Linear Algebra

Lecture Notes

Rostyslav Hryniv

Ukrainian Catholic University
Business Analytics and Computer Science Programmes

4^{th} term
Spring 2019

Lecture 13. Singular Value Decomposition

Outline

The Spectral Theorem

- Holds for symmetric/Hermitian, skew-Hermitian, orthogonal/unitary matrices
- claims existence of an orthogonal basis of eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ for eigenvalues $\lambda_1, \ldots, \lambda_n$
- spectral decomposition:

$$\boxed{A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\top + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^\top}$$

- orthogonally diagonalizable: with
  - $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$
  - P with columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$

$$P^{-1} A P = P^\top A P = \Lambda \iff A = P \Lambda P^\top$$

- λ_j are
  - real for symmetric/Hermitian matrices
  - purely imaginary for anti-symmetric/skew-Hermitian ones
  - unimodular for orthogonal/unitary matrices

Applications of the Spectral Theorem

- to construct symmetric or anti-symmetric or orthogonal (Hermitian, skew-Hermitian or unitary) matrix with prescribed spectrum and eigenvectors
- to construct low-rank approximation of A:
  - if $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ and $\lambda_{k+1}, \ldots, \lambda_n$ are small compared to $\lambda_1, \ldots, \lambda_k$, then
  $$A_k = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\top + \cdots + \lambda_k \mathbf{u}_k \mathbf{u}_k^\top$$
  is a good rank k approximation of A
  - in what norm? use the so-called Frobenius norm
  $$\|A\|_F^2 := \sum_{j,k=1}^{n} |a_{jk}|^2 = \mathrm{tr}(A^* A)$$
  - then $\|A\|_F^2 = \sum_{j=1}^n \lambda_j^2 \,\mathrm{tr}(\mathbf{u}_j \mathbf{u}_j^\top) = \sum_{j=1}^n \lambda_j^2$

    and $\|A - A_k\|_F^2 = \sum_{j=k+1}^n \lambda_j^2$
  - in fact, the SVD says A_k is the best rank k approximation!

Example of low-rank approximation

Let

$$A = \begin{pmatrix} 15 & 10 \\ 10 & 0 \end{pmatrix}$$

- Eigenvalues: $\lambda_1 = 20$, $\lambda_2 = -5$ ($\lambda_1 + \lambda_2 = 15$, $\lambda_1 \lambda_2 = -100$)
- eigenvectors: $\mathbf{u}_1 = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)^\top$, $\mathbf{u}_2 = \left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)^\top$
- rank-one approximation:

$$A_1 = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^\top = \begin{pmatrix} 16 & 8 \\ 8 & 4 \end{pmatrix}, \qquad A - A_1 = \begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}$$

- Frobenius norms:
  - $\|A\|_F^2 = 15^2 + 10^2 + 10^2 = 425$,
  - $\|A - A_1\|_F^2 = 25$ is just $1/17$ of $\|A\|_F^2$

Why low-rank?

Why low-rank matrix approximation are important?

- In reality, deal with huge matrices (sizes 10^3–10^6 or larger)
- Sending and efficient storing becomes an issue!
- Low-rank approximations are much easier for storing and sending!

Cost comparison:

- full $m \times n$ matrix requires mn numbers to store;
- rank one matrix requires only $m + n + 1$
- important e.g. for image compressions

What if A is non-symmetric or non-square?

- If A is non-symmetric, then
  - its eigenvectors need not be orthogonal, or even
  - too few eigenvectors (have to use generalized EVc's)
- If A is non-square, there are no eigenvalues and eigenvectors at all!

However, low rank approximations in (Frobenius norm) exist;
what is the best one?

