Numerical Linear Algebra
Vector Spaces

Franck K. Mutombo
franckm@aims.ac.za

Professor of Mathematics
University of Lubumbashi

use this email to contact me:
franck.mutombo@aims.ac.rw

October 2024

Linear Algebra II

Outline

Matrices

Introduction

Matrix operations

- A matrix is a rectangular array of numbers arranged by rows and columns

Example

An $m \times n$ matrix $\mathbf{A}$, denoted $[a_{ij}]$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \tilde{\mathbf{a}}_2^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix}$$

$\mathbf{a}_j$ for $j = 1, \ldots, n$ are $m \times 1$ vectors; $\tilde{\mathbf{a}}_i$ for $i = 1, \ldots, m$ are $n \times 1$ vectors

Matrix operations

Example (Matrix-vector multiplication)

- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} x_j \end{bmatrix} = \sum_{j=1}^{n} \mathbf{a}_j x_j = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{x} \end{bmatrix} \in \mathbb{R}^m$$

- Complexity: $\sim m(2n-1) = \mathcal{O}(mn)$ floating point operations (FLOPS)

Matrix operations

Example (Matrix-matrix multiplication)

- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} \in \mathbb{R}^{n \times p}$ and $\mathbf{b}_i \in \mathbb{R}^n$

$$\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} \sum_{j=1}^{n} a_{1j} b_{1j} & \cdots & \sum_{j=1}^{n} a_{1j} b_{pj} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{mj} b_{1j} & \cdots & \sum_{j=1}^{n} a_{mj} b_{pj} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \cdots & \mathbf{Ab}_p \end{bmatrix} \\
&= \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{B} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{B} \end{bmatrix}
\end{aligned}$$

- Complexity: $\sim mp(2n-1) = \mathcal{O}(mnp)$ FLOPS

Special matrices

Definition (Identity matrix)

The identity matrix (denoted $\mathbf{I}_n \in \mathbb{R}^{n \times n}$) is a **square** matrix of zero entries except on the *main diagonal*, which has ones on it. For $\mathbf{A} \in \mathbb{R}^{m \times n}$ we have

$$\mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A}.$$

Special matrices

Definition (Identity matrix)

The identity matrix (denoted $\mathbf{I}_n \in \mathbb{R}^{n\times n}$) is a **square** matrix of zero entries except on the *main diagonal*, which has ones on it. For $\mathbf{A} \in \mathbb{R}^{m\times n}$ we have

$$\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}.$$

Definition (Inverse of a matrix)

The inverse of a **square** matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ (denoted $\mathbf{A}^{-1}$) satisfies the following.

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

- When the inverse of a matrix **exists** we say it is invertible
- When the inverse of a matrix does **not exist** we say it is non-invertible
- **Non-invertible** matrices are popularly referred to as singular matrices

Orthogonal matrices

Definition (Orthogonal (or Unitary) matrix)

A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is

orthonormal, if $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n$, $\quad$ ($\mathbf{A}^T$ denotes transpose of $\mathbf{A}$)
unitary, if $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^* = \mathbf{I}_n$, $\quad$ ($\mathbf{A}^*$ denotes conjugate transpose of $\mathbf{A}$)

- Orthonormal matrices are **orthogonal** with normalized columns and rows.
- Orthogonal matrices have several nice properties.
  1. They preserve inner products
  $$(\mathbf{Ax})^T(\mathbf{Ay}) = \mathbf{x}^T\mathbf{A}^T\mathbf{Ay} = \mathbf{x}^T\|\mathbf{y} = \mathbf{x}^T\mathbf{y}$$
  2. Hence they preserve 2-norms
  $$\|\mathbf{Ax}\|_2 = \sqrt{\mathbf{x}^T\mathbf{A}^T\mathbf{Ax}} = \sqrt{\mathbf{x}^T\mathbf{x}} = \|\mathbf{x}\|_2$$
  3. $\Rightarrow$ multiplication by orthogonal $\mathbf{A}$ is a transformation that preserves **length**

Basic matrix definitions

Definition (Nullspace of a matrix)

The **nullspace** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, (denoted by $\text{null}(\mathbf{A})$) is defined as

$$\text{null}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$$

- $\text{null}(\mathbf{A})$ is the set of vectors mapped to zero by $\mathbf{A}$.
- $\text{null}(\mathbf{A})$ is the set of vectors orthogonal to the rows of $\mathbf{A}$.

