

# CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin  
Spring 2024

# Lecture 1: Introduction and Overview

# Outline

- 1 Class Topics and Logistics
- 2 Introduction - Vector spaces and matrices
- 3 Eigenvalues and singular values
- 4 Vector and matrix norms

# Data Deluge

- Modern applications involve *large dimensional datasets* (matrices and beyond!).
- New technologies - generation and collection of *large volumes of scientific data*.
- Algorithms - Inexpensive, scalable; parallel and online/streaming.

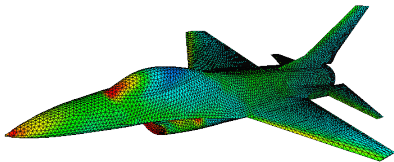


# A Multi-Dimensional World

- Much of real-world **data** is inherently **multidimensional**



- Many **operators** and **models** are natively **multi-way**



# Algorithms for Data

- Growing demands of data science and artificial intelligence and the need to handle *large and high dimensional* data have ushered in a “new era” for algorithms research.
- Today’s *data problems* are two folds:
  - ▶ **Computational** issues in handling large and high dimensional data.
  - ▶ **Representational** challenges in order to capture multi-dimensional correlation structure.
- Typical data applications require combining a diverse set of algorithmic tools. Most are not heavily covered in traditional algorithms curriculum.

# Class topics

- The class topics are divided into two parts:
  - ① **Randomized matrix computations**
  - ② **Tensor algebraic methods**
- *Randomized linear algebra* - Approximate computational paradigm through the interplay between statistics, algebra and geometry.
- *Tensor algebra* - algebraic constructs that represent and manipulate natively high-dimensional entities, while preserving their multi-dimensional integrity.
- We will cover theory, matlab/Python implementations, and applications.
- Focus on the tools to design new algorithms.
- Will need strong background in *linear algebra* and *probability*.

## Course webpage:

<https://shashankaubaru.github.io/Teaching/CSE392-2024.html>

You will find all information related to the course.

**Instructor:** Shashanka Ubaru

- *Email:* shashanka.ubaru@austin.utexas.edu or @ibm.com
- *Office hours:* Wednesdays 3:00pm - 4:00pm.
- *Location:* POB 3.134

## Class time and Location:

Mondays and Wednesdays, 11:00am - 12:30pm, GDC 2.402.



## Class Logistics II

- Syllabus, schedule, lecture notes and other information can all be found in the *class webpage*.
- Assignments are to be submitted through Canvas, and should be individual work. You can discuss the problems, but should submit individually. Preferably typewritten.
- The programming languages for the course will be Matlab and/or Python.
- Some of the assignments and exercises will involve programming and code submission.
- We will use *Canvas* for grades, submissions, etc.

## Grading:

- **Scribing** - 10% : Participation and scribed notes preparation. 1 - 2 lectures, depending on the class strength. LaTeX template is available in the webpage.
- **Assignments** - 50% : Around 5 problem sets each contributing an equal amount to the grade. Will include programming exercises.
- **Class Project** - 40% : Teams of two. There will be a final presentation of the projects during the last week of the semester.

*Relevant resources* will be posted on the webpage.

Questions?

## General Introduction

- Background: Linear algebra and numerical linear algebra.
- Mathematical background - vector spaces, matrices, rank.
- Types of matrices, structured matrices.
- eigenvalues, singular values.
- Inner products, norms.

# Vector spaces and matrices

- A **vector subspace** of  $\mathbb{R}^n$  is a subset of  $\mathbb{R}^n$  that is also a real vector space.
- The set of all linear combinations of a set of vectors  $\mathbb{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_q\}$  of  $\mathbb{R}^n$  is a vector subspace called the linear span of  $\mathbb{A}$ .
- If the  $\mathbf{a}_i$ 's are linearly independent, then each vector of  $\text{span}(\mathbb{A})$  admits a unique expression as a linear combination of the  $\mathbf{a}_i$ 's. The set  $\mathbb{A}$  is then called a *basis*.

# Vector spaces and matrices

- A **vector subspace** of  $\mathbb{R}^n$  is a subset of  $\mathbb{R}^n$  that is also a real vector space.
- The set of all linear combinations of a set of vectors  $\mathbb{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_q\}$  of  $\mathbb{R}^n$  is a vector subspace called the linear span of  $\mathbb{A}$ .
- If the  $\mathbf{a}_i$ 's are linearly independent, then each vector of  $\text{span}(\mathbb{A})$  admits a unique expression as a linear combination of the  $\mathbf{a}_i$ 's. The set  $\mathbb{A}$  is then called a *basis*.
- A matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is an  $m \times n$  array of real numbers

$$a_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

- A matrix represents a linear mapping between two vector spaces of finite dimension  $n$  and  $m$ :

$$\mathbf{x} \in \mathbb{R}^n \longrightarrow \mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$$

- **Notation** :  $\mathcal{A}^{n_1 \times n_2 \dots \times n_d}$  -  $d^{th}$  order tensor

- ▶  $0^{th}$  order tensor - scalar



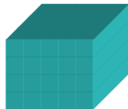
- ▶  $1^{st}$  order tensor - vector



- ▶  $2^{nd}$  order tensor - matrix



- ▶  $3^{rd}$  order tensor ...



