Quantitative Trading Series

Quantitative and Qualitative Treatments to Capital Markets

and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022∼

Quantitative Trading Series

Quantitative and Qualitative Treatments to Capital Markets

and related bodies of knowledge

By

HangukQuant

Private Notes,

Quantitative Research

2022∼

Abstract

This book is designed to be a practical handbook for all finance professionals, practitioner or academic. It is an organization of the various knowledge domains, with a focus on drawing links in the intricate web between the theory and practice of finance that market participants seek to unfold. It contains discussions of trading anomalies, premias and inefficiencies. It contains discussions in discretionary and quantitative trading. Discussion stretches across theoretical work, such as statistical methods, linear algebra and financial mathematics. Applied work in equity research, quantitative research and macroeconomic theory is involved.

This work is attributed to the brilliant writers, academics, scientists and traders before me. Although we have tried to credit the referenced work where relevant, to give a complete reference for its source is impossible. The work has been organized and compiled from various texts, lecture notes, journals, blogs, personal communications and even scraps of scribbled notes from the author's time in college. These contain notes from blogs referencing journals, journals referencing blogs, blogs referencing blogs referring journals - you name it. We apologise if we have failed to credit your work. Other work is original. Readers may reach us at hangukquant@gmail.com. The updated notes are released at `hangukquant.substack.com`.

Faith is to have believe without seeing. This work is dedicated to those who placed their faith in me. To Jeong(s), Choi, Julian and my dearest friends who have shaped my world view and colored it rainbow.

Keywords:

Linear Algebra
Calculus Methods
Computer Methods
Global Macro Trading
Quantitative Research
Statistics & Probability Theory
Risk Premia and Market Inefficiencies
Equities Trading and Other Asset Classes

Table of Contents

Chapter 1

Introduction

1.1 Guidelines for Reviewing Work

The following are the stages of alpha formulations.

Idea 1 (This means to further explore the idea creatively. This is a precursor to a **Test**.).

Test 1 (This refers to parameterized research idea that is to be verified as a **Strategy**.).

Strategy 1 (This explores the implementation and characteristics of a **Test**.).

The following are the stages of theoretical formulations.

Definition 1 (Standard conventions and formal nomenclature are introduced.).

Problem 1 (A formalization of the problem statement is provided).

Exercise 1 (An example or working problem to demonstrate concepts discussed).

The following are stages of theoretical derivations

Lemma 1 (An important result used as is or for other derivations.).

Corollary 1 (An important aside of the theoretical work.).

Theorem 1 (A central result with derivations).

Result 1 (A central result without proof.).

The following are for declarative statements.

Proposition 1 (An opinion of sorts.).

Fact 1 (A statement of (almost) undeniable truth.).

Chapter 2

Ordinary Calculus

Theorem 2 (Integration By Parts)**.** *The integration by parts formula takes form*

$$\int udv = uv - \int vdu \tag{1}$$

Theorem 3 (L'Hopital's Rule)**.**

Chapter 3

Linear Algebra

Here we discuss concepts in linear algebra - notably the literature on this subject is divided into two different schools. One introduces linear algebra as the mathematics and computation of multiply defined linear equations. Here the focus is on teaching linear algebra as a tool for manipulation and computation in multi-dimensional spaces. Determinants are introduced early on, and focuses are on matrix operations. The second approach is to treat matrices as abstract objects, laying focus to the structure of linear operators and vector spaces. Determinants and matrices are only introduced later. Here we provide both - the first will focus on the linear algebraic manipulation of matrices on finite-dimensional, Euclidean spaces. The second treatment will focus on the underlying mathematics of the structure of linear operators and their properties, including the mathematics in infinite dimensional vector spaces and over complex fields. Some of these treatments and notes on Linear Algebra herein are adapted from the texts from Ma et al. [6], Axler [1] and Roman [10].

3.1 Computational Methods in the Euclidean Space

3.1.1 Linear Systems

Definition 2 (Linear Equation)**.** *A linear equation is one in which for variables* $\{x_1, \cdots, x_n\}$*, equation takes form*

$$\sum_{i=1}^{n} a_i x_i = b \tag{2}$$

where $a_i \in \mathbb{R}, i \in [n]$ *and* $b \in \mathbb{R}$*.*

Definition 3 (Zero Equation)**.** *A zero equation is a linear equation (see Definition 2) where all* $i \in [n], a_i = 0$ *and* $b = 0$*. That is,*

$$0x_1 + 0x_2 + \cdots 0x_n = 0. \tag{3}$$

The variables $x_i, i \in [n]$ in Definition 2 are not known and it is our task to solve for the solutions to these. The number of variables defines the dimensionality of our problem setting. For instance, see that the equation $ax + by + cz = d$ specify variables in the three-dimensional space $(x, y, z) \in \mathbb{R}^3$. For instance, the linear equation $z = 0$ specifies an xy-plane inside the xyz-space.

Definition 4 (Solution and Solution Sets to a Linear Equation). *A solution to a linear equation (see Definition 2 is a set of numbers* $\{x_1 = s_1, x_2 = s_2, \cdots, x_n = s_n\}$ *that satisfies the linear equation s.t.*

$$\sum_{i=1}^{n} a_i s_i = b. \tag{4}$$

The set of all such solutions is called a solution set to the equation. When the solution set is expressed by equations representing exactly the equations in the solution set, these set of expressions are known as the general solution.