Basic matrix definitions

Definition (Range of a matrix)

The **range** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, (denoted by $\mathrm{range}(\mathbf{A})$) is defined as

$$\mathrm{range}(\mathbf{A}) = \{\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$$

- $\mathrm{range}(\mathbf{A})$ is the span of the columns (or the **column space**) of $\mathbf{A}$.
- $\mathrm{range}(\mathbf{A})$ is the set of vectors $\mathbf{y} = \mathbf{A}\mathbf{x}$ for which the system has a solution.

Basic matrix definitions

Definition (Rank of a matrix)

The **rank** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, (denoted by $\mathrm{rank}(\mathbf{A})$) is defined as the maximum number of linearly independent columns (or rows) of $\mathbf{A}$.

- $\mathrm{rank}(\mathbf{A}) = \mathbf{dim}\,(\mathrm{range}(\mathbf{A}))$
- $\mathrm{rank}(\mathbf{A}) \leq \min(m, n)$
- $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}^T)$
- $\mathrm{rank}(\mathbf{A}) + \mathbf{dim}\,(\mathrm{null}(\mathbf{A})) = \mathbf{dim}\,(\mathrm{range}(\mathbf{A})) + \mathbf{dim}\,(\mathrm{null}(\mathbf{A})) = n$

Eigenvalues & Eigenvectors

Eigenvalues & Eigenvectors

Definition (Eigenvalues & Eigenvectors)

The vector $\mathbf{x}$ is an **eigenvector** of a *square* matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ if $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ where $\lambda \in \mathbb{R}$ or $\mathbb{C}$ is called an **eigenvalue** of $\mathbf{A}$.

- In particular, λ is an eigenvalue of $\mathbf{A} \Leftrightarrow \mathbf{A} - \lambda\mathbf{I}$ is singular $\Leftrightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

Example

Compute the eigenvalues and eigenvectors of the following matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}.$$

1. By hand
2. In Python

Eigenvalues & Eigenvectors

Proposition

Let $\mathbf{x}$ be an eigenvectors of $\mathbf{A}$ with corresponding eigenvalues λ. Then

1. *For any $\gamma \in \mathbb{R}$, $\mathbf{x}$ is an eigenvector of $\mathbf{A} + \gamma\mathbf{I}$ with eigenvalue $\gamma + \lambda$.*
2. *If $\mathbf{A}$ is invertible, then $\mathbf{x}$ is an eigenvector of $\mathbf{A}^{-1}$ with eigenvalue λ^{-1}.*
3. *$\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ for any $k \in \mathbb{Z}$ (where $\mathbf{A}^0 = \mathbf{I}$ by definition).*

Proof.

See jupyter notebook □

Matrix decomposition

Definition (Eigenvalue (spectral) decomposition)

The **eigenvalue decomposition** of a square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is given by:

$$\mathbf{A} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{-1}$$

- the columns of $\mathbf{X} \in \mathbb{R}^{n \times n}$, i.e., $\mathbf{x}_i$, are eigenvectors of $\mathbf{A}$
- $\boldsymbol{\Lambda} = \mathbf{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right)$ where λ_i are eigenvalues of $\mathbf{A}$

- Decomposition on exist only if $\mathbf{X}$ is **invertible**; $\Rightarrow$ the eigenvectors of $\mathbf{A}$ are linearly independent
- When ordered $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, $\lambda_i(\mathbf{A})$ denotes the i^{th} largest eigenvalue of $\mathbf{A}$, i.e.
  - $\lambda_p(\mathbf{A}) = \lambda_{\min}(\mathbf{A})$ is the **minimum** eigenvalue of $\mathbf{A}$
  - $\lambda_1(\mathbf{A}) = \lambda_{\max}(\mathbf{A})$ is the **maximum** eigenvalue of $\mathbf{A}$

