AMS526: Numerical Analysis I
(Numerical Linear Algebra for Computational and Data Sciences)

Lecture 1: Course Overview; Matrix Multiplication

Xiangmin Jiao

Stony Brook University
Outline

1 Course Overview

2 Matrix Notation and Basic Operations (MC §1.1)

3 Range, Null Space and Rank (NLA §1)
Course Description

- What is numerical linear algebra?
  - Solving linear algebra problems using efficient algorithms on computers

- Topics: Direct and iterative methods for solving simultaneous linear equations, least squares problems, computation of eigenvalues and eigenvectors, and singular value decomposition

- Required textbooks

- Additional supplementary material will be provided as needed

- Course webpage:
  http://www.ams.sunysb.edu/~jiao/teaching/ams526
Prerequisite

- **This MUST NOT** be your first course in linear algebra, or you will get lost

- **Prerequisite/Co-requisite:**
  - AMS 510 (linear algebra portion) or equivalent undergraduate-level linear algebra course. Familiarity with following concepts is assumed: Vector spaces, Gaussian elimination, Gram-Schmidt orthogonalization, and eigenvalues/eigenvectors
  - AMS 595 (co-requisite for students without programming experience)

- To review fundamental concepts of linear algebra, see textbook such as
Why Learn Numerical Linear Algebra?

- Foundation of scientific computations and data sciences
- Many problems ultimately reduce to linear algebra concepts or algorithms, either analytical or computational
- Examples: Finite-element analysis, data fitting, PageRank (Google)

Focus: Fundamental concepts, efficiency and stability of algorithms, and programming

New focus: relevance to computational and data sciences
Course Outline

- **Fundamentals** (matrix notation and basic operations; vector spaces; algorithmic considerations; norms and condition numbers; decomposition of matrices)
- **Linear systems** (triangular systems; Gaussian elimination; accuracy and stability; Cholesky factorization; sparse linear systems)
- **QR factorization and least squares** (Gram-Schmidt orthogonalization; QR factorization with Householder reflection; updating QR factorization with Givens rotation; stability; least squares problems; rank-revealing QR; SVD and low-rank approximations)
- **Eigenvalue problems** (eigenvalues and invariant spaces; classical eigenvalue methods; QR algorithms; two-stage methods; Arnoldi and Lanczos iterations)
- **Iterative Methods for linear systems** (basic iterative methods; conjugate gradient methods; minimal residual style methods; bi-Lanczos iterations; preconditioners)
- **Special topics** (multigrid methods; compressed sensing; under-determined systems, etc., if time permits)
Course Policy

- Assignments (written or programming)
  - Assignments are due in class one to two weeks after assigned
  - You can discuss course materials and homework problems with others, but you must write your answers completely independently
  - Do NOT copy solutions from any source. Do NOT share your solutions with others

- Exams and tests
  - All exams are closed-book
  - However, one-page cheat sheet is allowed

- Grading
  - Assignments: 30%
  - Two midterm exams: 40%
  - Final exam: 30%
Outline

1. Course Overview

2. Matrix Notation and Basic Operations (MC §1.1)

3. Range, Null Space and Rank (NLA §1)
Matrices and Vectors

- Denote vector space of all $m$-by-$n$ real matrices by $\mathbb{R}^{m \times n}$.

$$A \in \mathbb{R}^{m \times n} \iff A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- Denote vector space of all real $n$-vectors by $\mathbb{R}^n$, or $\mathbb{R}^{n \times 1}$

$$x \in \mathbb{R}^n \iff x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Transposition ($\mathbb{R}^{m \times n} \to \mathbb{R}^{n \times m}$): $C = A^T \Rightarrow c_{ij} = a_{ji}$

- Row vectors are transpose of column vectors and are in $\mathbb{R}^{1 \times n}$
Matrix Operations

- **Addition and subtraction** ($\mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{m\times n}$):
  \[ C = A \pm B \Rightarrow c_{ij} = a_{ij} \pm b_{ij} \]

- **Scalar-matrix multiplication or scaling** ($\mathbb{R} \times \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{m\times n}$):
  \[ C = \alpha A \Rightarrow c_{ij} = \alpha a_{ij} \]

