CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2024

Lecture 1: Introduction and Overview

Outline

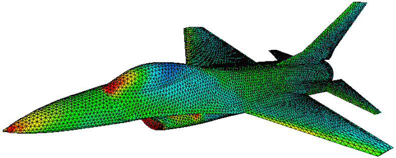
Data Deluge

- 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 Logistics

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.

Class Logistics III

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?

This lecture

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

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

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

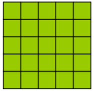
$$\boldsymbol{x} \in \mathbb{R}^n \longrightarrow \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$$

Tensors

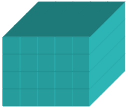
- **Notation** : $\mathcal{A}^{n_1 \times n_2 \ldots, \times n_d}$ - d^{th} order tensor
  - 0^{th} order tensor - **scalar**

  - 1^{st} **order tensor - vector**

  - 2^{nd} **order tensor - matrix**



  - 3^{rd} **order tensor ...**



Matrix operations

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

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

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

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

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

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

Matrix operations

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

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

- **Transpose conjugate:** for complex matrices

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

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

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

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

Questions and Exercises

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

Range, rank, and null space

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

$$\mathrm{rank}(\boldsymbol{A}) \leq \min\{m, n\}.$$

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

Rank - Nullity Theorem

- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$:
$$\dim(\text{Ran}(\boldsymbol{A})) + \dim(\text{Null}(\boldsymbol{A})) = n$$
Also
$$\dim(\text{Ran}(\boldsymbol{A}^\top)) + \dim(\text{Null}(\boldsymbol{A}^\top)) = m$$
- $\dim(\text{Null}(\boldsymbol{A}))$ is called the **nullity** or **co-rank** of $\boldsymbol{A}$.
- $\text{rank}(\boldsymbol{A}) + \text{nullity}(\boldsymbol{A}) = n$.

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

Types of matrices

- **Orthonormal :** $\boldsymbol{U} \in \mathbb{R}^{m\times n}$ is orthonormal if $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$.
- If $\boldsymbol{U}$ is square, then it is orthogonal (or **unitary** if complex), and $\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{I}$.
- A square matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is,
  Symmetric : $\boldsymbol{A}^\top = \boldsymbol{A}$, **Skew-symmetric :** $\boldsymbol{A}^\top = -\boldsymbol{A}$,
  Hermitian: $\boldsymbol{A}^H = \boldsymbol{A}$, **Skew-Hermitian :** $\boldsymbol{A}^H = -\boldsymbol{A}$, **Normal:** $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^H$.
- Matrix is *non-negative* if $a_{ij} \geq 0, i, j = 1, \ldots, n$.
- A symmetric matrix $\boldsymbol{P}$ of the form $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{U}^\top$ is a projection matrix, and $\boldsymbol{P}\boldsymbol{P} = \boldsymbol{P}$.
- **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.

Reference

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`.

Eigenvalues and Eigenvectors

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

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

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

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

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

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

Eigenvalues and properties

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

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

  where $\boldsymbol{U}$ is an orthogonal matrix with eigenvectors $\boldsymbol{u}_i$ as columns, and Λ is diagonal matrix with eigenvalues λ_i on the diagonal.
- **Spectral radius**: The maximum modulus of the eigenvalues

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

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

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

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

Singular values

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

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

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

where $\boldsymbol{U} \in \mathbb{R}^{m \times m}, \boldsymbol{V} \in \mathbb{R}^{m \times n}$ are an orthogonal matrices , and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal matrix with singular values σ_i on the diagonal ordered nonincreasingly:
$\sigma_1 \geq \sigma 21 \geq \cdots \geq \sigma_m \geq 0$.

Questions and Exercises

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

Inner products and norms

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

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

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

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

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

Vector norms

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

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

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

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^2 \right)^{1/p}$$

- **Cauchy-Schwartz inequality:**

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

- **Hölder inequality:**

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

Matrix norms

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

$$\|\boldsymbol{A}\|_p = \max_{\boldsymbol{x}\mathbb{C}^m, \boldsymbol{x}\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}$$

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

$$\|\boldsymbol{A}\boldsymbol{B}\|_p \leq \|\boldsymbol{A}\|_p \|\boldsymbol{B}\|_p$$

- **Frobenius norm** of a matrix:

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

Expressions of standard matrix norms

- Recall for a square matrix, we have $\rho(\boldsymbol{A}) = \max_{\lambda \in \Lambda(\boldsymbol{A})} |\lambda|$ and $\mathrm{Tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$.
- Then the matrix norms are:

$$\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|,$$
$$\|\boldsymbol{A}\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|,.$$
$$\|\boldsymbol{A}\|_2 = [\rho(\boldsymbol{A}^H \boldsymbol{A})]^{1/2} = [\rho(\boldsymbol{A}^H \boldsymbol{A})]^{1/2}.$$
$$\|\boldsymbol{A}\|_F = [\mathrm{Tr}(\boldsymbol{A}^H \boldsymbol{A})]^{1/2} = [\mathrm{Tr}(\boldsymbol{A}^H \boldsymbol{A})]^{1/2}.$$

In terms of singular values

- For $\boldsymbol{A}$, 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

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

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

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

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

Questions and Exercises

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