Peilin Wu

MA Courses Review Notes
MA2503
Linear Algebra

August 27, 2017

CityU

Contents

Chapter 1
Linear Equations

1.1 Linear System

1.1.1 Introduction

The general problem is to calculate, if possible, a common solution for a system of m linear algebraic equations in n unknowns:

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{gathered}$$

For which:

- The x_i's are the unknowns
- The a_{ij}'s are the coefficients of the system
- The b_i's are the right-hand side of the system

At this stage, we can introduce the concept of matrix. For the coefficient, we can define a coefficient matrix as following:

Definition:
An $m \times n$ matrix A is a rectangular array of real or complex numbers (called scalars) with m rows and n columns:

$$A_{m,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

Whenever $m = n$, A is called a square matrix, otherwise A is said to be rectangular. Matrices consisting of a single row or a single column are called row and column vectors, respectively.

1.1.2 Rewriting the linear system

By the introducing of the matrix, we can rewrite it into the following matrix form:

$$A_{m,n}X_{n,1} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = B_{m,1}$$

In this way, for comprehensive understanding, matrix represents a particular transform to the vector space where the vector lying in. Each linear system con be considered as a process of space transform and its aim is to identify which vector in space will be transformed into the right hand side constant vector during the very process.

1.2 Gaussian Elimination and Matrices

1.2.1 Three possibilities

As the transform occurred, there are three possibilities for the system:

- unique solution
- no solution
- infinitely many solutions
  note that: if a system has more than one solution, then it necessarily has infinitely many solutions.

1.2.2 Gaussian Elimination

The key idea of Gaussian elimination is to systematically transform one system into another simpler, but equivalent, a system by successively eliminating unknowns, eventually arriving at a system that is easily solvable.

First we write the augmented matrix $[A|b]$like ths following:

$$\left[\begin{array}{ccc|c} a_{11} & a_{12} & \cdots & b_1 \\ a_{21} & a_{12} & \cdots & b_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{12} & \cdots & b_m \end{array}\right]$$

The elimination process relies on three elementary operations:

- interchanging the i-th and j-th equations
- replacing the i-th equation by a nonzero multiple of itself
- replacing the j-th equation by a combination of itself plus a multiple of the i-th equation

After the elimination, the original matrix should transform into a upper triangular matrix like the following:

$$\left[\begin{array}{ccc|c} a'_{11} & a'_{12} & \cdots & b'_1 \\ * & a'_{12} & \cdots & b'_2 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & b'_n \end{array}\right]$$

One can solve the triangularized system using back substitution.
Note: see the connection in the elementary matrix in the LU factorization.

1.3 Gauss-Jordan Method

The Gauss-Jordan method is a variation of Gaussian elimination and is distinct in the following two aspects:

- at each step, the **pivot** element is forced to be **1**
- at each step, all terms **above and below the pivot are eliminated**

Intuitively, Gauss-Jordan method is the strengthened version of Gaussian elimination adding up the back substitution.

After the Gauss-Jordan method, one would in the following form:

$$\left[\begin{array}{ccc|c} 1 & * & \cdots & b_1'' \\ * & 1 & \cdots & b_2'' \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & b_n'' \end{array}\right]$$

1.3.1 Operation counts

In spite of its apparent similarity to Gaussian elimination, the Gauss-Jordan method is more expensive to apply, and it is 50% more work compared with Gaussian elimination. Nevertheless, the method has some theoretical advantages and is useful in other contexts, such as **matrix inversion**.

	multiplications/divisions	additions/subtractions
Gaussian elimination	$\frac{n^3}{3} + n^2 - \frac{n}{3} \sim \frac{n^3}{3}$	$\frac{n^3}{3} - \frac{n^2}{2} - \frac{5n}{6} \sim \frac{n^3}{3}$
Gauss-Jordan method	$\frac{n^3}{2} + \frac{n^2}{2} \sim \frac{n^3}{2}$	$\frac{n^3}{2} - \frac{n^2}{2} \sim \frac{n^3}{2}$

Table 1.1 Operation counts of two methods

1.4 Making Gaussian Elimination Work

1.4.1 Floating-point Numbers

Definition:

$$\pm d_1 d_2 \cdots d_t \times \beta^\epsilon = f \in \mathcal{F}(\beta, t, L, U)$$
$$0 \leq d_t < \beta,\ d_1 \neq 0,\ \epsilon \in [L, U]$$

- β is the base
- t is the precision
- $\epsilon \in [L, U]$ is the exponent range

Rules for floating-point numbers:

- floating-point addition and multiplication are both commutative.
- however,they are neither associative nor distributive.

1.5 Ill-Conditioned Systems

An 2×2 ill-conditioned system where small perturbations in the system can lead to large changes in the solution, corresponds to two straight lines that are almost parallel. General ill-conditioned systems can be characterized in similar ways, with lines replaced by planes and hyperplanes.

Chapter 2
Rectangular Systems and Echelon Forms

2.1 Row Echelon Form and Rank

Analysis general linear systems of m equations in n unknowns, where m may be different from n. Rectangular systems with $m \neq n$ arise naturally when:

1. The number of constraints in the system exceeds the number of unknowns ($m > n$)
   e.g. in the determination of currents in terms of resistances and electromotive forces in an electrical circuit
2. the number of unknowns in the system exceeds the number of constraints ($n > m$)
   e.g. in the computation of geographical locations in a Global Positioning System (GPS)

2.1.1 Modified Gaussian Elimination

Let U be the augmented matrix associated with the system after $i-1$ elimination steps, to execute the i-th step:

- locate the first column in U that contains a nonzero entry on or below the i-th row, say it is U_{*j} ($i \leq j$)
- the pivot position for the i-th step is the (i, j)-position
- if necessary, interchange the i-th row with a lower row to bring a nonzero number into the (i, j)-position, and then eliminate all entries below this pivot
- if row U_{i*} as well as all rows in U below U_{*i} consist entirely of zeros, then the elimination process is complete.

2.2 Reduced Row Echelon Form

2.2.1 Definition of Row Echelon Form

An $m \times n$ matrix E with rows E_{i*} and columns E_{*j} is said to be in row echelon form provided that:

- if E_{i*} consists entirely of zeros, then all rows below E_{i*} are also zero
- if the first nonzero entry in E_{i*} lies in the j-th column, then all entries below the i-th row in columns $E_{*1}, E_{*2}, \cdots, E_{*j}$, are zero.