- **Matrix-matrix multiplication/product** ($\mathbb{R}^{m\times p} \times \mathbb{R}^{p\times n} \rightarrow \mathbb{R}^{m\times n}$):
  \[ C = A \cdot B \Rightarrow c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} \text{ (denoted by } A \ast B \text{ in MATLAB)} \]

- Each operation also applies to vectors. In particular,
  - **Inner product** is row vector times column vector, i.e., $c = x^T y$
    (it is called *dot product* in vector calculus and denoted as $x \cdot y$)
  - **Outer product** is column vector times row vector, i.e., $C = xy^T$
    (it is a special case of Kronecker product and denoted as $x \otimes y$)

- **Element-wise multiplication and division** ($\mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \rightarrow \mathbb{R}^{m\times n}$)
  - $C = A \ast B \Rightarrow c_{ij} = a_{ij} b_{ij}$
  - $C = A/\cdot B \Rightarrow c_{ij} = a_{ij} / b_{ij}$, where $b_{ij} \neq 0$

- **Matrix inversion** ($A^{-1}$) and **division** ($A/B$ and $A\backslash B$) to be defined later
Notation of Matrices and Vectors

- **Matrix notation**
  - Capital letters (e.g., A, B, Δ, etc.) for matrices
  - Corresponding lower case with subscript \(ij\) (e.g., \(a_{ij}\), \(b_{ij}\), \(\delta_{ij}\)) for \((i,j)\) entry; sometimes with notation \([A]_{ij}\) or \(A(i,j)\)

- **Vector notation**
  - Lowercase letters (e.g., \(x\), \(y\), etc.) for vectors
  - Corresponding lower case with subscript \(i\) for \(i\)th entry (e.g., \(x_i\), \(y_i\))

- Lower-case letters for scalars (e.g., \(c\), \(s\), \(\alpha\), \(\beta\), etc.)

- Some books suggest using boldface lowercase (e.g., \(x\)) for vectors, regular lowercase (e.g., \(c\)) for scalars, and boldface uppercase for matrices

- A matrix is a collection of column vectors or row vectors

\[
A \in \mathbb{R}^{m \times n} \iff A = [c_1 | c_2 | \ldots | c_n], \quad c_k \in \mathbb{R}^m
\]

\[
A \in \mathbb{R}^{m \times n} \iff A = \begin{bmatrix} r_1^T \\ \vdots \\ r_m^T \end{bmatrix}, \quad r_k \in \mathbb{R}^n
\]
Complex Matrices

- Occasionally, complex matrices are involved
- Vector space of $m$-by-$n$ complex matrices is designated by $\mathbb{C}^{m \times n}$
  - Scaling, addition, multiplication of complex matrices correspond exactly to real case
  - If $A = B + iC \in \mathbb{C}^{m \times n}$, then $\text{Re}(A) = B$, $\text{Im}(A) = C$, and conjugate of $A$ is $\overline{A} = (\overline{a}_{ij})$
  - Conjugate transpose is defined as $A^H = B^T - iC^T$, or
    $$G = A^H \Rightarrow g_{ij} = \overline{a}_{ji}$$
    (also called adjoint, and denoted by $A^*; (AB)^* = B^*A^*$)
- Vector space of complex $n$-vectors is designated by $\mathbb{C}^n$
  - Inner product of complex $n$-vectors $x$ and $y$ is $s = x^H y$
  - Outer product of complex $n$-vectors $x$ and $y$ is $S = xy^H$
- We will primarily focus on real matrices
Matrix-Vector Product

- Matrix-vector product $b = Ax$ is a special case of matrix-matrix product

$$b_i = \sum_{j=1}^{n} a_{ij}x_j$$

- For $A \in \mathbb{R}^{m \times n}$, $Ax$ is a mapping $x \mapsto Ax$ from $\mathbb{R}^n$ to $\mathbb{R}^m$

- This map is **linear**, which means that for any $x, y \in \mathbb{R}^n$ and any $\alpha \in \mathbb{R}$

$$A(x + y) = Ax + Ay$$
$$A(\alpha x) = \alpha Ax$$
Linear Combination

- Let $a_j$ denote $j$th column of matrix $A$
  - Alternative notation is colon notation: $A(:,j)$ or $a::j$
  - Use $A(i,:)$ or $a_{i,:}$ to denote $i$th row of $A$