Best rank one approximation to a generic A

Problem

What is the best rank one approximation $\mathbf{u}\mathbf{v}^\top$ of an $m \times n$ matrix A in Frobenius norm? (WLOG assume $\|\mathbf{v}\| = 1$)

The matrix $\mathbf{u}\mathbf{v}^\top$ has rows $u_1\mathbf{v}^\top, \ldots, u_m\mathbf{v}^\top$, if the rows of A are $\mathbf{a}_1^\top, \ldots, \mathbf{a}_m^\top$, then

$$\|A - \mathbf{u}\mathbf{v}^\top\|_F^2 = \sum_{j=1}^{m} \|\mathbf{a}_j^\top - u_j\mathbf{v}^\top\|^2 = \sum_{j=1}^{m} \|\mathbf{a}_j - u_j\mathbf{v}\|^2$$

This is minimal if $u_j\mathbf{v}$ is the projection $P_{\|}\mathbf{a}_j$ of $\mathbf{a}_j$ onto $\mathrm{ls}(\mathbf{v})$:

$$\sum_{j=1}^{m} \|\mathbf{a}_j - P_{\|}\mathbf{a}_j\|^2 = \sum_{j=1}^{m} \|P_\perp\mathbf{a}_j\|^2 = \sum_{j=1}^{m} \|\mathbf{a}_j\|^2 - \sum_{j=1}^{m} \|P_{\|}\mathbf{a}_j\|^2$$

Thus need to maximize

$$\sum_{j=1}^{m} \|P_{\|}\mathbf{a}_j\|^2 = \sum_{j=1}^{m} |\mathbf{a}_j^\top\mathbf{v}|^2 = \sum_{j=1}^{m} |\mathbf{v}^\top\mathbf{a}_j|^2 = \|A\mathbf{v}\|^2$$

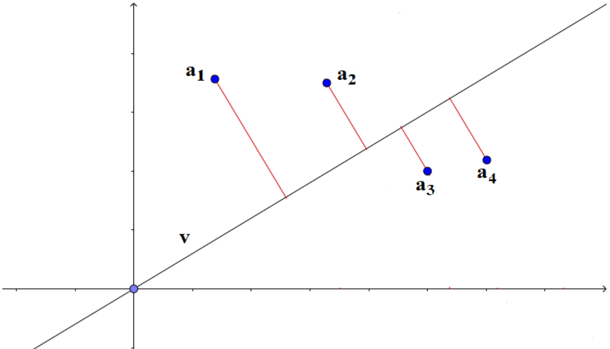
The trolley-line-location problem

We reduced the above problem to the following one:

Maximize $\|A\mathbf{v}\|$ under the restriction that $\|\mathbf{v}\| = 1$

This is what we get in the trolley-line-location problem:

Choose a direction $\mathbf{v}$ to minimize the sum of squared distances from $\mathbf{a}_1, \ldots, \mathbf{a}_m$ to the line



The trolley line problem

Problem:

For the given vectors $\mathbf{a}_1, \mathbf{a}_2, \dots \mathbf{a}_m$ in $\mathbb{R}^n$, find their best line fit ℓ
The objective function to be minimized:

$$f(\ell) := \sum_{k=1}^{m} \operatorname{dist}^2(\mathbf{a}_j, \ell)$$

- $\mathbf{v}$ is the unit vector on ℓ and $P_{\mathbf{v}} := \mathbf{v}\mathbf{v}^\top$ the orthogonal projector;
- then $\operatorname{dist}(\mathbf{a}_k, \ell) = \|\mathbf{a}_k - P_{\mathbf{v}}\mathbf{a}_k\|$, so that

$$f(\ell) = \sum \|\mathbf{a}_k - P_{\mathbf{v}}\mathbf{a}_k\|^2 = \sum \|\mathbf{a}_k\|^2 - \sum \|P_{\mathbf{v}}\mathbf{a}_k\|^2$$