2.2.2 Definition of Rank of a Matrix

Suppose $A_{m \times n}$ is reduced by row operations to an echelon form E. The rank of A is defined to be:

$$\begin{aligned}\textbf{rank} &= \text{number of pivots}\\ &= \text{number of nonzero rows in } E\\ &= \text{number of basic columns in } A\end{aligned}$$

where the **basic columns** of A are defined to be those columns in A that contain the pivot positions.

2.2.3 Definition of Row Echelon Form determining

A matrix $E_{m \times n}$ is said to be in reduced row echelon form provided that:

- E is in row echelon form
- the first nonzero entry in each row (i.e. each pivot) is 1
- all entries above each pivot are 0

2.2.4 Properties of Reduced Row Echelon Forms

Unlike row echelon forms, the reduced row echelon form derived from a matrix A is **uniquely** determined by A; we denote this unique reduced row echelon form by E_A

Relationship between Basci Column with Reduced Row Echelon Forms:
Each nonbasic column of A can be expressed as a linear combination of the basic columns $[A_{*i}, A_{*j}, A_{*k}]$, and exactly the same relationships hold in E_A, in addition, the multipliers appearing in these relationships are precisely the nonzero entries in the two nonbasic columns of E_A, furthermore, only the basic columns to the left of a given nonbasic column are needed in the representation formula.

2.3 Consistency of Linear Systems

2.3.1 Definition

A system of m linear equations in n unknowns is said to be a **consistent** system if it possesses **at least one solution**. If there are **no solutions**, then the system is called **inconsistent**.

2.3.2 Criteria for Determining Consistency

- let $[A|b]$ be the augmented matrix associated with the system and reduce it by row operations to a row echelon form;
- if, somewhere in the elimination process, a situation arises in which the only nonzero entry in a row appears on the right-hand side:

$$\left(\begin{array}{cccccc|c} * & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \\ 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \alpha \neq 0$$

  then the original system must be inconsistent.
- This method is also equivalent to identifying pivot in the last column.

- Conclusion:
  If the system is consistent, then b must be a nonbasic column of $[A|b]$, or equivalently b must be a linear combination of the (basic) columns from A
- consistency is equivalent to:

$$rank[A|b] = rank(A)$$

2.4 Homogeneous Systems

2.4.1 Definition

All the RHS entries are zero (homogeneous system)

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

2.4.2 Trivial vs. Nontrivial Solutions

Consistency is never an issue for homogeneous systems since the zero solution:

$$x_1 + x_2 + \cdots + x_n = 0$$

is always a solution, the so-called trivial solution.

Other than the trivial solution would be the nontrivial solution which can be obtained by the Gaussian elimination.

2.4.3 Example

For the following system,

$$\begin{aligned} 1x_1 + 2x_2 + 2x_3 + 3x_4 &= 0 \\ 2x_1 + 3x_2 + 1x_3 + 3x_4 &= 0 \\ 3x_1 + 6x_2 + 1x_3 + 4x_4 &= 0 \end{aligned}$$

Written in augmented matrix and conduct Gaussian elimination:

$$\left(\begin{array}{cccc|c} 1 & 2 & 2 & 3 & 0 \\ 2 & 1 & 1 & 3 & 0 \\ 3 & 6 & 1 & 4 & 0 \end{array}\right) \rightarrow \left(\begin{array}{cccc|c} 1 & 2 & 2 & 3 & 0 \\ 0 & 0 & -3 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

After Gaussian elimination, the matrix is equivalent to the following equation system:

$$\begin{aligned} 1x_1 + 2x_2 + 2x_3 + 3x_4 &= 0 \\ -3x_3 - 3x_4 &= 0 \end{aligned}$$

Which leads to the nontrivial solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

with the understanding that x_2 and x_4 are free variables that can range over all real numbers, this representation will be called the general solution of the homogeneous system.

2.4.4 General Cases

Now consider a general homogeneous system $[A|0]$ of m linear equations in n unknowns; if $rank(A) = r$, then:

- there are exactly r basic variables and $n - r$ free variables corresponding to the r basic and $n - r$ nonbasic columns of A;
- reducing A to a row echelon form using Gaussian elimination and solving for the basic variables in terms of the free variables produces the general solution:

  $$x = x_{f_1}h_1 + x_{f_2}h_2 + \cdots + x_{f_{n-r}}h_{n-r}$$

  where $x_{f_1}h_1, x_{f_2}h_2, \cdots, x_{f_{n-r}}h_{n-r}$ are the free variables and $h_1, h_2, \cdots, h_{n-r}$ are $n \times 1$ columns that represent particular solutions of the system; the h_is are **independent** of the particular row echelon form used;
- the system possesses a unique solution (the trivial solution) if and only if $rank(A) = n$.

2.5 Nonhomogeneous Systems

2.5.1 Definition

If at least one of the RHS entries is nonzero, the system is nonhomogeneous:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Assuming the existence of a solution, the solutions of a (consistent) nonhomogeneous system can be obtained by exactly the same method used for homogeneous systems.

2.5.2 Example

For the following system,

$$\begin{aligned} 1x_1 + 2x_2 + 2x_3 + 3x_4 &= 4 \\ 2x_1 + 4x_2 + 1x_3 + 3x_4 &= 5 \\ 3x_1 + 6x_2 + 1x_3 + 4x_4 &= 7 \end{aligned}$$

Written in augmented matrix and reduced to row echelon form:

$$\left(\begin{array}{cccc|c} 1 & 2 & 2 & 3 & 4 \\ 2 & 1 & 1 & 3 & 5 \\ 3 & 6 & 1 & 4 & 7 \end{array}\right) \rightarrow \left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

After Gaussian elimination, the matrix is equivalent to the following equation system:

$$\begin{aligned} x_1 + 2x_2 + x_4 &= 2 \\ x_3 + x_4 &= 1 \end{aligned}$$

Which leads to the nontrivial solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

Note: In the exam, you should write the answer in form of:

$$answer = particular\ solution + homogenous\ solution$$

Chapter 3
Matrix Algebra

3.1 Introduction

3.1.1 Function of Matrix

- the elementary row operations in Gaussian elimination can be realized using a chain of multiplications by **elementary matrices**;
- besides linear systems, linear functions that **map** m-vectors to n-vectors can also be represented by and manipulated using matrices;
- even problems involving **graphs** or **Markov chain processes** can be conveniently analyzed using matrices.