For instance, in the xy-space, solutions to the equation $x + y = 1$ are points taking the form $(x, y) = (1-s, s)$ where $s \in \mathbb{R}$. In the xyz-space, the solutions to the same equation are points $(x, y, z) = (1-s, s, t)$ where $s, t \in \mathbb{R}$. The solution set form points on a plane. The solution set to the zero equation (see Definition 3) is the entire space $\mathbb{R}^n$ corresponding to the number of dimensions in the linear equation. The solution set to $\sum_i^n 0x_i \neq 0$ is $\emptyset$.

Definition 5 (Linear System). *A finite set of* m *equations in* n *variables* $x_1, \cdots x_n$ *is called a linear system and may be represented*

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i, \qquad i \in [m] \tag{5}$$

where $a_{ij}, i \in [m], j \in [n] \in \mathbb{R}$.

Definition 6 (Zero System). *A zero system is a linear system (see Definition 5) where all the constants* $a_{ji}, b_j, i \in [n], j \in [m]$ *are zero.*

Definition 7 (Solution and Solution Sets to a Linear System). *A solution to a linear system (see Definition 5) is a set of numbers* $\{x_1 = s_1, x_2 = s_2, \cdots x_n = s_n\}$ *that satisfies all linear equations (i.e)*

$$\sum_{i=1}^{n} a_{ji} s_i = b_j, \qquad j \in [m] \tag{6}$$

The set of all such solutions is called a solution set to the system. When the solution set is expressed by equations representing exactly the equations in the solution set, these set of expressions are known as the general solution.

Definition 8 (Consistency of Systems). *A system of linear equations that has solution set* $\neq \emptyset$ *is said to be consistent. Otherwise it is inconsistent.*

Every system of linear equations will either be consistent or inconsistent. Consistent systems have either a unique solution or infinitely many solutions.

Exercise 2. *Show that a linear system* $Ax = b$ *has either no solution, only one solution or infinitely many.*

Proof. If the linear system is not consistent then it must have no solution. Otherwise, it may have a unique solution, or more than one solution. Suppose there are two solutions $u \neq v$ and $Au = Av = b$. Then we may write

$$A(tu + (1-t)v) = tAu + (1-t)Av = tb + (1-t)b = tb + b - tb = b. \tag{7}$$

This is valid for all $t \in \mathbb{R}$, and so we have infinitely many solutions. □

For example, a system of two linear equations in two-dimensional space each representing a line has infinite solutions if they are the same line, no solution if they are parallel but different lines, and exactly one solution otherwise.

Exercise 3. *In the xyz-space, the two equations*

$$a_1x + b_1y + c_1z = d_1, \qquad (E_1) \tag{8}$$

$$a_2x + b_2y + c_2z = d_2, \qquad (E_2) \tag{9}$$

where $\exists a_1, b_1, c_1 \neq 0 \wedge \exists a_2, b_2, c_2 \neq 0$ *represents two planes. The solution to the system is the intersection between the two planes. Logicize that there is either no solution* $(E_1 // E_2)$ *or infinite number of solutions* $((E_1 = E_2) \vee (E_1$ *intersects* E_1 *on a line)).*

3.1.1.1 Elementary Row Operations (EROs)

Definition 9 (Augmented Matrix Representation of Linear Systems)**.** *See that the system of linear equations (Definition 5) given*

$$\forall j \in m, \qquad \sum_{i=1}^{n} a_{ji}x_i = b_j \tag{10}$$

may be represented by the rectangular array of numbers

$$\left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \cdots & \cdots & \cdots & \cdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right] \tag{11}$$

and we call this the augmented matrix of the system. We denote this $(A|b)$*. Sometimes, we omit this representation and just assign a single letter, say* M*, to represent the entire matrix.*

Definition 10 (Elementary Row Operations)**.** *When we solve for a linear system, we implicitly or explicitly perform the following operations; i) multiply equation by some non-zero* $k \in \mathbb{R}$*, (ii) interchange two equations, (iii) add a multiple of one equation to another. In the augmented matrix (see Definition 9), these operations correspond to multiplying a row by a non-zero constant, swapping two rows and adding a multiple of one row to another row respectively. These three operations are collectively known as the elementary row operations. We adopt the following notations*

1. $kR_i \equiv$ *multiply row* i *by* k*.*
2. $R_i \leftrightarrow R_j \equiv$ *swap rows* i, j*.*
3. $R_j + kR_i \equiv$ *add* k *times of row* i *to row* j*.*

Definition 11 (Row Equivalent Matrices)**.** *Two matrices* A, B *are said to be row equivalent if one may be obtained by another from a series of EROs. We denote this by*

$$A \overset{\mathcal{R}}{\equiv} B. \tag{12}$$

Theorem 4 (Solution Sets of Row Equivalent Augmented Matrix Represented Linear Systems)**.** *Two linear systems (Definition 5) with augmented matrix representations* $(A|b), (C|d)$ *have the same solution set if* $(A|b) \overset{\mathcal{R}}{\equiv} (C|d)$*.*

Proof. See proof in Exercise 14 using block matrix notations. □

3.1.1.2 Row-Echelon Forms

Definition 12 (Leading Entry)**.** *The first non-zero number in a row of the matrix is said to be the leading entry of the row.*

Definition 13 (Zero Row)**.** *Let the row representing a zero equation (see Definition 3) be called the zero row.*

Definition 14 (Zero Column)**.** *Let the column representing all zero coefficients in the representative linear system for some variable (see Definition 6) be called the zero column. That is, the column has all zeros.*

Definition 15 (Row-Echelon Form (REF))**.** *A matrix is said to be row-echelon if the following properties hold:*

1. *Zero rows (Definition 13) are grouped at the bottom of the matrix.*
2. *If any two successive rows are non-zero rows, then the higher row has a leading entry (Definition 12) occurring at a column that is to the left of the lower row.*

For matrix A, we denote its matrix REF as $REF(A)$.