- thus one needs to maximize the sum

$$\sum \|P_{\mathbf{v}}\mathbf{a}_k\|^2 = \sum \|\mathbf{v}\mathbf{v}^\top\mathbf{a}_k\|^2 = \sum |\mathbf{a}_k^\top\mathbf{v}|^2 = \|A\mathbf{v}\|^2,$$

where A has rows $\mathbf{a}_1^\top, \mathbf{a}_2^\top, \dots, \mathbf{a}_m^\top$

Solution to the rank-one approximation problem

Consider the quadratic form

$$Q(\mathbf{v}) := \|A\mathbf{v}\|^2 = (A\mathbf{v})^\top (A\mathbf{v}) = \mathbf{v}^\top A^\top A\mathbf{v}$$

and denote

- the largest eigenvalue by σ_1^2
- corresponding eigenvector (the first principal axis) by $\mathbf{v}_1$

then $A^\top A\mathbf{v}_1 = \sigma_1^2 \mathbf{v}_1$

$$\max\{Q(\mathbf{v}) \mid \|\mathbf{v}\| = 1\} = Q(\mathbf{v}_1) = \sigma_1^2$$

and $\mathbf{u}_1 := A\mathbf{v}_1$ satisfies $A^\top \mathbf{u}_1 = \sigma_1^2 \mathbf{v}_1$

Solution to the rank-one approximation problem:

In Frobenius norm, the best rank-one approximation of A is $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$

This leads to the notion of singular values of A

Definition

- Let A be any $m \times n$ matrix
- $B := A^\top A$ is $n \times n$ and nonnegative:

$$\mathbf{x}^\top B\mathbf{x} = \mathbf{x}^\top (A^\top A)\mathbf{x} = (A\mathbf{x})^\top (A\mathbf{x}) = \|A\mathbf{x}\|^2 \geq 0$$

- denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ the EV's of B
- $\sigma_j := \sqrt{\lambda_j}$ are called the singular values of A
- notice that there are $r := \operatorname{rank} B = \operatorname{rank} A$ positive σ_j

Example

$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$; $B = A^\top A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has EV's $\lambda_1 = 3$ and $\lambda_2 = 1$; thus $\sigma_1 = \sqrt{3}, \sigma_2 = 1$

SVD theorem

Theorem (SVD)

Every $m \times n$ matrix A can be written as

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

where U and V are orthogonal and Σ is an $m \times n$ matrix with singular values of A on its main diagonal and zeros otherwise

Remark

This is an analogue of the diagonalization $A = UDU^\top$ of a symmetric matrix A

SVD theorem

Theorem (SVD — expanded form)

Every $m \times n$ matrix A of rank r can be written as $A = U\Sigma V^\top$, where

$$U = (\mathbf{u}_1 \dots \mathbf{u}_r | \mathbf{u}_{r+1} \dots \mathbf{u}_m),$$

$$V = (\mathbf{v}_1 \dots \mathbf{v}_r | \mathbf{v}_{r+1} \dots \mathbf{v}_n),$$

- *Σ has σ_j on its main diagonal and zeros otherwise*
- *$\mathbf{v}_j$ are eigenvectors of $A^\top A$ with EV's σ_j^2:* $\quad A^\top A\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j$
- *$\mathbf{u}_j := A\mathbf{v}_j / \|A\mathbf{v}_j\| = A\mathbf{v}_j/\sigma_j$ for $j = 1, \dots, r$ is an ONB for the column space of A*
- *$\mathbf{u}_1, \dots, \mathbf{u}_m$ is an ONB for $\mathbb{R}^m$*
- $$\boxed{A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \dots + \sigma_r \mathbf{u}_r \mathbf{v}_r^\top}$$

The vectors $\mathbf{u}_1, \dots, \mathbf{u}_r$ are the left singular vectors of A;
$\mathbf{v}_1, \dots, \mathbf{v}_r$ are the right singular vectors of A

Remark: $\quad A\mathbf{v}_j = \sigma_j \mathbf{u}_j$, $A^\top \mathbf{u}_j = \sigma_j \mathbf{v}_j$