# Matrix operations

- **Addition:**  $C = A + B$ , where  $A, B, C \in \mathbb{R}^{m \times n}$  with

$$c_{ij} = a_{ij} + b_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

- **Scalar multiplication:**  $C = \alpha A$ , where

$$c_{ij} = \alpha a_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

- **Matrix-matrix multiplication:**  $C = AB$ , where  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{n \times p}$ ,  $C \in \mathbb{R}^{m \times p}$  with

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \quad i = 1, \dots, m, \quad j = 1, \dots, p.$$



# Matrix operations

- **Transposition:**  $C = A^\top$ , where  $A \in \mathbb{R}^{m \times n}$ ,  $C \in \mathbb{R}^{n \times m}$  with

$$c_{ij} = a_{ji}, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

- **Transpose conjugate:** for complex matrices

$$A^H = \bar{A}^\top = \bar{A}^\top.$$

- **Kronecker product:** For  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{p \times q}$

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \cdots & \cdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

In Matlab and Numpy: `kron(A,B)`. Size = ??

## Questions and Exercises

- $(\mathbf{A}^\top)^\top = ?$   $(\mathbf{AB})^\top = ?$   $(\mathbf{A}^H)^H = ?$   
 $(\mathbf{A}^H)^\top = ?$   $(\mathbf{ABC})^\top = ?$
- When is  $\mathbf{AA}^\top = \mathbf{A}^\top \mathbf{A}$ ?
- What are the computational complexity of (a) matrix addition,  
(b) matrix-vector product (matvec), and  
(c) matrix-matrix product?
- If  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , then what are the sizes of  $\mathbf{u}^\top \mathbf{v}$  and  $\mathbf{uv}^\top$ ?  
What are these called?
- **Exercise 1:** Show that for  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , we have  $\mathbf{v}^\top \otimes \mathbf{u} = \mathbf{uv}^\top$ .

# Range, rank, and null space

- **Range:**  $\text{Ran}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$
- **Null Space:**  $\text{Null}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\} \subseteq \mathbb{R}^n$
- Range = linear span of the columns of  $\mathbf{A}$
- **Rank** of a matrix  $\text{rank}(\mathbf{A}) = \dim(\text{Ran}(\mathbf{A})) \leq n$
- $\text{Ran}(\mathbf{A}) \subseteq \mathbb{R}^m \rightarrow \text{rank}(\mathbf{A}) \leq m \rightarrow$

$$\text{rank}(\mathbf{A}) \leq \min\{m, n\}.$$

- $\text{rank}(\mathbf{A}) =$  number of linearly independent columns of  $\mathbf{A} =$  number of linearly independent rows of  $\mathbf{A}$ .
- $\mathbf{A}$  is of *full rank* if  $\text{rank}(\mathbf{A}) = \min\{m, n\}$ . Otherwise it is *rank-deficient*.

# Rank - Nullity Theorem

- For  $\mathbf{A} \in \mathbb{R}^{m \times n}$ :

$$\dim(\text{Ran}(\mathbf{A})) + \dim(\text{Null}(\mathbf{A})) = n$$

Also

$$\dim(\text{Ran}(\mathbf{A}^\top)) + \dim(\text{Null}(\mathbf{A}^\top)) = m$$

- $\dim(\text{Null}(\mathbf{A}))$  is called the **nullity** or **co-rank** of  $\mathbf{A}$ .
- $\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n$ .

**Question:** If  $\text{rank}(\mathbf{A}) = r$ , what are  $\text{rank}(\mathbf{A}^\top)$ ,  $\text{rank}(\bar{\mathbf{A}})$ ,  $\text{rank}(\mathbf{A}^H)$ .  
Explore **rank** function in Matlab or numpy.