Definition 16 (Pivot Points/Columns)**.** *The leading entries (Definition 12) of row-echelon matrices (Definition 15) are called pivot points. The column of a row-echelon form containing a pivot point is called a pivot column, and is otherwise a non-pivot column.*

Definition 17 (Reduced Row-Echelon Form (RREF))**.** *A reduced row-echelon-form matrix is a row-echelon-form matrix that has*

1. *All leading entries of non-zero row equal to one. (Definitions 12 and 13)*
2. *In each pivot column, all entries other than the pivot point is equal to zero. (Definition 16)*

For matrix A, we denote its matrix RREF as $RREF(A)$.

Note that a zero system is an REF (and also an RREF) by the Definitions 15 and 17. We show that obtaining the REF and RREF gives us an easy way to obtain the solution set to a linear system.

Exercise 4 (Finding solutions to REF, RREF Representations of Linear Systems; Back-Substitution Method)**.** *Find the solution set to the linear systems represented by the following augmented matrices. (see Definitions 9, 5 and 4)*

1.

$$\left[\begin{array}{ccc:c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right] \tag{13}$$

2.

$$\left[\begin{array}{ccccc:c} 0 & 2 & 2 & 1 & -2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 2 & 4 \end{array}\right] \tag{14}$$

3.

$$\left[\begin{array}{cccc:c} 1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{15}$$

4.

$$\left[\begin{array}{ccc:c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{16}$$

5.

$$\left[\begin{array}{cc:c} 3 & 1 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{array}\right] \tag{17}$$

Proof. 1. It is easy to see that $x_1 = 1, x_2 = 2, x_3 = 3$ is the unique solution this linear system.

2. Since this represents the linear system

$$2x_2 + 2x_3 + x_4 - 2x_5 = 2, \tag{18}$$
$$x_3 + x_4 + x_5 = 3, \tag{19}$$
$$2x_5 = 4, \tag{20}$$

solve. We let the solutions to variables of non-pivot columns be arbitrary. That is, $x_1 \in \mathbb{R}$. The third equation says $x_5 = 2$. Substituting into the second equation, get

$$x_3 + x_4 + 2 = 3, \tag{21}$$

so $x_3 = 1 - x_4$. Substituting into first equation,

$$2x_2 + 2(1 - x_4) + x_4 - 2 \cdot 2 = 2, \tag{22}$$

so $x_2 = 2 + \frac{1}{2}x_4$. So there are two free parameters, and we arrive at the general solution $(x_1, x_2, x_3, x_4, x_5) = (s, 2 + \frac{1}{2}t, 1 - t, t, 2)$, where $s, t \in \mathbb{R}$. This technique is known as the back-substitution method.

3. By the same back-substitution method, arrive at the general solution $(x_1, x_2, x_3, x_4) = (-2 + s - 3t, s, 5 - 2t, t)$ where $s, t \in \mathbb{R}$.

4. The solution set is $(r, s, t) = \mathbb{R}^3$.

5. This system is inconsistent! (Definition 8)

□

3.1.1.3 Gaussian Elimination Methods

Let $A \stackrel{\mathcal{R}}{\equiv} R$. If R is (R)REF, R is said to (reduced) row-echelon form of A and A is said to have (R)REF form R.

Theorem 5 (Gaussian Elimination/Row Reduction and Gauss-Jordan Elimination)**.** *We outline the algorithm to reduce a matrix A to its REF B.*

1. *Locate the leftmost non-zero column (see Definition 14).*
2. *If this happens to be the top-most column, then continue. Else, swap the top row with the row corresponding to the leading entry (Definition 12) found in the previous step.*
3. *For each row below the top row, add a suitable multiple so that all leading entries below the leading entry of the top row equals zero.*
4. *From the second row onwards, repeat algorithm from step 1 applied to the submatrix until REF is obtained.*

To further get a RREF from REF obtained,

5. *Multiply a suitable constant to each row so that all the leading entries become one.*
6. *From the bottom row onwards, add suitable multiples of each row such that all rows above the leading entries at pivot columns (Definition 16) are all zero.*

Steps $1-4$ *are known as Gaussian Elimination. Obtaining the RREF via Steps* $1-6$ *is known as Gauss-Jordan elimination.*

Exercise 5. *Obtain the RREF of the following augmented matrix*

$$\left[\begin{array}{ccccc:c} 0 & 0 & 2 & 4 & 2 & 8 \\ 1 & 2 & 4 & 5 & 3 & -9 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right] \tag{23}$$

via Gauss-Jordan Elimination (see Theorem 5).