3.2 Addition and Transposition

3.2.1 Notation

- a scalar is a complex number (unless otherwise stated)
- $\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers, respectively
- $\mathbb{R}^n$ and $\mathbb{C}^n$ denote the set of all n-tuples of real and complex numbers,respectively
- $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ denote the set of all $m\times n$ matrices containing real and complex numbers, respectively

3.2.2 Equal Matrix

Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be equal if they have the **same shape** and **the corresponding entries are equal**.

Additionally, matrix can be considered as the composition of several column vectors or row vector.

3.2.3 Addition of Matrices

If A and B are $m \times n$ matrices, the sum of A and B is defined to be the $m \times n$ matrix $A + B$ obtained by adding the corresponding entries, that is:

$$[A + B]_{ij} = [A]_{ij} + [B]_{ij} \quad i = 1, 2, \cdots , m; j = 1, 2, \cdots , n$$

3.2.4 Additive Inverse

The matrix $-A$, called the additive inverse of A, is defined to be the matrix obtained by negating each entry of A, that is,

$$[-A]_{ij} = -[A]_{ij} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

3.2.5 Subtraction of Matrices

This allows matrix subtraction to be defined in the natural way: if A and B are $m \times n$ matrices, the difference of A and B is defined to be the $m \times n$ matrix $A - B = A + (-B)$ so that:

$$[A - B]_{ij} = [A]_{ij} + [-B]_{ij} = [A]_{ij} - [B]_{ij} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

3.2.6 Properties of Matrix Addition

For $m \times n$ matrices A, B, and C, the following properties hold:

- **Closure under addition:** $A + B$ is again an $m \times n$ matrix
- **Associative law:** $(A + B) + C = A + (B + C)$
- **Commutative law:** $A + B = B + A$
- **Existence of additive identity:** there exists one and only one element 0 in $\mathbb{C}^{m \times n}$ such that $A+0 = A$ for all $A \in \mathbb{C}^{m \times n}$
- **Existence of additive inverse:** for each $A \in \mathbb{C}^{m \times n}$, there exists one and only one element (-A) in $\mathbb{C}^{m \times n}$, such that $A + (-A) = 0$

3.2.7 Scalar Multiplication

The scalar multiplication of α scalar a and a matrix A, denoted by αA, is defined to be the matrix obtained by multiplying each entry of A by α, that is,

$$[\alpha A]_{ij} = \alpha[A]_{ij} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

3.2.8 Properties of Scalar Multiplication

For $m \times n$ matrices A, B, and scalar α, the following properties hold:

- **Closure under addition:** αA is again an $m \times n$ matrix
- **Associative law:** $(\alpha\beta)A = \alpha(\beta A)$
- **Distributive law one:** $\alpha(A + B) = \alpha A + \alpha B$
- **Distributive law two:** $(\alpha + \beta) = \alpha A + \beta A$
- **Scalar identity:** $1 \cdot A = A$ for all $A \in \mathbb{C}^{m \times n}$

3.2.9 Transpose and Conjugate Transpose

If A is an $m \times n$ matrix:

- **Transpose Matrix:** A^T

$$[A^T]_{ij} = [A]_{ji} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

- **Conjugate Matrix:** $\overline{A}$

$$[\overline{A}]_{ij} = \overline{[A]_{ij}} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

- **Conjugate Transpose Matrix:** A^* or $A^\dagger$

$$[A^*]_{ij} = \overline{[A]_{ij}} \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$$

3.2.10 Properties of Matrix Transposition

For $m \times n$ matrices A, B and scalar α, the following holds:

- $(A+B)^T = A^T + B^T$, $(A+B)^* = A^* + B^*$
- $(\alpha A)^T = \alpha A^T$, $(\alpha A)^* = \alpha A^*$

Proof: for $i = 1, 2, \cdots, m; j = 1, 2, \cdots, n$

$$\begin{aligned}
[(A+B)^T]_{ij} &= [A+B]_{ji} = [A]_{ji} + [B]_{ji} = [A^T]_{ij} + [B^T]_{ij} = [A^T + B^T]_{ij} \\
[(A+B)^*]_{ij} &= \overline{[A+B]_{ij}} = \overline{[A]_{ij}} + \overline{[B]_{ij}} = [A^*]_{ij} + [B^*]_{ij} = [A^* + B^*]_{ij} \\
[(\alpha A)^T]_{ij} &= [\alpha A]_{ji} = \alpha [A]_{ji} = \alpha [A^T]_{ij} \Rightarrow (\alpha A)^T = \alpha A^T \\
[(\alpha A)^*]_{ij} &= \overline{[\alpha A]_{ij}} = \overline{\alpha [A]_{ij}} = \alpha [A^*]_{ij} \Rightarrow (\alpha A)^* = \alpha A^*
\end{aligned}$$

3.2.11 Symmetries

Let A be an $n \times n$ **square matrix**:

- **Symmetric:** $A^T = A$

$$[A^T]_{ij} = [A]_{ji} = [A]_{ij} \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, n$$

- **Skew-symmetric:** $A^T = -A$

$$[A^T]_{ij} = [A]_{ji} = -[A]_{ij} \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, n$$

- **Hermitian:** $A^* = A$

$$[A^*]_{ij} = \overline{[A]_{ij}} = [A]_{ij} \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, n$$

- **Skew-hermitian:** $A^* = -A$

$$[A^*]_{ij} = \overline{[A]_{ij}} = -[A]_{ij} \quad i = 1, 2, \cdots, n; j = 1, 2, \cdots, n$$

3.3 Linearity

3.3.1 Definition

Suppose that $\mathcal{D}$ and $\mathcal{R}$ are two sets equipped with an addition and a scalar multiplication operation (consider, for example, $\mathcal{D} = \mathbb{C}^n$ and $\mathcal{R} = \mathbb{C}^m$). A function f that maps points in $\mathcal{D}$ to points in $\mathcal{R}$ is said to be a linear function if f satisfies:

$$f(x+y) = f(x) + f(y) \ (\textbf{additivity})$$
$$f(\alpha x) = \alpha f(x) \ (\textbf{homogeneity} \text{ of degree } 1)$$

or equivalently,

$$f(\alpha x + y) = \alpha f(x) + f(y)$$

for all $x, y \in \mathcal{D}$ and all scalars α.