Matrix decomposition

Example

Let $\mathbf{A}$ be the following matrix (it was generated at random from normal distributions with variance 0.5)

$$\mathbf{A} = \begin{pmatrix} -0.64 & 0.14 & 1.47 & 1.16 \\ -0.28 & -0.65 & 0.97 & 0.81 \\ 0.73 & 0.23 & 0.74 & 0.65 \end{pmatrix}.$$

1. Display the four 3-dimensional vectors in a 3D plot.
2. Compute $\mathbf{B} = \mathbf{A} \cdot \mathbf{A}^T$ and its spectral decomposition.
3. Plot the eigenvalues of the $\mathbf{B}$ matrix in decreasing order.

Trace & Determinant

Trace

Definition

The trace of a square matrix, $\mathbf{A} \in \mathbb{R}^{n\times n}$, (denoted $\mathrm{tr}(\mathbf{A})$) is the sum of its diagonal entries or its eigenvalues, i.e.

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i(\mathbf{A}).$$

- The trace has several nice algebraic properties.
  1. $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$.
  2. $\mathrm{tr}(\alpha\mathbf{A}) = \alpha\mathrm{tr}(\mathbf{A})$.
  3. $\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A})$.
  4. $\mathrm{tr}(\mathbf{ABCD}) = \mathrm{tr}(\mathbf{BCDA}) = \mathrm{tr}(\mathbf{CDAB}) = \mathrm{tr}(\mathbf{DABC})$.
- 1. and 2. imply that the **trace** is a linear map.
- 4. is known as invariance under cyclic permutations.

Determinant

Definition

The determinant of a square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, (denoted $\det(\mathbf{A})$) is the product of its eigenvalues, i.e.

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i(\mathbf{A}).$$

- The determinant has several nice properties.
  1. $\det(\mathbf{I}) = 1$.
  2. $\det(\mathbf{A}^T) = \det(\mathbf{A})$.
  3. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.
  4. $\det\left(\mathbf{A}^{-1}\right) = \det(\mathbf{A})^{-1}$.
  5. $\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$.

Symmetric matrices

Symmetric matrices

Definition (Symmetric matrix)

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if $\mathbf{A} = \mathbf{A}^T$, i.e. $a_{ij} = a_{ji}$.

- Symmetric matrices have many nice properties
- Most important of all being the spectral theorem.

Theorem (Spectral Theorem)

If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, then there exists an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.

Symmetric matrices

- A practical application of theorem is the spectral decomposition.

Definition (Spectral decomposition)

The **spectral decomposition** of a square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is given by:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

- the columns of $\mathbf{Q} \in \mathbb{R}^{n \times n}$, i.e., $\mathbf{x}_i$, are eigenvectors of $\mathbf{A}$
- $\mathbf{\Lambda} = \mathbf{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right)$ where $\lambda_i \in \mathbb{R}$ are eigenvalues of $\mathbf{A}$

Problem

This decomposition is always possible, why?

Symmetric matrices

- A practical application of theorem is the spectral decomposition.

Definition (Spectral decomposition)

The **spectral decomposition** of a square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is given by:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$$

- the columns of $\mathbf{Q} \in \mathbb{R}^{n \times n}$, i.e., $\mathbf{x}_i$, are eigenvectors of $\mathbf{A}$
- $\mathbf{\Lambda} = \mathbf{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right)$ where $\lambda_i \in \mathbb{R}$ are eigenvalues of $\mathbf{A}$

Problem

This decomposition is always possible, why?