# Types of matrices

- **Orthonormal** :  $U \in \mathbb{R}^{m \times n}$  is orthonormal if  $U^\top U = I$ .
- If  $U$  is square, then it is orthogonal (or **unitary** if complex), and  $UU^\top = I$ .
- A square matrix  $A \in \mathbb{C}^{n \times n}$  is,  
**Symmetric** :  $A^\top = A$ , **Skew-symmetric** :  $A^\top = -A$  ,  
**Hermitian**:  $A^H = A$ , **Skew-Hermitian** :  $A^H = -A$ , **Normal**:  $A^H A = A A^H$ .
- Matrix is *non-negative* if  $a_{ij} \geq 0, i, j = 1, \dots, n$ .
- A symmetric matrix  $P$  of the form  $P = UU^\top$  is a projection matrix, and  $PP = P$ .
- **Structured matrices**: Diagonal, Upper (U) and Lower (L) triangular, U & L bidiagonal, tridiagonal, and U & L Hessenberg.
- **Special matrices**: Toeplitz, Hankel, and circulant matrices.
- **Sparse matrices** Many of the large matrices encountered in applications are sparse. Sparse matrix computations can be a separate course.

## **Recommended reading:**

If these topics are not familiar, refer to sections 1.1 to 1.6 in Dr. Yousef Saad's text book:

[http://www.cs.umn.edu/~saad/eig\\_book\\_2ndEd.pdf](http://www.cs.umn.edu/~saad/eig_book_2ndEd.pdf).

# Eigenvalues and Eigenvectors

A complex scalar  $\lambda$  is called an *eigenvalue* of a square matrix  $\mathbf{A} \in \mathbb{C}^{n \times n}$  if there exists a nonzero vector  $\mathbf{u} \in \mathbb{C}^n$  such that

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}.$$

The vector  $\mathbf{u}$  is called an *eigenvector* of  $\mathbf{A}$  associated with  $\lambda$ .

- The set of all eigenvalues of  $\mathbf{A}$ , denoted  $\Lambda(\mathbf{A})$ , is the *spectrum* of  $\mathbf{A}$ .
- An eigenvalue is a root of the *characteristic polynomial*:

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

- **Diagonalization:** Two matrices  $\mathbf{A}, \mathbf{B}$  are *similar* if there exists a nonsingular matrix  $\mathbf{X}$  such that:  $\mathbf{A} = \mathbf{X}\mathbf{B}\mathbf{X}^{-1}$ .  
 $\mathbf{A}$  is diagonalizable if it is similar to a diagonal matrix

# Eigenvalues and properties

- For every square symmetric matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , we can compute eigendecomposition:

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top,$$

where  $\mathbf{U}$  is an orthogonal matrix with eigenvectors  $\mathbf{u}_i$  as columns, and  $\mathbf{\Lambda}$  is diagonal matrix with eigenvalues  $\lambda_i$  on the diagonal.

- **Spectral radius:** The maximum modulus of the eigenvalues

$$\rho(\mathbf{A}) = \max_{\lambda \in \Lambda(\mathbf{A})} |\lambda|$$

- **Trace** of  $\mathbf{A}$  is the sum of diagonal elements

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i$$

sum of all the eigenvalues of  $\mathbf{A}$  counted with their multiplicities.

- Note  $\det(\mathbf{A}) =$  product of all the eigenvalues of  $\mathbf{A}$  counted with their multiplicities.



# Singular values

- Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$  or  $\mathbb{C}^{m \times n}$ .
- The eigenvalues of  $\mathbf{A}^H \mathbf{A}$  and  $\mathbf{A} \mathbf{A}^H$  are real and  $\geq 0$ .
- Let  $\sigma_i = \sqrt{\lambda_i(\mathbf{A}^H \mathbf{A})}$  if  $n \leq m$   
else  $\sigma_i = \sqrt{\lambda_i(\mathbf{A} \mathbf{A}^H)}$  for  $i = 1, \dots, \min\{n, m\}$ .
- These  $\sigma_i$ 's are called the **singular values** of  $\mathbf{A}$ .

**Singular value decomposition:** For every matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , we have

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

where  $\mathbf{U} \in \mathbb{R}^{m \times m}$ ,  $\mathbf{V} \in \mathbb{R}^{n \times n}$  are orthogonal matrices, and  $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$  is diagonal matrix with singular values  $\sigma_i$  on the diagonal ordered nonincreasingly:  
 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$ .

## Questions and Exercises

- Given a symmetric matrix  $\mathbf{A}$  with eigen-decomposition  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ , then
  - ① What are the eigenvalues/eigenvectors of  $\mathbf{A}^q$  for a given integer power  $q$ ?
  - ② If  $\mathbf{A}$  is nonsingular what are the eigenvalues/eigenvectors of  $\mathbf{A}^{-1}$ ?
  - ③ What are the eigenvalues/eigenvectors of  $p(\mathbf{A})$  for a polynomial  $p(\cdot)$ ?
- Similarly, for a general matrix  $\mathbf{A} \in \mathbb{R}^{n \times d}$ , with SVD  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ , what are the eigen-values of  $\mathbf{A}^\top \mathbf{A}$ ?