Proof of the SVD decomposition

- Start with normalized eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and eigenvalues $\sigma_1^2, \ldots, \sigma_n^2$ of $A^\top A$
- Then $\|A\mathbf{v}_j\|^2 = \mathbf{v}_j^\top A^\top A\mathbf{v}_j = \sigma_j^2$, $j = 1, \ldots, n$
- Form $\mathbf{u}_j := A\mathbf{v}_j / \|A\mathbf{v}_j\| = A\mathbf{v}_j/\sigma_j, \quad j = 1, \ldots, r (= \operatorname{rank} A)$
- $\mathbf{u}_i^\top \mathbf{u}_j = \mathbf{v}_i^\top A^\top A\mathbf{v}_j/(\sigma_i\sigma_j) = \mathbf{v}_i^\top (A^\top A)\mathbf{v}_j/(\sigma_i\sigma_j) = \frac{\sigma_j}{\sigma_i}\mathbf{v}_i^\top \mathbf{v}_j = \delta_{ij}$
- complete with $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ to an ONB of $\mathbb{R}^m$
- Now

$$\begin{aligned} U\Sigma &= (\sigma_1\mathbf{u}_1 \ \ldots \ \sigma_r\mathbf{u}_r \ \underbrace{0\ldots0}_{n-r}) \\ &= (A\mathbf{v}_1 \ \ldots \ A\mathbf{v}_r \ \underbrace{0\ldots0}_{n-r}) = A(\mathbf{v}_1 \ \ldots \ \mathbf{v}_n) = AV \end{aligned}$$

- since V is orthogonal, $VV^\top = I$ yields $A = U\Sigma V^\top$

Example

For $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$, we find that

- $\sigma_1 = \sqrt{3}$ and $\sigma_1 = 1$
- $\mathbf{v}_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$ and $\mathbf{v}_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^\top$
- $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(\sqrt{2}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$,
  $\mathbf{u}_2 = (0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^\top$,
  $\mathbf{u}_3 = \frac{1}{\sqrt{3}}(-1, 1, 1)^\top$
- $$\sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^\top = \begin{pmatrix} 1 & 1 \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = A$$
- $\sigma_1 \mathbf{u}_1 \mathbf{v}_1^\top$ is the best rank one approximation of A in the Frobenius norm $\sum (a_{ij} - b_{ij})^2$

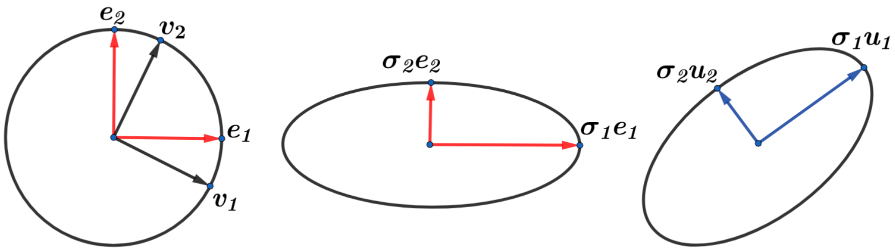
Interpretation of SVD

$A = U\Sigma V^\top$ implies decomposition of $\mathbf{x} \mapsto A\mathbf{x}$ into

$$\mathbf{x} \mapsto \mathbf{y} := V^\top \mathbf{x}, \qquad \mathbf{y} \mapsto \mathbf{z} := \Sigma \mathbf{y}, \qquad \mathbf{z} \mapsto A\mathbf{x} = U\mathbf{z}$$

- $\mathbf{x} \mapsto \mathbf{y}$ finds the coordinates of the vector $\mathbf{x}$ in terms of one orthonormal basis $(\mathbf{v}_1, \dots, \mathbf{v}_n)$
- $\mathbf{y} \mapsto \mathbf{z}$ scales those coordinates
- $\mathbf{z} \mapsto A\mathbf{x}$ find the vector with the scaled coordinates over another orthonormal basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$