Solution

The eigenvectors are orthogonal $\Rightarrow$ *linearly independent*

Rayleigh quotients

Rayleigh quotients

- More on the properties of symmetric matrices: Rayleigh quotients

Definition (Quadratic forms)

Given a square matrix, $\mathbf{A} \in \mathbb{R}^{n\times n}$, a quadratic form is the following

$$\mathbf{x}^T\mathbf{A}\mathbf{x}$$

- An interesting connection between the quadratic form of a symmetric matrix and its eigenvalues

Definition (Rayleigh quotient)

The Rayleigh quotient (denoted by $R_{\mathbf{A}}$) is a function of $\mathbf{x}$ and is defined as

$$R_{\mathbf{A}}(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

Properties of Rayleigh quotients

(i) Scale invariance: for any $\mathbf{x} \neq \mathbf{0}$ and any $\alpha \neq 0$, $R_{\mathbf{A}}(\alpha \mathbf{x}) = R_{\mathbf{A}}(\mathbf{x})$

(ii) If $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ, then $R_{\mathbf{A}}(\mathbf{x}) = \lambda$

(ii) $R_{\mathbf{A}}(\mathbf{x})$ is **bounded** by the smallest and largest eigenvalues of $\mathbf{A}$

Theorem (Min-max theorem)

For all $\mathbf{x} \neq \mathbf{0}$

$$\lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x}) \leq \lambda_{\max}(\mathbf{A})$$

with equality if and only if $\mathbf{x}$ *is a corresponding eigenvector.*

Proof.

The proof is a corollary of the following **Proposition**. □

Properties of Rayleigh quotients

Proposition

For any $\mathbf{x}$ *such that* $\|\mathbf{x}\|_2 = 1$

$$\lambda_{\min}(\mathbf{A}) \leq \mathbf{x}^T \mathbf{A}\mathbf{x} \leq \lambda_{\max}(\mathbf{A})$$

Problem (1)

Prove the above Proposition.

- Proof of min-max theorem: follows from the **scale invariance** of $R_{\mathbf{A}}(\mathbf{x})$
- The **min-max theorem** is referred to as the variational characterization of eigenvalues for expressing the smallest (largest) eigenvalues in terms of a minimization (maximization) problem

$$\lambda_{\min\,/\,\max}(\mathbf{A}) = \min_{\mathbf{x}\neq\mathbf{0}} / \max_{\mathbf{x}\neq\mathbf{0}} \; R_{\mathbf{A}}(\mathbf{x})$$

Positive (semi-) definite matrices

Positive (semi-) definite matrices

- More on symmetric matrices: Positive (semi-) definite matrices

Definition (Positive (semi-) definite matrices)

A **symmetric matrix**, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is positive semi-definite (denoted $\mathbf{A} \succcurlyeq 0$) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \quad \text{for all } \ \mathbf{x} \in \mathbb{R}^n,$$

and is positive definite (denoted $\mathbf{A} \succ 0$) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \quad \text{for all } \ \mathbf{x} \in \mathbb{R}^n,$$

- The following **proposition** relate these properties to eigenvalues

Proposition

A symmetric matrix is positive semi-definite ***iff*** *all of its eigenvalues are nonnegative, and positive definite* ***iff*** *all of its eigenvalues are positive.*

Positive (semi-) definite matrices

Problem (2)

Prove the above Proposition.

- An example how these matrices arise

Proposition

Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then $\mathbf{A}^T\mathbf{A}$ is positive semi-definite. If null$(\mathbf{A}) = \mathbf{0}$, then $\mathbf{A}^T\mathbf{A}$ is positive definite

Proof.

Let $\mathbf{A}$ be positive semi-definite and let $\epsilon > 0$, we have for any $\mathbf{x} \neq 0$ that

$$\mathbf{x}^T(\mathbf{A} + \epsilon\mathbf{I})\mathbf{x} = \mathbf{x}^T\mathbf{A}\mathbf{x} + \epsilon\mathbf{x}^T\mathbf{I}\mathbf{x} = \mathbf{x}^T\mathbf{A}\mathbf{x} + \epsilon\|\mathbf{x}\|_2^2 > 0$$

□

- Consequence of two propositions is that $\mathbf{A} + \epsilon\mathbf{I}$ is positive definite (and invertible) for any matrix $\mathbf{A}$ and any $\epsilon > 0$.