- $b$ is a linear combination of column vectors of $A$, i.e.,

$$b = Ax = \sum_{j=1}^{n} x_j a_j = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1m} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

- In summary, two different views of matrix-vector products:
  1. Scalar operations: $b_i = \sum_{j=1}^{n} a_{ij} x_j$: $A$ acts on $x$ to produce $b$
  2. Vector operations: $b = \sum_{j=1}^{n} x_j a_j$: $x$ acts on $A$ to produce $b$
Matrix-Matrix Multiplication

- Computes $C = AB$, where $A \in \mathbb{R}^{m \times r}$, $B \in \mathbb{R}^{r \times n}$, and $C \in \mathbb{R}^{m \times n}$
- Element-wise, each entry of $C$ is
  \[ c_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj} \]
- Column-wise, each column of $C$ is
  \[ c_j = Ab_j = \sum_{k=1}^{r} b_{kj} a_k; \]
  in other words, $j$th column of $C$ is $A$ times $j$th column of $B$
Outline

1. Course Overview
2. Matrix Notation and Basic Operations (MC §1.1)
3. Range, Null Space and Rank (NLA §1)
Terminology of Vector Space

- **Vector space** is closed under addition and scalar multiplication, with zero vector as a member.
- Vector space *spanned* by a set of vectors \( \{a_j\} \) is

\[
\text{span}\{a_1, \ldots, a_n\} = \left\{ \sum_{j=1}^{n} \beta_j a_j \mid \beta_j \in \mathbb{R} \right\}
\]

- Space spanned by \( n \)-vectors is a *subspace* of \( \mathbb{R}^n \).
- If \( \{a_1, \ldots, a_n\} \) is *linearly independent*, then the \( a_j \) are the *basis* of \( S = \text{span}\{a_1, \ldots, a_n\} \), *dimension* of \( S \) is \( \dim(S) = n \), and each \( b \in S \) is unique linear combination of the \( a_j \).
- If \( S_1 \) and \( S_2 \) are two subspaces, then \( S_1 \cap S_2 \) is a subspace, so is \( S_1 + S_2 = \{b_1 + b_2 \mid b_1 \in S_1, b_2 \in S_2\} \).
  - Note: \( S_1 + S_2 \) is different from \( S_1 \cup S_2 \); latter may not be a subspace.
- Two subspaces \( S_1 \) and \( S_2 \) of \( \mathbb{R}^n \) are *complementary subspaces* of each other if \( S_1 + S_2 = \mathbb{R}^n \) and \( S_1 \cap S_2 = \{0\} \).
  - In other words, \( \dim(S_1) + \dim(S_2) = n \) and \( S_1 \cap S_2 = \{0\} \).
Definition

The *range* of a matrix $A$, written as range$(A)$ or ran$(A)$, is the set of vectors that can be expressed as $Ax$ for some $x$

\[
\{ y \in \mathbb{R}^m | y = Ax \text{ for some } x \in \mathbb{R}^n \}.
\]
## Range and Null Space

### Definition

The *range* of a matrix $A$, written as $\text{range}(A)$ or $\text{ran}(A)$, is the set of vectors that can be expressed as $Ax$ for some $x$

$$\{ y \in \mathbb{R}^m | y = Ax \text{ for some } x \in \mathbb{R}^n \}.$$  

### Theorem

$\text{range}(A)$ is the space spanned by the columns of $A$.

Therefore, *range* of $A$ is also called the *column space* of $A$. 

Range and Null Space

**Definition**

The *range* of a matrix $A$, written as $\text{range}(A)$ or $\text{ran}(A)$, is the set of vectors that can be expressed as $Ax$ for some $x$

$$\{ y \in \mathbb{R}^m | y = Ax \text{ for some } x \in \mathbb{R}^n \}.$$  

**Theorem**

$\text{range}(A)$ is the space spanned by the columns of $A$. 
Therefore, range of $A$ is also called the *column space* of $A$.

**Definition**

The *null space* of $A \in \mathbb{R}^{m\times n}$, written as $\text{null}(A)$, is the set of vectors $x$ that satisfy $Ax = 0$. 