Proof. Recall the notations for EROs (see Definition 10). We perform the following steps;

$$\left[\begin{array}{ccccc:c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right] \qquad R_1 \leftrightarrow R_2, \tag{24}$$

$$\left[\begin{array}{ccccc:c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 3 & 6 & 9 & -12 \end{array}\right] \qquad R_3 + 2 \cdot R_1, \tag{25}$$

$$\left[\begin{array}{ccccc:c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 0 & 0 & 6 & -24 \end{array}\right] \qquad R_3 - \frac{3}{2} \cdot R_2, \tag{26}$$

$$\left[\begin{array}{ccccc:c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 1 & 2 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right] \qquad \frac{1}{2}R_2, \qquad \frac{1}{6}R_3, \tag{27}$$

$$\left[\begin{array}{ccccc:c} 1 & 2 & 4 & 5 & 0 & 3 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right] \qquad R_2 - 1 \cdot R_3, \qquad R_1 - 3 \cdot R_3, \tag{28}$$

$$\left[\begin{array}{ccccc:c} 1 & 2 & 0 & -3 & 0 & -29 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right] \qquad R_1 - 4 \cdot R_2. \tag{29}$$

[1] □

[1] we thank reader Irena for the correction of errata in the Gaussian Elimination workings.

Result 2 (REF and their Interpretations for Solution Sets)**.** *Consider the REF $(A|b)$ augmented matrix form (see Definition 9). Note that every matrix has a unique RREF but can have many different REFs. If a linear system is not consistent (Definition 8), then the last column of the REF form of the augmented matrix is a pivot column. In other words, there will be a row representing an equation where $\sum_i^n 0x_i = c$, but $c \neq 0$. There is no solution to this linear system. A consistent linear system has a unique solution if except the last column b, every column is a pivot column. This system has as many variables in the linear system as the number of nonzero rows in the REF. If there exists a non-pivot column in the REF that is not the last one (b), then this consistent linear system has infinitely many solutions. This linear system has number of variables greater than the number of non-zero rows in the REF.*

Note that when solving for linear systems in which the contents are unknown constants, then we need to be careful about performing illegal row operations. That is, assume an augmented matrix

$$\left[\begin{array}{ccc:c} a & 1 & 0 & a \\ 1 & 1 & 1 & 1 \\ 0 & 1 & a & b \end{array}\right] \tag{30}$$

and in order to make the second row leading entry 0, we would perhaps like to perform $R_2 - \frac{1}{a}R_1$. However, we do not know that $a \neq 0$. In this case, we can consider either first swapping the first and second row and progressing, or perform a by-case method.

3.1.1.4 Homogeneous Linear Systems

Definition 18 (Homogeneous Linear Systems)**.** *A linear system (Definition 18) is homogeneous (HLS) if it has augmented matrix representation $(A|b)$ where $b = 0$ and all constants $a_{ij}, \in \mathbb{R}, \forall i \in [m], \forall j \in [n]$.*

See that the HLS is always satisfied by $x_i = 0, i \in [n]$ and we call this the trivial (sometimes, zero) solution. A non-trivial solution is any other solution that is not trivial.

Exercise 6. *See that in the xy-plane, the equations*

$$a_1x + b_1y = 0, \tag{31}$$

$$a_2x + b_2y = 0 \tag{32}$$

where a_1, b_1 not both zero and a_2, b_2 not both zero each represent straight lines through the origin, The system has only the trivial solution when the two equations are not the same line, otherwise they have infinitely many solutions. In the xyz-space, a system of two such linear equations passing through the origin always has infinitely many (non-trivial) solutions in addition to the trivial one, since they are either the same plane or intersect at a line passing through the origin at $(0, 0, 0)$.

Lemma 2. *A HLS (Definition 18) has either only the trivial solution or infinitely many solutions in addition to the trivial solution. A HLS with more unknowns than equations has infinitely many solutions.*

Proof. The first assertion is trivial since the zero solution satisfies it. The second assertion follows from considering the REF of the augmented matrix representation of a HLS with more unknowns than equations, then apply Result 2. □

Exercise 7. *For a HLS $Ax = 0$ (Definition 18) with non-zero solution, show that the system $Ax = b$ has either no solution or infinitely many solutions.*

Proof. By Theorem 2, a HLS system has no solution, one solution or infinite solutions. But suppose there is some solution u s.t. $Au = b$. Let v be non-zero solution for the HLS s.t. $Av = 0$, $v \neq 0$. Then $A(u+v) = Au + Av = b + 0 = b$, so $u + v$ is solution and $u + v \neq u$. But by Lemma 2, the solution space for $Ax = 0$ must have infinitely many vectors if such a v exists. It follows $Ax = b$ has infinitely many solutions if $\exists u$ s.t $Au = b$. □

3.1.2 Matrices

We formally defined augmented matrices in Definition 9. In the earlier theorems, we also referred to generalized matrix representations of numbers. We provide formal definition here.

Definition 19 (Matrix)**.** *A matrix is a rectangular array (or array of arrays) of numbers. The numbers are called entries. The size of a matrix is given by the rectangle's sides, and we say a matrix A is $m \times n$ for m rows and n column matrix. We can denote the entry at the i-th row and j-th coordinate by writing $A_{(i,j)} = a_{ij}$. This is often represented*

$$A = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ \cdots & \cdots\cdots & \cdots \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{bmatrix}, \tag{33}$$

and for brevity we also denote this $A = (a_{ij})_{m \times n}$, and sometimes we drop the size all together and write $A = (a_{ij})$.