Singular Value Decomposition

Singular Value Decomposition

- Singular value decomposition (SVD) is one of the most widely applicable tools in linear algebra

Definition (Singular Value Decomposition (SVD))

Consider a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. There exists an SVD of $\mathbf{A}$ of the form

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T,$$

where

- $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an orthogonal (unitary) matrix known as the matrix of left-singular vectors
- $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r) \in \mathbb{R}^{m \times n}$ is a diagonal matrix with non-negative real values $\sigma_i, \ i = 1, \ldots, r$ $(r = \min(m, n))$ known as the matrix of singular values,
- $\mathbf{V} \in \mathbb{R}^{n \times n}$ is an orthogonal (unitary) matrix known as the matrix of right-singular vectors

- One may think of the SVD as a generalization of the eigenvalue decomposition to non-square matrices.

Singular Value Decomposition

- Computing the SVD of a matrix is fairly simple and follows the recipe:
    1. The left-singular vectors of $\mathbf{A}$ are orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^T$
    2. The right-singular vectors of $\mathbf{A}$ are orthonormal eigenvectors of $\mathbf{A}^T\mathbf{A}$
    3. The non-zero singular values of $\mathbf{A}$ are the (positive) square roots of the non-zero eigenvalues of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$

Example

1. Compare the SVD of the matrix $\mathbf{A}$ of the previous example with the spectral decomposition of $\mathbf{B}$. In particular, compare the singular values of $\mathbf{A}$ with the square roots of the eigen values of $\mathbf{B}$.
2. Now compare the SVD and the EVD of $\mathbf{B}$. Do we notice anything interesting?

Singular Value Decomposition

Definition (Reduced SVD)

The **reduced SVD** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, is given by:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

- $\text{rank}(\mathbf{A}) = r \leq \min(m, n)$ and σ_i is the i^{th} **singular value** of $\mathbf{A}$
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} **left** and **right singular vectors** of $\mathbf{A}$ respectively
- $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are **unitary** matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{I}$)
- $\boldsymbol{\Sigma} = \textbf{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_r\right) \in \mathbb{R}^{r \times r}$ where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$

Singular Value Decomposition

Definition (Reduced SVD)

The **reduced SVD** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, is given by:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

- $\operatorname{rank}(\mathbf{A}) = r \leq \min(m, n)$ and σ_i is the i^{th} **singular value** of $\mathbf{A}$
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} **left** and **right singular vectors** of $\mathbf{A}$ respectively
- $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are **unitary** matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{I}$)
- $\boldsymbol{\Sigma} = \mathbf{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_r\right) \in \mathbb{R}^{r \times r}$ where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$

- $\mathbf{v}_i$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$; $\sigma_i = \sqrt{\lambda_i\left(\mathbf{A}^T\mathbf{A}\right)}$ (and $\lambda_i\left(\mathbf{A}^T\mathbf{A}\right) = 0$ for $i > r$) since $\mathbf{A}^T\mathbf{A} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right)^T\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\right) = \left(\mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^T\right)$

Singular Value Decomposition

Definition (Reduced SVD)

The **reduced SVD** of a matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, is given by:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

- $\operatorname{rank}(\mathbf{A}) = r \leq \min(m, n)$ and σ_i is the i^{th} **singular value** of $\mathbf{A}$
- $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i^{th} **left** and **right singular vectors** of $\mathbf{A}$ respectively
- $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ are **unitary** matrices (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{I}$)
- $\mathbf{\Sigma} = \mathbf{diag}\left(\sigma_1, \sigma_2, \ldots, \sigma_r\right) \in \mathbb{R}^{r \times r}$ where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$