# Inner products and norms

- **Inner product** of two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ :

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\top \mathbf{v} = \sum_{i=1}^n u_i v_i$$

- For complex numbers?
- Given  $\mathbf{A} \in \mathbb{C}^{m \times n}$  then,

$$\langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{A}^H \mathbf{v} \rangle.$$

- **Vector norm** on a vector space  $\mathbb{X}$  is a real-valued function on  $\mathbb{X}$ , which satisfies the following three conditions:
  1.  $\|\mathbf{x}\| \geq 0, \forall \mathbf{x} \in \mathbb{X}$ , and  $\|\mathbf{x}\| = 0$  iff  $\mathbf{x} = \mathbf{0}$ .
  2.  $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|, \forall \mathbf{x} \in \mathbb{X}, \forall \alpha \in \mathbb{C}$ .
  3.  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{X}$ .

# Vector norms

- **Euclidean norm** on  $\mathbb{X} = \mathbb{C}^n$ ,

$$\|\mathbf{x}\|_2 = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = \sqrt{\sum_{i=1}^n |x_i|^2}$$

- Most common vector norms in numerical linear algebra: for  $p \geq 1$  (Hölder norms)

$$\|\mathbf{x}\|_p = \left( \sum_i |x_i|^p \right)^{1/p}$$

- **Cauchy-Schwartz inequality:**

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

- **Hölder inequality:**

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \text{ with } \frac{1}{p} + \frac{1}{q} = 1$$

# Matrix norms

- **Matrix norm** by treating  $m \times n$  matrices as vectors in  $\mathbb{C}^{mn}$ :
  1.  $\|\mathbf{A}\| \geq 0, \forall \mathbf{A} \in \mathbb{C}^{m \times n}$ , and  $\|\mathbf{A}\| = 0$  iff  $\mathbf{A} = \mathbf{0}$ .
  2.  $\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\|, \forall \mathbf{A} \in \mathbb{C}^{m \times n}, \forall \alpha \in \mathbb{C}$ .
  3.  $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|, \forall \mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$ .
- Given  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , we define a set of *matrix norms* :

$$\|\mathbf{A}\|_p = \max_{\mathbf{x} \in \mathbb{C}^n, \mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

- **Consistency / sub-multiplicativity of matrix norms:**

$$\|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{B}\|_p$$

- **Frobenius norm** of a matrix:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}.$$

# Expressions of standard matrix norms

- Recall for a square matrix, we have

$$\rho(\mathbf{A}) = \max_{\lambda \in \Lambda(\mathbf{A})} |\lambda| \text{ and } \text{Tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i.$$

- Then the matrix norms are:

$$\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^m |a_{ij}|,$$

$$\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|,$$

$$\|\mathbf{A}\|_2 = [\rho(\mathbf{A}^H \mathbf{A})]^{1/2} = [\rho(\mathbf{A}^H \mathbf{A})]^{1/2}.$$

$$\|\mathbf{A}\|_F = [\text{Tr}(\mathbf{A}^H \mathbf{A})]^{1/2} = [\text{Tr}(\mathbf{A}^H \mathbf{A})]^{1/2}.$$

## In terms of singular values

- For  $\mathbf{A}$ , assume we have  $r$  nonzero singular values (with  $r \leq \min\{m, n\}$ ) :

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

- Then, we have

$$\|\mathbf{A}\|_2 = \sigma_1 \quad \text{and} \quad \|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$$

- Schatten  $p$ -norms for  $p \geq 1$

$$\|\mathbf{A}\|_{*,p} = \left[ \sum_{i=1}^r \sigma_i^p \right]^{1/p}$$

- In particular:  $\|\mathbf{A}\|_{*,1} = \sum_{i=1}^r \sigma_i$  is called the **nuclear norm** and is denoted by  $\|\mathbf{A}\|_*$ .

## Questions and Exercises

- For an orthogonal matrix  $\mathbf{U}$ , show that  $\|\mathbf{U}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ .
- **Exercise 2:** Show that for any  $\mathbf{x}$ :  $\frac{1}{\sqrt{n}}\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ .
- **Exercise 3:** Prove that the Frobenius norm is consistent [Hint: Use Cauchy-Schwartz]
- Let  $\mathbf{A} = \mathbf{u}\mathbf{v}^\top$ . Then,  $\|\mathbf{A}\|_2 = \|\mathbf{u}\|_2\|\mathbf{v}\|_2$ .
- **Exercise 4:** Prove the above.  
What is  $\|\mathbf{A}\|_F = ?$