Definition 20. *For brevity, given a matrix A (Definition 19) we refer to its size by using the notation $nrows(A)$ and $ncols(A)$ to indicate the number of rows in A and number of columns in A respectively. That is, A is a matrix size $nrows(A) \times ncols(A)$.*

Definition 21 (Column, Row Matrices/Vectors)**.** *A column matrix (vector) is a matrix with only a single column. A row matrix (vector) is a matrix with only one row.*

Definition 22 (Square Matrix)**.** *A square matrix is a matrix (Definition 19) that is square (number of rows is equivalent to the number of rows). We say $A_{n \times n}$ square matrix is of order n.*

Definition 23 (Diagonal Matrix)**.** *A square matrix A of order n (Definition 22) is diagonal matrix if all entries that are not along the diagonal are zero. That is,*

$$a_{ij} = 0 \qquad \text{when } i \neq j. \tag{34}$$

Definition 24 (Scalar Matrix)**.** *A diagonal matrix (Definition 23) is scalar matrix if all diagonal entries are the same, that is*

$$a_{ij} = \begin{cases} 0 & i \neq j \\ c & i = j, \end{cases} \tag{35}$$

for some constant $c \in \mathbb{R}$.

Definition 25 (Identity Matrix)**.** *Scalar matrix (Definition 24) is identity matrix if the diagonals are all one, that is $c = 1$. We often denote this as $\mathbb{1}$. If the size needs to be specified, we add subscript $\mathbb{1}_n$ to indicate order n.*

Definition 26 (Zero Matrix)**.** *Arbitrary matrix $m \times n$ is zero matrix if all entries are zero.*

Definition 27 (Symmetric Matrix)**.** *A square matrix A (Definition 22) is symmetric if $a_{ij} = a_{ji}$ for all $i, j \in [n]$.*

Definition 28 (Triangular Matrix)**.** *A square matrix A (Definition 22) is upper triangular if $a_{ij} = 0$ whenever $i > j$, and is lower triangular if $a_{ij} = 0$ whenever $i < j$.*

3.1.2.1 Operations on Matrices

Definition 29 (Matrix Addition, Subtraction and Scalars)**.** *The following are defined for operations on matrices:*

1. *Scalar Multiplication: $cA = (ca_{ij})$.*
2. *Matrix addition: $A + B = (a_{ij} + b_{ij})$.*
3. *Matrix subtraction: $A - B = (a_{ij} - b_{ij})$.* [2] *We denote $-A = -1 \cdot A$.*

Definition 30 (Matrix Equality)**.** *To show that two matrices A, B are equal, we have to show their their size is the same, and their entries $a_{ij} = b_{ij}$ $\forall i, \forall j$.*

Theorem 6 (Properties of Matrix Operators)**.** *Define matrices A, B, C of the same size and let $c, d \in \mathbb{R}$. Then the following properties hold:*

1. *Commutativity: $A + B = B + A$.*
2. *Associativity: $A + (B + C) = (A + B) + C$.*
3. *Linearity: $c(A + B) = cA + cB$.*
4. *Linearity: $(c + d)A = cA + dA$.*
5. *$c(dA) = (cd)A = d(cA)$.*
6. *$A + 0 = 0 + A = A$.*
7. *$A - A = 0$.*
8. *$0A = 0$.*

Proof. To show equality of matrices, we have to show their size is the same and that their corresponding entries match (see Definition 30). The proofs for the above theorems are rather trivial, and we show the associativity law (other proofs are of the same stripe). Proof of associativity: Let $A = (a_{ij}), B = (a_{ij}), C = (a_{ij})$, then

$$\begin{aligned} A + (B + C) &= (a_{ij}) + (B + C) && (36) \\ &= (a_{ij}) + (b_{ij} + c_{ij}) && (37) \\ &= (a_{ij} + b_{ij}) + (c_{ij}) && (38) \\ &= (A + B) + (c_{ij}) && (39) \\ &= (A + B) + C. && (40) \end{aligned}$$

That is, we rely on the associativity on addition of real numbers to prove the associativity on addition of matrices. Finally, see that their sizes match. □

[2] note that the matrix subtraction can be defined as the addition of a matrix A with a matrix B that has first been operated on a by scalar multiplication of $c = -1$.

Definition 31 (Matrix Multiplication)**.** *For matrices $A = (a_{ij})_{m\times p}, B = (b_{ij})_{p\times n}$, the matrix product AB is defined to be the $m \times n$ matrix s.t.*

$$C = A \times B = (c_{ij})_{m\times n} = \sum_{k=1}^{p} a_{ik} b_{kj}. \tag{41}$$

The matrix multiplication AB is only possible when $ncols(A) = nrows(B)$.

Exercise 8. *Show that matrix multiplication (Definition 31) is not commutative.*

Proof. Prove by counterexample. For matrices

$$A = \begin{pmatrix} -1 & 0 \\ 2 & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}, \tag{42}$$

see that

$$AB = \begin{pmatrix} -1 & -2 \\ 11 & 4 \end{pmatrix} \neq \begin{pmatrix} 3 & 6 \\ -3 & 0 \end{pmatrix} = BA. \tag{43}$$

□

Since the matrix multiplication is not commutative, when describing in words, we say that AB is the pre-multiplication of A to B and BA as the post-multiplication of A to B to prevent ambiguity.