Additionally, linearity comes from the linear (homomorphism) map $Hom(\mathcal{D}, \mathcal{R})$ (or $Hom(\mathbb{C}^n, \mathbb{C}^m)$) in the vector space.

3.3.2 Linear Operators Examples

- **The zero map** between two left-modules (or two right-modules) over the same ring is always linear.
- **The identity map** on any module is a linear operator.
- Any **homothecy** centered in the origin of a vector space, $v \mapsto cv$ where c is a scalar, is a linear operator. This does not hold in general for modules, where such a map might only be **semilinear**.
- For real numbers, the map $x \mapsto x+1$ is **not** linear.
- **Differentiation** defines a linear map from the space of all differentiable functions to the space of all functions. It also defines a linear operator on the space of all smooth functions.
- **The (definite) integral** over some interval I is a linear map from the space of all real-valued integrable functions on I to R.
- etc.

Inclass **important examples** of linear operators (proof needed):

- **the transposition function** $f(X_{mn}) = X^T$ is linear since:
$$(A+B)^T = A^T + B^T \quad (\alpha A)^T = \alpha A^T$$
- **the trace function** of an $n \times n$ matrix $X = [x_{ij}]$, defined by:
$$f(X_{n\times n}) = trace(X) := x_{11} + x_{22} + \cdots + x_{nn} = \sum_{i=1}^{n} x_{ii}$$
is linear since:
$$f(\alpha A + B) = \sum_{i=1}^{n}(\alpha a_{ii} + b_{ii}) = \alpha \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = \alpha f(A) + f(B)$$
for all $A = [a_{ij}], B = [b_{ij}] \in \mathbb{C}^{n\times n}$ and all scalars α.
- **Linear system** is also linear.
By considering the linear system as a function (linear transform function):
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= u_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= u_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= u_m \end{aligned}$$
to be a function $u = f(x)$ that maps $x = (x_1, x_2, \cdots, x_n)^T \in \mathbb{C}^n$ to $u = (u_1, u_2, \cdots, u_m)^T \in \mathbb{C}^m$; $f(x)$ is linear since:
$$f(\alpha x + y) = \sum_{j=1}^{n}(\alpha x_j A_{*j} + y_j A_{*j}) = \alpha \sum_{i=1}^{n} x_j A_{*j} + \sum_{i=1}^{n} y_j A_{*j} = \alpha f(x) + f(y)$$
Notes the following: in the above derivation, we have used the fact that for any $z = (z_1, z_2, \cdots, z_n)^T \in \mathbb{C}^n$
$$f(x) = \begin{pmatrix} a_{11}z_1 + a_{12}z_2 + \cdots + a_{1n}z_n \\ a_{21}z_1 + a_{22}z_2 + \cdots + a_{2n}z_n \\ \vdots \\ a_{n1}z_1 + a_{n2}z_2 + \cdots + a_{nn}z_n \end{pmatrix} = \begin{pmatrix} a_{11}z_1 \\ a_{21}z_1 \\ \vdots \\ a_{n1}z_1 \end{pmatrix} + \begin{pmatrix} a_{12}z_2 \\ a_{22}z_2 \\ \vdots \\ a_{n2}z_2 \end{pmatrix} + \cdots + \begin{pmatrix} a_{1n}z_n \\ a_{2n}z_n \\ \vdots \\ a_{nn}z_n \end{pmatrix} = \sum_{j=1}^{n} z_j A_{*j}$$

3.3.3 Linear Combination

The following terminology will be used in subsequent development: for scalars α_j and matrices X_j, the expression:
$$\alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha_{1n}X_n = \sum_{j=1}^{n} \alpha_j X_{*j}$$
is called a linear combination of the X_js.

3.4 Matrix Multiplication

3.4.1 Motivation

Before proceeding to defining matrix multiplication, lets consider the problem of composing two linear functions:

$$f(x) = f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}$$
$$g(x) = g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_{11}x_1 + b_{12}x_2 \\ b_{21}x_1 + b_{22}x_2 \end{pmatrix}$$

Then, we can construct another function:

$$h(x) := f(g(x)) = \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21})x_1 + (a_{11}b_{11} + a_{12}b_{22})x_2 \\ (a_{21}b_{11} + a_{22}b_{21})x_1 + (a_{21}b_{12} + a_{22}b_{22})x_2 \end{pmatrix}$$

Using matrix to simplify those function:

$$F = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad G = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Therefore, we can define the multiplication of matrix as the following:

$$h(x) := f(g(x)) = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{11} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

3.4.2 General Definition of Matrix Multiplication

Matrices A and B are said to be conformable for multiplication in the order AB if A has exactly as many columns as B has rows, i.e. A is $m \times p$ and B is $p \times n$ (Inner index are the same). For conformable matrices $A_{m\times p} = [a_{ij}]$ and $B_{p\times n} = [b_{ij}]$, the matrix product AB is defined to be the $m \times n$ matrix whose (i, j)-entry is the inner product of the i-th row of A with the j-th column in B, that is,

$$[AB]_{ij} = A_{i*}B_{*j} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

In case A and B fail to be conformable, then no product AB is defined.

