 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \Sigma^{\dagger} = \begin{bmatrix} 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.68 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

U 7×7
Σ 7×2
V 2×2

Step 2) $\mathbf{z} = U^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} 54.5 \\ 51.1 \\ 3.2 \\ -15.6 \\ 9.6 \\ 15.2 \\ 10.8 \end{bmatrix}$

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

Step 4) $\mathbf{x} = V\mathbf{y} = \begin{bmatrix} -0.23 \\ 34.8 \end{bmatrix}$ $\Rightarrow$ best fit line: $x_2 = -0.23x_1 + 34.8$

Least Squares

The normal equations give a best approximation

$$\mathbf{x} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b} \quad \text{to the original problem} \quad A\mathbf{x} = \mathbf{b}$$

by considering $\mathbf{b}'$ in the subspace of A called $\mathcal{V}$
Substitute this expression for $\mathbf{x}$ into $A\mathbf{x} = \mathbf{b}'$:

$$\mathbf{b}' = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b} = AA^{\dagger}\mathbf{b} = \mathrm{proj}_{\mathcal{V}}\mathbf{b}$$

— Goal is to project $\mathbf{b}$ into $\mathcal{V}$ $\Rightarrow$ $AA^{\dagger}$ 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 σ_i

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

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

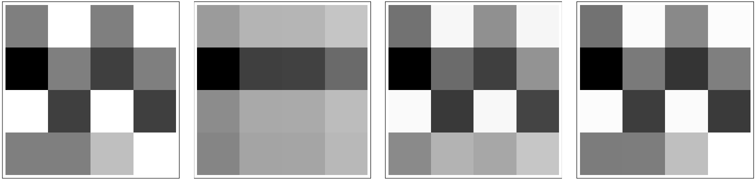
$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathrm{T}} + \ldots + \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}$$

Use this for image compression

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

— Figure (left): input image with 4×4 pixels

— Each grayscale associated with a number $\Rightarrow$ grid is a matrix



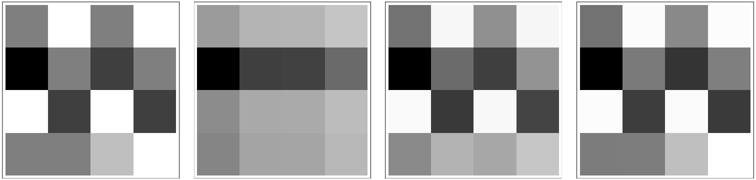
Application: Image Compression

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

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}} \quad \Rightarrow \quad \text{image } I_1$$
$$A_2 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathrm{T}} \quad \Rightarrow \quad \text{image } I_2$$

Original image nearly replicated incorporating only half the singular values
$\Rightarrow \sigma_1$ and σ_2 large in comparison to σ_3 and σ_4

Image I_3 created from $A_3 = A_2 + \sigma_3 \mathbf{u}_3 \mathbf{v}_3^{\mathrm{T}}$
Image I_4 is not displayed — identical to I_0



Application: Image Compression

The change in an image by adding the smallest σ_i can be hardly noticeable
$\Rightarrow$ Omitting images I_k corresponding to small σ_k amounts to compressing the original image

Chapter introduction Figure: 8×8 matrix
$\sigma_i = 6.2,\ 1.7,\ 1.49,\ 0, \ldots, 0$
— Figure illustrates images corresponding to each non-zero σ_i
— Last image is identical to the input
$\Rightarrow$ the five remaining $\sigma_i = 0$ are unimportant to image quality

Principal Components Analysis

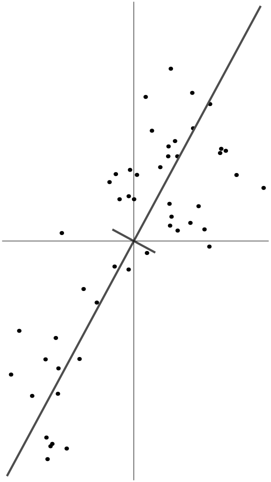
Scatter plot: data pairs recorded in Cartesian coordinates