- $\mathbf{v}_i$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$; $\sigma_i = \sqrt{\lambda_i\left(\mathbf{A}^T\mathbf{A}\right)}$ (and $\lambda_i\left(\mathbf{A}^T\mathbf{A}\right) = 0$ for $i > r$) since $\quad \mathbf{A}^T\mathbf{A} = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right) = \left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\right)$
- $\mathbf{u}_i$ are eigenvectors of $\mathbf{A}\mathbf{A}^T$; $\sigma_i = \sqrt{\lambda_i\left(\mathbf{A}\mathbf{A}^T\right)}$ (and $\lambda_i\left(\mathbf{A}\mathbf{A}^T\right) = 0$ for $i > r$) since $\quad \mathbf{A}\mathbf{A}^T = \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T = \left(\mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T\right)$

Operator & matrix norms

Operator & matrix norms

- Consider vector spaces U and V endowed with a norm
- The set of linear maps from U to V form another **vector space**
- The **norms** defined on U & V induce a norm on this space of linear maps

Definition (Operator norm)

If $T : U \to V$ is a linear map, then the operator norm is defined as

$$\|T\|_{\text{op}} = \|T\|_{U\to V} = \sup_{\mathbf{u}\in U\setminus\{\mathbf{0}\}} \frac{\|T(\mathbf{u})\|_V}{\|\mathbf{u}\|_U}$$

- An important class is when $U = \mathbb{R}^n$ and $V = \mathbb{R}^m$ and $p-$norms used

Definition (Matrix norm)

The matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ mapping $\mathbb{R}^n$ and $\mathbb{R}^m$ has an induced norm

$$\|\mathbf{A}\|_{p\to q} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_q}{\|\mathbf{x}\|_p}$$

Matrix norms

- When $p = q$, the matrix p-norm of $\mathbf{A}$ is given by

$$\|\mathbf{A}\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

- The special cases of $p = 1, 2, \infty$ give

$$\|\mathbf{A}\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |a_{ij}| \qquad \text{(maximum absolute column sum)}$$

$$\|\mathbf{A}\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |a_{ij}| \qquad \text{(maximum absolute row sum)}$$

$$\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A}) \qquad \text{(maximum singular value – \textbf{spectral norm})}$$

Matrix norms

- Induced matrix p-norms have the following important properties

$$\|\mathbf{Ax}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{x}\|_p$$
$$\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p \qquad \text{(submultiplicative property)}$$

- There are other matrix norms

Example (with $r = \min\{m, n\}$ and $\sigma_i = \sigma(\mathbf{A})_i$)

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^{r} (\sigma_i)^2} \equiv \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} \quad \text{(Frobenius norm)}$$

$$\|\mathbf{A}\|_* := \sum_{i=1}^{r} \sigma_i \equiv \operatorname{trace}\left(\sqrt{\mathbf{A}^T\mathbf{A}}\right) \quad \text{(Nuclear norm)}$$

Matrix norms

Definition (Unitary invariance)

A matrix norm $\|\cdot\|$ is unitary invariant if for any orthogonal (unitary) matrices $\mathbf{U}$ & $\mathbf{V}$ we have

$$\|\mathbf{UAV}\| = \|\mathbf{A}\|$$

- Unitary invariant norms depend only on the singular values of a matrix, i.e.

$$\|\mathbf{A}\| = \|\mathbf{U}\Sigma\mathbf{V}^T\|$$

- Three of the matrix norms discussed has the unitary invariance property

Proposition

The spectral norm, the Frobenius norm and Nuclear norm are unitary invariant.

Problem (3)

Prove the above Proposition.

Low rank approximation

Low rank approximation

- An important practical application of the **Singular Value Decomposition**
- Given a matrix, finding another matrix of the same dimensions but **lower rank** such that the two matrices are **close** in some norm
- This is used to reduce the amount of data needed to **store** a matrix, while retaining most of its information

Problem (Rank-r approximation)

Given $\mathbf{Y}$, find $\arg\min_{\mathbf{X}} \|\mathbf{X}-\mathbf{Y}\|_F$ subject to: $\operatorname{rank}(\mathbf{X}) \leq r$.

- By Eckart-Young-Mirsky theorem the **optimal matrix** can be computed from the SVD, for unitary invariant norms