3.4.3 An Illustration

$$[AB]_{ij} = A_{i*}B_{*j} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

$$\begin{pmatrix} * & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \vdots & a_{ip} \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & * \end{pmatrix}_{m\times p} \begin{pmatrix} * \cdots & b_{1j} & \cdots * \\ * \cdots & b_{2j} & \cdots * \\ \vdots & \vdots & \vdots \\ * \cdots & b_{pj} & \cdots * \end{pmatrix}_{p\times n} = \begin{pmatrix} * \cdots & * & \cdots * \\ \vdots & \vdots & \vdots \\ * \cdots & [AB]_{ij} & \cdots * \\ \vdots & \vdots & \vdots \\ * \cdots & * & \cdots * \end{pmatrix}_{m\times n}$$

3.4.4 Differences between Matrix and Scalar Multiplications

- matrix multiplication is a **noncommutative** operation: $AB \neq BA$
- the product AB of nonzero matrices $A \neq 0, B \neq 0$ can be 0
- the **cancellation law fails** for matrix multiplication

$$AB = AC, A \neq 0 \text{ do not necessarily implies } B = C$$

3.4.5 Rows and Columns of a Matrix Product

There are various ways to express the individual columns and rows of a matrix product: if $A = [a_{ij}]$ is $m \times p$ and $B = [b_{ij}]$ is $p \times n$, then:

$$[AB]_{i*} = A_{i*}B = a_{i1}B_{1*} + a_{i2}B_{2*} + \cdots + a_{ip}b_{p*} = \sum_{k=1}^{p} a_{ik}B_{k*}$$
$$[AB]_{*j} = AB_{*j} = A_{*1}b_{1j} + A_{*2}b_{2j} + \cdots + A_{*p}b_{pj} = \sum_{k=1}^{p} A_{*k}b_{kj}$$

In particular, these equations show that rows of AB are linear combinations of rows of B, while columns of AB are linear combinations of columns of A.

Verification:

- for example, in component form, $[AB]_{*j}$ can be written as:

$$[AB]_{*j} = \begin{pmatrix} [AB]_{1j} \\ [AB]_{2j} \\ \vdots \\ [AB]_{mj} \end{pmatrix} = \begin{pmatrix} a_{11}b_{1j} + a_{12}b_{2j} + \cdots + a_{1p}b_{pj} \\ a_{21}b_{1j} + a_{22}b_{2j} + \cdots + a_{2p}b_{pj} \\ \vdots \\ a_{m1}b_{1j} + a_{m2}b_{2j} + \cdots + a_{mp}b_{pj} \end{pmatrix} =$$
$$\begin{pmatrix} a_{11}b_{1j} \\ a_{21}b_{1j} \\ \vdots \\ a_{m1}b_{1j} \end{pmatrix} + \begin{pmatrix} a_{12}b_{2j} \\ a_{22}b_{2j} \\ \vdots \\ a_{m2}b_{2j} \end{pmatrix} + \cdots + \begin{pmatrix} a_{1p}b_{pj} \\ a_{2p}b_{pj} \\ \vdots \\ a_{mp}b_{pj} \end{pmatrix} = A_{*1}b_{1j} + A_{*2}b_{2j} + \cdots + A_{*p}b_{pj}$$

- in addition, using block matrix multiplication, the above expression can also be written as

$$[AB]_{*J} = A_{*1}b_{1j} + A_{*2}b_{2j} + \cdots + A_{*p}b_{pj}$$
$$= [A_{*1}|A_{*2}|\cdots|A_{*p}] \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{pj} \end{pmatrix} = AB_{*j}$$

3.4.6 Matrix Representation of Linear Systems

Matrix multiplication provides a convenient representation for linear systems: every system of m equations in n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as a single matrix equation $Ax = b$ in which:

$$A_{m,n} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

3.4.7 Application

a linear system $A_{m\times n}x_{n\times 1} = b_{m\times 1}$ is consistent if and only if b is a linear combination of the (basic) columns in A

3.5 Properties of Matrix Multiplication

Although not commutative and not satisfy the cancellation law, matrix multiplication is both distributive and associative; more precisely, for conformable matrices A, B, and C, there holds:

- **left-hand distributive law:** $A(B+C) = AB + AC$

$$[A(B+C)]_{ij} = A_{i*}(B+C)_{*j} = \sum_k [A]_{ik}[B+C]_{kj} = \sum_k [A]_{ik}([B]_{kj} + [C]_{kj}) = \sum_k ([A]_{ik}[B]_{kj} + [A]_{ik}[C]_{kj}) = \sum_k ([A]_{ik}[B]_{kj} + \sum_k [A]_{ik}[C]_{kj}) = [A]_{i*}[B]_{*j} + [A]_{i*}[C]_{*j} = [AB]_{ij} + [AC]_{ij} = [AB+AC]_{ij}$$

- **right-hand distributive law:** $(A+B)C = AC + BC$
- **associative law:** $A(BC) = (AB)C$

$$[A(BC)]_{ij} = [A]_{i*}[BC]_{*j} = [A]_{i*}\sum_k^q B_{*k}c_{kj} = \sum_k^q [A]_{i*}B_{*k}c_{kj} = \sum_k^q [AB]_{ik}c_{kj} = [AB]_{i*}C_{*j} = [(AB)C]_{ij}$$

3.5.1 Linearity of Matrix Multiplication

Let A be an $m \times n$ matrix, and f be the function defined by matrix multiplication:

$$f(X_{n\times p}) = AX$$

the left-hand distributive law guarantees that f is linear since

$$f(\alpha X + Y) = \alpha AX + AY = \alpha f(X) + f(Y)$$

for all scalars a and all $n \times p$ matrices X and Y.

3.5.2 Multiplicative Identity

$$I_{n\times n} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

for every $m \times n$ matrix A, there holds:

$$AI_n = A \qquad I_m A = A.$$

3.5.3 Powers of Square Matrices

$$A^0 = I_n$$
$$A^r A^s = A^{r+s}$$
$$(A^r)^s = A^{rs}$$

Note that powers of non-square matrices are never defined.
Note that: $(A+B)^2 = A^2 + BA + AB + B^2$ $(BA \neq AB)$

3.5.4 Transpose of a Matrix Product

The operation of transposition has an interesting effect upon a matrix product: a reversal of order occurs; more precisely, for conformable matrices A and B, there holds:

$$(AB)^T = B^T A^T \qquad (AB)^* = B^* A^*$$

Proof: By definition

$$(AB)^T_{ij} = [AB]ji = A_{j*}B_{*i}$$

Consider the (i, j)-entry of the matrix $B^T A^T$ and write:

$$[B^T A^T]_{ij} = (B^T)_{i*}(A^T)_{*j} = \sum_k [B^T]_{ik}[A^T]_{kj} = \sum_k [B]_{ki}[A]_{jk} = \sum_k [A]_{jk}[B]_{ki} = A_{j*}B_{*i}$$

Therefore, $(AB)^T_{ij} = [B^T A^T]_{ij}$.