Theorem 7 (Properties of Matrix Multiplication)**.** *Matrix multiplication (Definition 31) satisfies the following properties (we assume trivially that the size of the matrices are appropriate such that the matrix multiplication is legitimate) :*

1. *Associativity: $A(BC) = (AB)C$.*
2. *Distributivity: $A(B_1 + B_2) = AB_1 + AB_2$.*
3. *$c(AB) = (cA)B = A(cB)$.*
4. *$A0 = 0$, and $0A = 0$.*
5. *For identity matrix (Definition 25) of appropriate size ,$A\mathbb{1} = \mathbb{1}A = A$.*

Proof. Proof of the asserted statements follow directly form their definitions of matrices and matrix multiplications (Definitions 19, 31) and computing the resulting entries componentwise via the laws of algebra on real numbers (additionally, we also have to show that the sizes on the LHS and RHS are matching). □

Definition 32 (Powers of Square Matrices)**.** *For square matrix A and natural number $n \geq 0$, the power of A can be written*

$$A^n = \begin{cases} \mathbb{1} & \text{if } n = 0, \\ \underbrace{AA\cdots A}_{n \text{ number of times}} & \text{if } n \geq 1. \end{cases} \tag{44}$$

By associativity, $A^m A^n = A^{m+n}$. By non-commutativity $(AB)^n \neq A^n B^n$. See Theorem 7 for statements on properties of matrix multiplications.

Exercise 9. *Show that if $AB = BA$, then $(AB)^k = A^k B^k$.*

Proof. We proof by induction. Base case is when $k = 1$, so $(AB)^1 = AB = A^1B^1$. This statement is trivial. Now assume $(AB)^j = A^jB^j$ for $j < k$. Then $(AB)^{j+1} = (AB)^jAB = A^jB^jAB$. Define the swap operator $\psi : BA \to AB$, then apply $\psi^j(B^jA)$ to get AB^j. Then we have $A^j\psi^j(B^jA)B = A^jAB^jB = A^{j+1}b^{j+1}$ and by induction we are done. □

We may express rows, columns and even submatrices of a matrix by grouping together different entities. Here we show some examples.

Exercise 10 (Expressing Matrices as Block Matrices of Rows and Columns)**.** *For matrix* $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & 2 \end{pmatrix}$, *we may write*

$$A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \qquad B = \begin{pmatrix} b_1 & b_2 \end{pmatrix}, \tag{45}$$

$$a_1 = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}, \tag{46}$$

$$b_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \qquad b_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}. \tag{47}$$

See that the following relationships hold by direct computation

$$AB = \begin{pmatrix} Ab_1 & Ab_2 \end{pmatrix} = \begin{pmatrix} a_1B \\ a_2B \end{pmatrix}. \tag{48}$$

Exercise 11 (Block Matrix Operations)**.** *Let A be $m \times n$ matrix, and B_1, B_2 be $n \times p$, $n \times q$ matrices, C_1, C_2 be $r \times m$ matrices, and D_1, D_2 be $s \times m, t \times m$ matrices respectively. See which of the following block operations are valid:*

1. $A\begin{pmatrix} B_1 & B_2 \end{pmatrix} = \begin{pmatrix} AB_1 & AB_2 \end{pmatrix}$.
2. $\begin{pmatrix} C_1 & C_2 \end{pmatrix} A = \begin{pmatrix} C_1A & C_2A \end{pmatrix}$.
3. $\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} A = \begin{pmatrix} D_1A \\ D_2A \end{pmatrix}$.

Proof. Refer to Exercise 10 for operations on matrix blocks written as rows and columns.

1. If we write $B_1 = \begin{pmatrix} b_1 & \cdots & b_p \end{pmatrix}, B_2 = \begin{pmatrix} c_1 & \cdots & c_q \end{pmatrix}$. Then

$$A\begin{pmatrix} B_1 & B_2 \end{pmatrix} = \begin{pmatrix} Ab_1 & \cdots & Ab_p & Ac_1 & \cdots & Ac_q \end{pmatrix} \tag{49}$$

   and the relation is valid.

2. The matrix sizes do not permit a valid matrix operation.

3. If we let $D_1 = \begin{pmatrix} d_1 \\ \cdots \\ d_s \end{pmatrix}, D_2 = \begin{pmatrix} f_1 \\ \cdots \\ f_t \end{pmatrix}$, then

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \begin{pmatrix} d_1 \\ \cdots \\ d_s \\ f_1 \\ \cdots \\ f_t \end{pmatrix}. \tag{50}$$

Then we have

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} A = \begin{pmatrix} d_1 A \\ \cdots \\ d_s A \\ f_1 A \\ \cdots \\ f_t A \end{pmatrix} \tag{51}$$

and the relation is valid.

□

Recall the augmented matrix representation of linear systems (see Definition 9). We may define an equivalent form.

Definition 33 (Matrix Representation of Linear System)**.** *For system of linear equations*

$$\forall j \in [m], \qquad a_{j1}x_1 + a_{j2}x_2 + \cdots a_{jn}x_n = b_j, \tag{52}$$

we may represent the linear system by matrix multiplication

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix}}_{b}. \tag{53}$$

Then we say that A is the coefficient matrix, x is the variable matrix and that b is the constant matrix for the linear system specified. A solution to the linear system is a $n \times 1$ column matrix

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{pmatrix} \tag{54}$$

where $Au = b$. If we treat $A = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}$ where c_i represents the i-th column of A, then we may write

$$c_1x_1 + c_2x_2 + \cdots c_nx_n = \sum_{j=1}^{n} c_jx_j = b. \tag{55}$$

That is, the constant matrix is a linear combination of the columns of the coefficient matrix, where the weights are determined via the variable matrix.

Definition 34 (Matrix Transpose)**.** *For matrix $A = (a_{ij})_{m\times n}$, the matrix transpose of A is written $A' = (a'_{ij})_{n\times m}$ where the entry $a'_{ij} = a_{ji}$.*

See that the rows of A are the columns of A' and vice versa. See that a square matrix A is symmetric (Definition 27) *iff* $A = A'$.