Entries of $x \in \text{null}(A)$ give coefficient of $\sum x_i a_i = 0$. 
Note: The null space of $A$ is in general *not* a complementary subspace of $\text{range}(A)$. 
For real matrices $A \in \mathbb{R}^{m \times n}$
- $\text{null}(A)$ and range($A^T$) are complementary subspaces
- For symmetric matrices ($A = A^T$), $\text{null}(A)$ and range($A$) are complementary subspaces

For complex matrices $A \in \mathbb{C}^{m \times n}$
- $\text{null}(A)$ and range($A^H$) are complementary subspaces
- For Hermitian matrices ($A = A^H$), $\text{null}(A)$ and range($A$) are complementary subspaces
Rank

Definition

The *column rank* of a matrix is the dimension of its column space. The *row rank* is the dimension of the space spanned by its rows.

Question: Can the column rank and the row rank be different?

Answer: No. We will give a proof in later lectures. We therefore simply say the *rank* of a matrix.
Definition

The column rank of a matrix is the dimension of its column space. The row rank is the dimension of the space spanned by its rows.

Question: Can the column rank and the row rank be different?
Answer: No. We will give a proof in later lectures.

- We therefore simply say the rank of a matrix.

Question: Given $A \in \mathbb{R}^{m \times n}$, what is dim(null($A$)) + rank($A$) equal to?
Definition

The *column rank* of a matrix is the dimension of its column space. The *row rank* is the dimension of the space spanned by its rows.

**Question:** Can the column rank and the row rank be different?

**Answer:** No. We will give a proof in later lectures.

- We therefore simply say the *rank* of a matrix.

**Question:** Given $A \in \mathbb{R}^{m \times n}$, what is $\dim(\text{null}(A)) + \text{rank}(A)$ equal to?

**Answer:** $n$.

- A real matrix $A$ is rank-1 if it can be written as $A = uv^T$, where $u$ and $v$ are nonzero vectors.
Full Rank

Definition

A matrix has *full rank* if it has the maximal possible rank, i.e., \( \min\{m, n\} \).

Otherwise, it is called *rank deficient*.
Full Rank

Definition
A matrix has full rank if it has the maximal possible rank, i.e., \( \min\{m, n\} \).
Otherwise, it is called rank deficient.

Theorem
A matrix \( A \in \mathbb{R}^{m \times n} \) with \( m \geq n \) has full rank if and only if it maps no two distinct vectors to the same vector.

In other word, the linear mapping defined by \( Ax \) for \( x \in \mathbb{R}^n \) is one-to-one.
Full Rank

**Definition**

A matrix has *full rank* if it has the maximal possible rank, i.e., \( \min\{m, n\} \).

Otherwise, it is called *rank deficient*.

**Theorem**

A matrix \( A \in \mathbb{R}^{m \times n} \) with \( m \geq n \) has full rank if and only if it maps no two distinct vectors to the same vector.

In other word, the linear mapping defined by \( Ax \) for \( x \in R^n \) is one-to-one.

**Proof.**

(\( \Rightarrow \)) Column vectors of \( A \) forms a basis of \( \text{range}(A) \), so every \( b \in \text{range}(A) \) has a unique linear expansion in terms of the columns of \( A \).

(\( \Leftarrow \)) If \( A \) does not have full rank, then its column vectors are linear dependent, so its vectors do not have a unique linear combination.
Full Rank vs. Non-singularity

- If $A \in \mathbb{R}^{n \times n}$ and $AX = I$, then $X$ is the inverse of $A$, denoted by $A^{-1}$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^T = (A^T)^{-1} = A^{-T}$
  - $(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1} U)^{-1}V^T A^{-1}$
    for $U, V \in \mathbb{R}^{n \times k}$ (Sherman-Morrison-Woodbury formula)

- If $A^{-1}$ exists, $A$ is nonsingular. $A$ is square and has full rank.

- In $A$ is nonsingular, linear system $Ax = b$ results in $x = A^{-1}b$, and it is the inverse problem of matrix-vector multiplication

- If $A \in \mathbb{R}^{m \times n}$ where $n \neq m$ and $A$ has full rank, what is the inverse problem of matrix-vector multiplication?

- If $A$ is rank deficient, what is the inverse problem of matrix-vector multiplication?

- We will need progressively more advanced linear algebra concepts to answer these questions in later part of this semester.