Application: For every $m \times n$ matrix A, the products $(A^T A)_{nn}$ and $(AA^T)_{mm}$ are both symmetric matrices because:

$$\begin{aligned}(A^T A)^T &= A^T (A^T)^T = A^T A \\ (AA^T)^T &= (A^T)^T A^T = AA^T\end{aligned}$$

The matrix $A^T A$ is significant because it not only contains the singular values of A but also plays a central role in the least-squares solution of the overdetermined system $Ax = b$.

3.5.5 Trace of a Matrix Product

For $m \times n$ matrix A and $n \times m$ matrix B, the products AB and BA are both defined but may not be equal; nevertheless,

$$trace(AB) = trace(BA)$$

Proof: By definition

$$trace(AB) = \sum_{i=1}^{m} A_{i*}B_{*i} = \sum_{k=1}^{m} A_{k*}B_{*k} = trace(BA)$$

Note: This is true in spite of the fact that AB is $m \times m$ while BA is $n \times n$. Furthermore, this result can be extended to say that any product of conformable matrices can be permuted **cyclically** without altering the trace of the product. For example,

$$\begin{gathered} trace(ABC) = trace(BCA) = trace(CAB) \\ trace(ABC) \neq trace(BAC). \end{gathered}$$

3.5.6 Block Matrix Multiplication

Submatrices - matrix contained within another matrix

A useful technique in matrix multiplication is to partition one or both factors into submatrices, or simply blocks: if

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,r} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ A_{s,1} & A_{s,2} & \cdots & A_{s,r} \end{pmatrix} \qquad B = \begin{pmatrix} B_{1,1} & B_{1,2} & \cdots & B_{1,t} \\ B_{2,1} & B_{2,2} & \cdots & B_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ B_{r,1} & B_{r,2} & \cdots & B_{r,t} \end{pmatrix}$$

are two conformably partitioned block matrices in which each pair (A_{ik}, B_{kj}) is conformable, then the (i, j)-block in AB is given by:

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ir}B_{rj}$$

3.6 Matrix Inversion

3.6.1 Definition of Matrix Inversion

For a given square matrix $A_{n\times n}$, the matrix $B_{n\times n}$ that satisfies the conditions:

$$AB = I_n \qquad BA = I_n$$

is called the inverse of A and is denoted by $B = A^{-1}$. An invertible matrix is said to be **nonsingular**, and a square matrix with no inverse (such as the zero matrix) is called a **singular** matrix.

Note that matrix inversion is defined for square matrices only: the condition $AA^{-1} = A^{-1}A$ rules out inverses of non-square matrices.

3.6.2 Example

Inverse of 2 2 Matrices:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \textit{and} \quad \delta = ad - cb \neq 0$$
$$A^{-1} = \tfrac{1}{\delta}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

3.6.3 Solving Matrix Equations

- Uniqueness of Inverse Matrix:

$$X_1 = X_1 A X_2 = X_2$$

- Matrix Equations
  - if A is nonsingular, then the matrix equation $A_{n\times n}X_{n\times p} = B_{n\times p}$ has a unique solution given by:

$$X = A^{-1}B$$

  - in particular, for a system of n linear equations in n unknowns, written in matrix form as $A_{n\times n}x_{n\times 1} = b_{n\times 1}$, there exists a unique solution $x = A^{-1}b$ if (and only if) A is nonsingular.

3.6.4 Existence of Matrix Inverse

Theorem 4 (Characterization of nonsingular matrices)
For an $n \times n$ matrix A, the following statements are equivalent:

1. A^{-1} exists (A is nonsingular);
2. $rank(A) = n$;
3. $A \xrightarrow{Gauss-Jordan} I$;
4. $Ax = 0$ has only the trivial solution $x = 0$.

3.6.5 Computing Matrix Inverse: the Algorithm

To compute the inverse X of a nonsingular matrix $A_{n\times n}$:

1. recall first that determining A^{-1} is equivalent to solving $AX = I$, which is the same as solving the n linear systems $Ax = I_{*j}$;

2. if A is nonsingular, then Gauss-Jordan reduces $[A|I_{*j}] to [I|X_{*j}]$;
3. by applying Gauss-Jordan to $[A|I]$, the n linear systems $Ax = I_{*j}$ can be solved simultaneously, and the end result is $[I|A^{-1}]$;
4. the only way for Gauss-Jordan to fail is for a row of zeros to emerge in the left-hand side of $[A|I]$ at some point during the elimination process, and this occurs if and only if A is singular.

Example:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

if it is nonsingular. Applying Gauss-Jordan to $[A|I]$ yields:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 2 & -1 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right)$$

so A^{-1} exists and equals:

$$A^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

Operation counts:
It is not difficult to show that computing the inverse of an $n \times n$ nonsingular matrix by Gauss-Jordan requires n^3 multiplications/divisions and $n^3 + 2n^2 + n$ additions/subtractions.

3.6.6 Properties of Matrix Inversion

For nonsingular $n \times n$ matrices A and B, the following holds:

- $(A^{-1})^{-1} = A$;
- the product AB is also nonsingular;
- $(AB)^{-1} = B^{-1}A^{-1}$ (the reverse order law);
- $(A^{-1})^T = (A^T)^{-1} \qquad and \qquad (A^{-1})^* = (A^*)^{-1}$.

Proof:
The first property follows directly from the definition of inversion.

For the second and third one:
Let $X = B^{-1}A^{-1}$ and verify that $(AB)X = I$ by writing:

$$(AB)X = (AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = A(I)A^{-1} = AA^{-1} = I$$

and also from the previous conclusion, it guaranteed that $X(AB) = I$.
For the last property, let $X = (A^{-1})^T$ and verify that $A^T X = I$:

$$A^T X = A^T(A^{-1})^T = (AA^{-1})^T = I^T = I$$

and the proof for conjugate is similar.