Theorem 8 (Properties of the Matrix Transpose)**.** *The matrix transpose follows the following properties*

1. $(A')' = A$.
2. $(A+B)' = A' + B'$.
3. *For* $c \in \mathbb{R}$, $(cA)' = cA'$.
4. $(AB)' = B'A'$.

Proof. The proof of the first three parts are fairly straightforward by direct computation of the algebraic properties of real numbers that follow from their Definitions. We show the last assertion. Denote the sizes of matrix A to be $m \times n$ and that of B to be $n \times p$ so that the matrix multiplications (Definition 31) are defined. Then AB has size $m \times p$, so that its transpose has size $p \times m$. B' has size $p \times n$, A' has size $n \times m$, so $B'A'$ has size $p \times m$. We show they are componentwise equivalent. Since $(AB)_{ij} = \sum_l^n a_{il} b_{lj}$. Then $(AB)'_{ij} = \sum_l^n a_{jl} b_{li}$. On the other hand, we have $A'_{ij} = a_{ji}, B'_{ij} = b_{ji}$, so that $(B'A')_{ij} = \sum_l^n b'_{il} a'_{lj} = \sum_l^n b_{li} a_{jl}$. We have showed that the corresponding entries are the same. □

3.1.2.2 Invertibility of Matrices

Definition 35 (Invertibility of Square Matrix)**.** *Let A be square matrix of order n (Definition 22), then we say that A is invertible if $\exists$ square matrix B of order n s.t. $AB = \mathbb{1}_n = BA$. The matrix B is said to be the inverse of A. We denote this A^{-1}. There is no ambiguity; we shall see that the inverse of a matrix is unique (Theorem 9).*

Definition 36 (Singularity of Square Matrix)**.** *A matrix that does not have an inverse (Definition 35) is said to be singular.*

Exercise 12. *Show that the matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ is singular.*

Proof. Suppose not. Then let the inverse be $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then by Definition 35, we have

$$BA = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a+b & 0 \\ c+d & 0 \end{pmatrix}. \tag{56}$$

Then $1 = 0$. Contradiction. □

Theorem 9 (Uniqueness of Inverses)**.** *If B, C are inverses of square matrix A, then $B = C$.*

Proof. Write

$$AB = \mathbb{1} \implies CAB = C\mathbb{1} \implies \mathbb{1}B = C \implies B = C. \tag{57}$$

□

Exercise 13 (Conditions for Invertibility of Square Matrix Order Two)**.** *In the case for square matrix A of order two, denote*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{58}$$

State the conditions for invertibility and find the matrix inverse.

Proof. Define $B = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$, which is defined only if $ad - bc \neq 0$. We may compute directly the matrices $AB = BA = \mathbb{1}$ (we show how to explicitly compute matrix inverses such as B later on). □

Theorem 10 (Properties of Matrix Inverse)**.** *Let A, B be two invertible matrices (Definition 35), and $c \neq 0, \in \mathbb{R}$. Then the following properties hold*

1. *cA is invertible, in particular $(cA)^{-1} - \frac{1}{c}A^{-1}$.*
2. *A' is invertible, and $(A')^{-1} = (A^{-1})'$.*
3. *A^{-1} is invertible and $(A^{-1})^{-1} = A$.*
4. *AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. -

1. We can write

$$(cA)(\frac{1}{c}A^{-1}) = \left(c\frac{1}{c}\right) AA^{-1} = \mathbb{1}, \tag{59}$$

$$(\frac{1}{c}A^{-1})(cA) = (\frac{1}{c}c)A^{-1}A = \mathbb{1}, \tag{60}$$

   and the result immediately follows.

2. We show this by verifying that $(A^{-1})'$ is the inverse of A', which confirms the assertion that A' is invertible. In particular, by properties of matrix transpose (Theorem 8), write

$$A'(A^{-1})' = (A^{-1}A)' = \mathbb{1}' = \mathbb{1}, \tag{61}$$

$$(A^{-1})'A' = (AA^{-1})' = \mathbb{1}' = \mathbb{1}. \tag{62}$$

   Then A' is invertible, and the inverse is $(A^{-1})'$.

3. See that $A^{-1}A = \mathbb{1}, AA^{-1} = \mathbb{1}$ and by definition of inverse (Definition 35), the result follows.

4. Since A, B invertible, write

$$(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = A\mathbb{1}A^{-1} = AA^{-1} = \mathbb{1}. \tag{63}$$

   Also

$$(B^{-1}A^{-1})(AB) = \mathbb{1} \tag{64}$$

   by similar reasoning.

□

Definition 37 (Negative Powers of a Square Matrix)**.** **??** *For an invertible matrix A, we may define negative powers for a square matrix given $n \in \mathtt{Z}^+$ as the matrix power (Definition 32) of the inverse. That is,*

$$A^{-n} = (A^{-1})^n. \tag{65}$$

See that if A^n is invertible, then $(A^n)^{-1} = A^{-n}$ for any $n \in \mathtt{Z}$.

3.1.2.3 Elementary Matrices

One may notice that the elementary row operations (see Definition 10) may be considered as the pre-multiplication of some matrix to the matrix being operated on. For instance, see that

$$A = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \overset{2R_2}{\to} B = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 4 & -2 & 6 & 12 \\ 1 & 4 & 4 & 0 \end{pmatrix}, \tag{66}$$

and see that

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{E_1} A = B. \tag{67}$$

In particular, the ERO kR_i (Definition 10) may be performed by the pre-multiplication of matrix E_k, where E_k is a diagonal matrix (Definition 23) of order $nrows(A)$, where all the entries along the diagonal are one except for the i-th row, where the entry is k. If $k \neq 0$, and since performing kR_i, $\frac{1}{k}R_i$ in sequence gives us back the same matrix - see that the E_k is invertible and that E_k^{-1} is the diagonal matrix with all ones along the diagonal except for $\frac{1}{k}$ entry on the i-th row.