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 $\mathbf{x}_1, \ldots, \mathbf{x}_n$
such that $\mathbf{x}_1 + \ldots + \mathbf{x}_n = \mathbf{0}$

Let $\mathbf{d}$ be a unit vector
Project $\mathbf{x}_i$ onto line containing $\mathbf{d}$
Result:
vector with (signed) length $\mathbf{x}_i \cdot \mathbf{d}$

$$l(\mathbf{d}) = [\mathbf{x}_1 \cdot \mathbf{d}]^2 + \ldots + [\mathbf{x}_n \cdot \mathbf{d}]^2$$

Rotate $\mathbf{d}$ around the origin
For each position compute $l(\mathbf{d})$
Directions corresponding to largest and smallest $l(\mathbf{d})$ are orthogonal
$\Rightarrow$ indicates variation in data

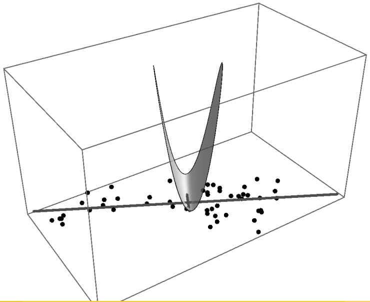
Principal Components Analysis

Arrange data $\mathbf{x}_i$ in a matrix

$$X = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{bmatrix} \quad \text{then} \quad l(\mathbf{d}) = \|X\mathbf{d}\|^2 = (X\mathbf{d}) \cdot (X\mathbf{d}) = \mathbf{d}^{\mathrm{T}} X^{\mathrm{T}} X \mathbf{d} \quad (*)$$

Let $C = X^{\mathrm{T}} X$ $\qquad$ C is a symmetric positive definite 2×2 matrix

$\Rightarrow$ (*) is a *quadratic form* — See Figure



Principal Components Analysis

For which $\mathbf{d}$ is $I(\mathbf{d})$ maximal?

Answer: $\mathbf{d}$ that corresponds to C's dominant eigenvector
And: $I(\mathbf{d})$ is minimal for $\mathbf{d}$ 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

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

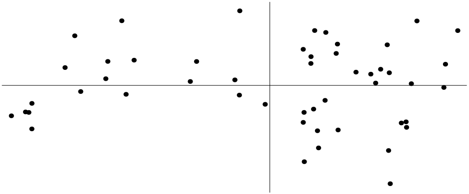
If each element of C is divided by n it is called the covariance matrix
— Summary of the variation in each coordinate and between coordinates
— Dividing by n will result in scaled eigenvalues
eigenvectors will not change

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 $V = [\mathbf{v}_1 \ \mathbf{v}_2]$ hold the normalized eigenvectors as column vectors
— $\mathbf{v}_1$ is the dominant eigenvector
Orthogonal transformation of the data X
— aligns $\mathbf{v}_1$ with $\mathbf{e}_1$ and $\mathbf{v}_2$ with $\mathbf{e}_2$

$$\hat{X} = XV \quad \Rightarrow \quad \hat{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i \cdot \mathbf{v}_1 \\ \mathbf{x}_i \cdot \mathbf{v}_2 \end{bmatrix}$$



Principal Components Analysis

Summary:

• Established a principal components coordinate system

— Defined by the eigenvectors of the covariance matrix

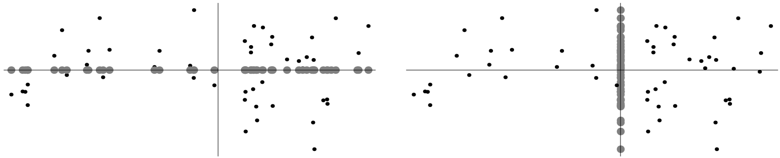
— Greatest variance corresponds to the first coordinate

• Data coordinates are now in terms of the trend lines

— Coordinates directly measure the distance from each trend line

$\Rightarrow$ 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 $V = \mathbf{v}_1$ (left Figure)
— Example: $V = \mathbf{v}_2$ (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

WYSK

- 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