3.6.7 Products of Nonsingular Matrices

In this regards, the class of rank-n (nonsingular) matrices in $\mathbb{C}^{n\times n}$ is special since the product of any two rank-n matrices again has rank n; more generally, if $A_1, A_2, \cdots, A_k \in \mathbb{C}^{n\times n}$ each has rank n, then their product also has rank n, and

$$(A_1 A_2 \cdots A_k) = A_k^{-1} \cdots A_2^{-1} A_1^{-1}$$

3.6.8 Inverses of Sums

Sherman-Morrison Formula: If $A_{n\times n}$ is nonsingular matrix and if c and d are $n \times 1$ columns such that:

$$1 + d^T A^{-1} = A^{-1} - \frac{A^{-1}cd^T A^{-1}}{1+d^T A^{-1}c}$$

The Sherman-Morrison-Woodbury formula is a generalization. If C and D are $n \times k$ such that $(I + D^T A^{-1} C)^{-1}$ exists, then:

$$(A + CD^T)^{-1} = A^{-1} - A^{-1}C(I + D^T A^{-1}C)^{-1}D^T A^{-1}$$

Neumann Series: If $\lim_{n\to\infty} A^n = 0$. then $I - A$ is nonsingular and

$$(I - A)^{-1} = I + A + A^2 + \cdots = \sum_{k=0}^{\infty} A^k$$

3.7 Elementary Matrices and Equivalence

3.7.1 Definition

Matrices of the form $I - uv^T$, where u and v are $n \times 1$ columns such that $v^T u \neq 1$ are called **elementary matrices**, and we know from the previous conclusion that all such matrices are nonsingular and

$$I + uv^T = I - \frac{uv^T}{v^T u - 1}$$

Notice that inverses of elementary matrices are elementary matrices.

3.7.2 Three types of the Elementary Matrices

- Type I is interchanging rows (columns) i and j.
- Type II is multiplying rows (columns) i by $\alpha \neq 0$.
- Type III is adding a multiple if rows (columns) i to rows (columns) j.

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix}$$

3.7.3 Properties of Elementary Matrices

- when used as a **left-hand multiplier**, an elementary matrix of type I, II, or III executes the corresponding **row operation**;
- when used as a **right-hand multiplier**, an elementary matrix of type I, II, or III executes the corresponding **column operation**.

Example:

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 13 \end{pmatrix} \xrightarrow{R_2 - 2R_1, R_3 - 3R_1, R_2 \leftrightarrow R_3, R_1 - 4R_2} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

This reduction can be accomplished by series left-hand multiplications with the corresponding elementary matrices.

$$\begin{pmatrix} 1 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} A = E_A$$

3.7.4 Products of Elementary Matrices

It turns out that a matrix A is nonsingular if and only if it is the product of elementary matrices of type I, II, or III.

Proof:
If A is nonsingular, then Gauss-Jordan reduces A to I by row operations with corresponding elementary matrices $G_1, G_2, \cdots, G_k$:
by Theorem.4 Part c

$$G_k \cdots G_2 G_1 A = I, \quad or \quad A = {G_1}^{-1} {G_2}^{-1} \cdots {G_k}^{-1}$$

Therefore, if $A = E_1 E_2 \cdots E_k$ is a product of elementary matrices, then A must be nonsingular because the E_is are nonsingular.

Notes: The inverse of elementary matrix is elementary matrix for sure.

3.7.5 Equivalence

whenever B can be derived from A by a combination of elementary row and column operations, we say that A and B are equivalent matrices and write $A \sim B$; in matrix terms,

$$\begin{array}{ll} A \sim B \iff PAQ = B & \text{for nonsingular P and Q} \\ A \sim^{row} B \iff PAQ = B & \text{for nonsingular P} \\ A \sim^{col} B \iff PAQ = B & \text{for nonsingular Q} \end{array}$$

3.7.6 Column and Row Relationship

- If $A \sim^{row} B$, then linear relationships existing among columns of A also hold among corresponding columns of B, that is,

$$B_{*k} = \sum_{j=1}^{n} \alpha_j B_{*j} \iff A_{*k} = \sum_{j=1}^{n} \alpha_j A_{*j}$$

- if $A \sim^{col} B$, then linear relationships existing among rows of A must also hold among corresponding rows of B;
- in summary, row equivalence preserves column relationships, and column equivalence preserves row relationships.

3.7.7 Rank Normal Form

Theorem 6(Rank normal form)
If A is an $m \times n$ matrix such that $rank(A) = r$, then

$$A \sim N_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where N_r is called the rank normal form of A. It is the end product of a complete reduction of A by using both row and column operations. Proof:

1. Since $A \sim^{row} E_A$, there is a nonsingular matrix P such that PA = EA.
2. If rank$(A) = r$, then E_A has r basic columns each of which is of the form e_i $(i = 1, \cdots, r)$. Apply column interchanges to E_A to move these columns to the far left of E_A and denote by Q_1 the product of the corresponding elementary matrices. The result is

$$PAQ_1 = \begin{pmatrix} I_r & J \\ 0 & 0 \end{pmatrix}$$

3. Multiplying both sides of this equation on the right by

$$Q_2 = \begin{pmatrix} I_r & -J \\ 0 & I \end{pmatrix}$$

produces $PAQ_1Q_2 = N_r$, which shows $A \sim N_r$

Example:
As an application, let's show:

$$rank \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = rank(A) + rank(B)$$

Indeed, if $rank(A) = r$ and $rank(B) = s$, then $A \sim N_r$ and $B \sim N_s$.
Consequently,

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \sim \begin{pmatrix} N_r & 0 \\ 0 & N_s \end{pmatrix}$$

which implies that

$$rank \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = r + s$$

3.7.8 Testing for Equivalence

Theorem 7 (Testing for equivalence):
For $m \times n$ matrices A and B the following statement are true:

- a.$A \sim B \iff rank(A) = rank(B)$
- b.$A \sim^{row} B \iff E_A = E_B$
- c.$A \sim^{col} B \iff E_{A^T} = E_{B^T}$

Note: in particular, that:

- either $A \sim^{row} B$ or $A \sim^{col} B$ implies $A \sim B$, but not vice versa
- multiplication by nonsingular matrices doesn't change rank.