Interpretation of SVD



$$\mathbf{v}_1 \mapsto \mathbf{e}_1 \mapsto \sigma_1 \mathbf{e}_1 \mapsto \sigma_1 \mathbf{u}_1$$
$$\mathbf{v}_2 \mapsto \mathbf{e}_2 \mapsto \sigma_2 \mathbf{e}_2 \mapsto \sigma_2 \mathbf{u}_2$$

Reduced SVD

- In the SVD representation, some part is uninformative:
  - $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are chosen arbitrarily in the nullspace of A
  - $\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$ are chosen arbitrarily in the nullspace of A^T
  - Σ has zero rows or columns
- The reduced SVD removes that uninformative part:

$$A = \underbrace{\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{pmatrix}}_{m \times r} \underbrace{\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \cdots & \sigma_r \end{pmatrix}}_{r \times r} \underbrace{\begin{pmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_r^\top \end{pmatrix}}_{r \times n}$$

- The reduced SVD of $A^\top$:

$$A^\top = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \cdots & \sigma_r \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^\top \\ \vdots \\ \mathbf{u}_r^\top \end{pmatrix}$$

Polar decomposition

Theorem (Polar decomposition)

Any square matrix A can be written as QS with orthogonal Q and symmetric positive semidefinite S

Why polar?

$z = re^{i\theta}$

Proof.

Write $A = U\Sigma V^\top = (UV^\top)(V\Sigma V^\top) =: QS$

$Q := UV^\top$ is orthogonal

$S := V\Sigma V^\top$ is symmetric and positive semidefinite $\square$

Image compression

Image compression

Instead of storing $m \times n$ numerical entries, can take the best rank-r approximation of A; need

$$r(1 + m + n)$$

numbers

Pseudo-inverse

A rectangular A cannot be inverted!

However, a pseudo-inverse A^+ can be defined s.t.

$$A^+A \approx I_n \text{ and } AA^+ \approx I_m \implies AA^+A \approx A \text{ and } A^+AA^+ \approx A^+$$

Definition (Moore–Penrose pseudo-inverse)

For an $m \times n$ matrix A, its Moore-Penrose pseudo-inverse is an $n \times m$ matrix A^+ satisfying

$$A^+A\,A^+ = A^+, \qquad AA^+\,A = A, \qquad (A^+A)^\top = A^+A, \qquad (AA^+)^\top = AA^+$$

Theorem

For every matrix A, its Moore–Penrose pseudo-inverse A^+ exists and is unique

Pseudo-inverse

Moore–Penrose pseudo-inverse

If $A = U\Sigma V^\top$ is the SVD of A, then the pseudo-inverse Σ^+ of Σ should satisfy $\Sigma^+\Sigma = I_r \oplus \mathbf{0}_{n-r}, \quad \Sigma\Sigma^+ = I_r \oplus \mathbf{0}_{m-r}$
Thus Σ^+ gets transposed and σ_j replaced with $1/\sigma_j$

Therefore, Moore–Penrose pseudo-inverse is $A^+ := V\Sigma^+U^\top$:

$$A^+A = V\Sigma^+(U^\top U)\Sigma V^\top = V\Sigma^+\Sigma V^\top = V(I_r \oplus \mathbf{0}_{n-r})V^\top$$
$$AA^+A = U\Sigma V^\top V(I_r \oplus \mathbf{0}_{n-r})V^\top = U\Sigma V^\top = A$$
$$A^+AA^+ = V(I_r \oplus \mathbf{0}_{n-r})V^\top V\Sigma^+U^\top = V\Sigma^+U^\top = A^+$$

Example

- If a rectangular A has linearly independent columns, then
$$A^+ = (A^\top A)^{-1}A^\top$$
is the left inverse of A (ie, $A^+A = I$) and AA^+ is an old friend...
- In terms of $A = QR$, $A^+ = R^{-1}Q^\top$