Proof:
To prove (a), let N_r, N_s be the rank normal form of A, B where $rank(A) = r$ and $rank(B) = s$. Then note that $A \sim B$ if and only if:

$$N_r \sim A \sim B \sim N_s$$

and $N_r \sim N_s$ if and only if $r = s$. To establish (b), let E_A and E_B be the unique reduced row echelon form of A, B, and note that:

$$A \sim^{row} B \sim^{row} \sim^{row} E_B \sim^{row} E_A$$

if and only if $E_A = E_B$ (by uniqueness). Finally, the proof of (c) follows by observing

$$A \sim^{col} B \iff A^T \sim^{row} B^T$$

3.7.9 Transposition and Rank

For all $m \times n$ matrices,

$$rank(A) = rank(A^T), \qquad rank(A) = rank(A^*)$$

Proof:
Let $rank(A) = r$ and N_r be the rank normal form of A. The proof of the first half of the statement follows easily by observing that $A \sim N_r$ if and only if $A_T \sim N_r^T$, and by using the fact that $rank(N_r) = rank(N_r^T)$. The proof of the second half is similar.

3.8 The LU Factorization

Example: constructing LU factorization
Consider the following linear system:

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{pmatrix} \xrightarrow{R_2 \to 2R_1, R_3 \to 3R_1, R_3 \to 4R_2} \begin{pmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 4 \end{pmatrix} = \mathcal{U}$$

To write the elimination process using matrices, note that:

- each of these (type III) row operations can be realized by means of a left-hand multiplication, with the corresponding elementary matrix G_k and their product given by:

$$G_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix}$$
$$G_2G_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$$

- denote by $T_1 = G_2G_1$, $T_2 = G_3$ the (product of) elementary matrices that are used to eliminate the entries below the first and second pivot, respectively; these matrices have the form

$$T_k = I - c_k e_k^T$$

  where e_k is the k-th unit column vector and c_k is a column vector with zeros in the first k positions; in our case, we have

$$c_1 = (0, 2, 3)^T, \qquad c_2 = (0, 0, 4)^T$$

  note also that the entries in c_k are precisely the multipliers used in the k-th step of the elimination process;
  Note: In general, matrices of the form $T_k = I - \tau_k e_k^T \in \mathbb{C}^{n\times n}$ are called **Gauss transformations** if the first k component of $\tau \in \ltimes$ are zero.
- by observing that $e_j^T c_k = 0$ whenever $j \leq k$, the inversion formula for elementary matrices produces

$$T_k^{-1} = I + c_k e_k^T$$
$$T_1^{-1}T_2^{-1} = I + c_1e_1^T + c_2e_2^T = \mathcal{L}$$

  where $\mathcal{L}$ is a unit **lower-triangular matrix**; in our case, we have:

$$\mathcal{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{pmatrix}$$

  note also that the (i, j)-th entry in $\mathcal{L}$ is precisely the multiplier used to annihilate the (i, j)-th position during the Gaussian elimination.
- the reduction from A to $\mathcal{U}$ by now row operation is equivalent to

$$G_3G_2G_1A = \mathcal{U}, \text{ or } T_2T_1A = \mathcal{U}$$
$$\implies A = T_1^{-1}T_2^{-1}\mathcal{U} = \mathcal{L}\mathcal{U}$$

  where $\mathcal{L}$ is the a lower-triangular matrix and $\mathcal{U}$ is an upper-triangular matrix; this is naturally called the $\mathcal{L}\mathcal{U}$ factorization of A.

For the general case, if A is an $n \times n$ matrix and a zero pivot is never encountered during Gaussian elimination, then A can be factored as $A = \mathcal{L}\mathcal{U}$ where

- $\mathcal{L}$ is lower triangular and $\mathcal{U}$ is upper triangular
- $l_{ii} = 1$ and $u_{ii} \neq 0$ for each $i = 1, 2, \cdots, n$
- below the diagonal of L, the entry l_{ij} is the multiple of row j that is subtracted from row i in order to annihilate the (i, j)-position during Gaussian elimination
- $\mathcal{U}$ is the final result of Gaussian elimination applied to A;
- $\mathcal{L}$ and $\mathcal{U}$ are uniquely determined by the first two properties.

Some remarks:

- it is a common practice to successively overwrite the entries in A with the corresponding entries in $\mathcal{L}$ and $\mathcal{U}$ as Gaussian elimination evolves; for example, for the 3×3 matrix considered in the previous example, the elimination would yield:

$$A = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{pmatrix} \xrightarrow{R_2 \to 2R_1, R_3 \to 3R_1, R_3 \to 4R_2} \begin{pmatrix} 2 & 2 & 2 \\ \boxed{2} & 3 & 3 \\ \boxed{3} & \boxed{4} & 4 \end{pmatrix}$$

  where the boxed entries represent the lower-triangular matrix $\mathcal{L}$;
- once the $\mathcal{L}\sqcap$ factorization of a nonsingular matrix $A_{n \times n}$ is obtained, the linear system $Ax = b$ can be easily solved by writing $Ax = b$ as

$$\mathcal{L}y = b, \text{ where } y = \mathcal{U}x$$

  the solution consists of two steps: the lower-triangular system $\mathcal{L}y = b$ is first solved for y by forward substitution, and then the upper-triangular system $\mathcal{U}x = y$ is solved for x using back substitution;
- the advantage of this approach is that, once the $\mathcal{LU}$ factors of A are computed, any other linear system $Ax = \tilde{b}$ can be solved with only n^2 multiplications/divisions and $n^2 - n$ additions/subtractions, in contrast to the $2n^3/3$ operations required by Gaussian elimination;
- if zero pivots are encountered in the elimination process, then row interchanges are needed to bring a nonzero number into the pivot position; the end result of this procedure is a factorization

$$PA = \mathcal{LU}$$

  where P is a permutation matrix (i.e. a product of elementary interchange matrices of type I); it can be shown that this is possible as long as A is nonsingular.

Chapter 4
Vector Space

4.1 Spaces and Subspaces

4.1.1 Definition of Vector Space

Ten axioms for the vector space:

-
-
-
-
-

4.2 Four Fundamental Subspaces

4.3 Linear Independence

4.4 Basis and Dimension