Theorem

$\hat{\mathbf{x}} = A^+\mathbf{b}$ is the best solution of $A\mathbf{x} = \mathbf{b}$

- A is invertible $\implies A^+ = A^{-1}$, and $\hat{\mathbf{x}}$ is the unique exact solution
- A has linearly independent columns $\implies A^+ = (A^\top A)^{-1}A^\top$;
  - if $\mathbf{b}$ is in the column space, then A^+ is the right inverse of A, and the unique exact solution $\mathbf{x}$ satisfies $\mathbf{x} = A^+A\mathbf{x} = A^+\mathbf{b} = \hat{\mathbf{x}}$
  - if $\mathbf{b}$ is not in the column space, then $\hat{\mathbf{x}} := A^+\mathbf{b}$ is the unique least square solution
- A has linearly dependent columns $\implies$ a solution (exact when $\mathbf{b} \in \mathrm{col}(A)$ or least square otherwise) is not unique
  - $\hat{\mathbf{x}} = A^+\mathbf{b}$ is then the shortest solution (ie, of the smallest norm)
  - Indeed, if $A = U\Sigma V^\top$, then $\|A\mathbf{x} - \mathbf{b}\| = \|\Sigma V^\top\mathbf{x} - U^\top\mathbf{b}\| = \|\Sigma\mathbf{y} - U^\top\mathbf{b}\|$ with $\mathbf{y} := V^\top\mathbf{x}$
  - $\Sigma\mathbf{y} - U^\top\mathbf{b}$ has the smallest norm when its first $r = \mathrm{rank}\, A$ entries are zero; the rest do not depend on $\mathbf{y}$ and are equal to those of $-U^\top\mathbf{b}$
  - this specifies the first r entries of $\mathbf{y}$ and leave the rest undefined
  - $\hat{\mathbf{y}} := \Sigma^+U^\top\mathbf{b}$ has the required first r entries and all the rest entries zero $\implies$ is of the shortest norm among all such $\mathbf{y}$
  - $\mathbf{x} = V\hat{\mathbf{y}} = V\Sigma^+U^\top\mathbf{b} = \hat{\mathbf{x}}$ is then the shortest one among all solutions

SVD vs PCA

- Observe that the largest value of $\|A\mathbf{x}\|$ with $\|\mathbf{x}\| \leq 1$ is obtained for $\mathbf{x} = \mathbf{v}_1$ and is equal to σ_1;
- this is the first principal axis for $A^T A$:
  - indeed, $A^T A = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T = VDV^T$ is the spectral decomposition of the symmetric matrix $B := A^T A$
  - B has eigenvalues σ_k^2 with eigenvectors $\mathbf{v}_k$
  - the quadratic form $Q(\mathbf{x}) := \mathbf{x}^T B\mathbf{x}$ is equal to $\|A\mathbf{x}\|^2$
  - by the minimax properties of the eigenvalues,

$$\begin{aligned}
\sigma_1^2 &= \max_{\|\mathbf{x}\|=1} \mathbf{x}^T B\mathbf{x} = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|^2, \\
\sigma_2^2 &= \max_{\substack{\|\mathbf{x}\|=1, \\ \mathbf{x}\perp\mathbf{v}_1}} \mathbf{x}^T B\mathbf{x} = \max_{\substack{\|\mathbf{x}\|=1, \\ \mathbf{x}\perp\mathbf{v}_1}} \|A\mathbf{x}\|^2, \\
\sigma_3^2 &= \dots
\end{aligned}$$

- $A^T A$ can be considered as a covariance matrix for the columns of A

The trolley line problem, revisited

Problem:

For the given vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_m$ in $\mathbb{R}^n$, find the best-fit subspace L of dimension k. The objective function to be minimized:

$$f(L) := \sum_{j=1}^{m} \mathrm{dist}^2(\mathbf{a}_j, L)$$

- $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is an ONB of L and $P_L := \sum \mathbf{u}_\ell \mathbf{u}_\ell^\top$ the orthoprojector;
- then $\mathrm{dist}(\mathbf{a}_j, L) = \|\mathbf{a}_j - P_L \mathbf{a}_j\|$, so that
$$f(L) = \sum \|\mathbf{a}_j - P_L \mathbf{a}_j\|^2 = \sum \|\mathbf{a}_j\|^2 - \sum \|P_L \mathbf{a}_j\|^2$$
- thus one needs to maximize the sum
$$\sum \|P_L \mathbf{a}_j\|^2 = \sum_{j,\ell} \|\mathbf{u}_\ell \mathbf{u}_\ell^\top \mathbf{a}_j\|^2 = \sum_{j,\ell} |\mathbf{a}_j^\top \mathbf{u}_\ell|^2 = \sum_\ell \|A\mathbf{u}_\ell\|^2,$$
where A has rows $\mathbf{a}_1^\top, \mathbf{a}_2^\top, \ldots, \mathbf{a}_m^\top$
- Solution: the subspace spanned by the first k right singular vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of A
- Indeed, $\|A\mathbf{u}\|^2 = \|\Sigma V^\top \mathbf{u}\|^2 = \|\Sigma \mathbf{w}\|^2 = \sum \sigma_1^2 w_1^2 + \cdots + \sigma_r^2 w_r^2 \ldots$

Best low-rank approximation of A

Frobenius norm of a matrix

$$\|A\|_F^2 = \sum_{i,j} |a_{ij}|^2 = \sum_i \|\mathbf{a}_i\|^2 = \sum_i \sigma_i^2$$

Indeed,

- pre-/post-multiplying by an orthogonal matrix does not change $\|\cdot\|_F$
- thus $A = U\Sigma V^T$ yields $\|A\|_F^2 = \|U^T A V\|_F^2 = \|\Sigma\|_F^2$
- another reason: $\|A\|_F^2 = \mathrm{tr}(A^T A)$; now
$$\mathrm{tr}(A^T A) = \mathrm{tr}(V\Sigma^T\Sigma V^T) = \mathrm{tr}(\Sigma^T\Sigma) = \sum \sigma_k^2$$

Best rank-one approximation of A in the Frobenius norm

For a rank-one operator $B = \mathbf{u}\mathbf{v}^T$, $\|B\|_F^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2$; thus ($\|\mathbf{u}\| = 1$)

$$\begin{aligned}\|A - \mathbf{u}\mathbf{v}^\top\|_F^2 &= \mathrm{tr}(A - \mathbf{u}\mathbf{v}^\top)^\top(A - \mathbf{u}\mathbf{v}^\top) \\ &= \ \ldots \ = \|A\|_F^2 - \|A\top\mathbf{u}\|^2 + \|A^\top\mathbf{u} - \mathbf{v}\|^2\end{aligned}$$

Thus: maximize $\|A\mathbf{u}\|$ and take $\mathbf{v} = A^\top\mathbf{u} \implies$

$$\mathbf{u} = \mathbf{u}_1 \text{ and } \mathbf{v} = A\mathbf{u}_1 = \sigma_1\mathbf{v}_1$$

What has not been covered (but could have been):

- Hamming codes as basis problem
- 2D image processing as change of basis problem
- Image rectification
- Iterative methods of solving $A\mathbf{x} = \mathbf{v}$:
  rewrite as $\mathbf{x} = B\mathbf{x} + \mathbf{x}_0$ and $\mathbf{x}_{n+1} = B\mathbf{x}_n + \mathbf{x}_0$
- Iterative methods for finding eigenvalues/eigenvectors
- PageRank as an eigenvalue problem
- Numerical issues (condition number, stability etc)
- LA and optimization problems

- and lots of other fun stuff . . .

Thanks for being with

Linear Algebra