Next, observe the ERO $R_i \leftrightarrow R_j$ (see Definition 10) on the following instance:

$$A = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \overset{R_2 \leftrightarrow R_3}{\to} B = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 1 & 4 & 4 & 0 \\ 2 & -1 & 3 & 6 \end{pmatrix}, \tag{68}$$

and see that

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{E_2} A = B. \tag{69}$$

In particular, the ERO $R_i \leftrightarrow R_j$ (Definition 10) may be performed by the pre-multiplication of matrix E_s, where E_s is a matrix that began with an identity matrix (Definition 25) of order $nrows(A)$ and has gone through precisely the row swap $R_i \leftrightarrow R_j$. See that swapping rows i and j and then swapping again rows i and j gives us back the original matrix. Then $E_s = E_s^{-1}$.

Last but not least, observe the ERO $R_i + kR_j$ (see Definition 10) on the following instance:

$$A = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \overset{R_3 + 2R_1}{\to} B = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 1 & 4 & 4 & 0 \\ 3 & 4 & 8 & 6 \end{pmatrix}, \tag{70}$$

and see that

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}}_{E_3} A = B. \tag{71}$$

In particular, the ERO $R_i + kR_j$ (Definition 10) may be performed by the pre-multiplication of matrix E_l, where E_l is a matrix that began with an identity matrix (Definition 25) of order $nrows(A)$ and has gone through precisely the row addition $R_i + kR_j$. As before, the (triangular, Definition 28) matrix E_l is invertible and E_l^{-1} represents the row-swap operation $R_i - kR_j$.

Definition 38 (Elementary Matrix)**.** *A square matrix (Definition 22) that can be obtained from an identity matrix (Definition 25) from a single elementary row operation (Definition 38) is called an elementary matrix.*

We saw that all elementary matrices (Definition 38) are invertible, and their inverses are also elementary matrices. The discussions thus far allow us to arrive at the following result:

Lemma 3. *The EROs (Definition 10) performed on arbitrary matrices correspond precisely to the pre-multiplication of an elementary matrix (Definition 38) obtained from performing the ERO on the identity matrix (Definition 25).*

For a series of EROs applied in sequence $O_1, O_2, \cdots O_k$, (Definition 10) applied on A, s.t.

$$A \overset{O_1}{\to}\overset{O_2}{\to} \cdots \overset{O_k}{\to} B, \tag{72}$$

and their corresponding elementary matrices $E_1, \cdots, E_k$, see that the relation

$$E_k E_{k-1} \cdots E_1 A = B \tag{73}$$

must hold. By the invertibility, we have the relation

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} B. \tag{74}$$

Exercise 14. *Prove the solution-set equivalency asserted in Theorem 4.*

Proof. We show that if there are two row equivalent (Definition 11) augmented matrices (Definition 9) $(A|b), (C|d)$, then the linear systems $Ax = b, Cx = d$ share solution set. By Lemma 3, see that $\exists E$ s.t.

$$(C|d) = E(A|b) = (EA|Eb), \tag{75}$$

which is valid by Exercise 11. Then if $Au = b$ (that is if u is solution), then

$$Au = b \implies EAu = Eb \implies Cu = d. \tag{76}$$

On the other hand, if $Cv = d$, then

$$Cv = d \implies EAv = Eb \implies E^{-1}EAv = E^{-1}Eb \implies \mathbb{I}Av = \mathbb{I}b \implies Av = b. \tag{77}$$

They share solution set. □

Theorem 11 (Invertibility of Square Matrices, 1)**.** *If A is square matrix order n, then the following statements are equivalent:*

1. A is invertible.

2. $Ax = 0$ has only the trivial solution.

3. RREF of A is identity $\mathbb{1}$ matrix.

4. A can be expressed as $\Pi_i^n E_i$, where E_i are elementary matrices.

Proof. It turns out that this theorem shows an easy way to compute the inverses of an invertible matrix A. To show

(i) $1 \implies 2$: if $Ax = 0$, then

$$x = \mathbb{I}x = A^{-1}Ax = A^{-1}0 = 0, \tag{78}$$

where the last step follows from Theorem 7.

(ii) $2 \implies 3$: $Ax = 0$ is the only trivial solution. Since A is square, $nrows(A) = ncols(A)$, then by Lemma 2, the RREF of A of $(A|0)$ has no zero rows. By definition of RREF (Definition 17), the RREF of A is identity (Definition 25).

(iii) $3 \implies 4$: Since RREF of A is $\mathbb{1}$, by Lemma 3, $\exists E_i, i \in [k]$ s.t.

$$E_k E_{k-1} \cdots E_1 A = \mathbb{1}. \tag{79}$$

Then $A = (E_k \cdots E_1)^{-1}\mathbb{1}$, and by inverse properties, Theorem 10, we have

$$A = E_1^{-1} \cdots E_k^{-1}. \tag{80}$$

(iv) $4 \implies 1$: Since A is product of invertible elementary matrices, A is invertible by Theorem 10.

□

Theorem 12 (Cancellation Law)**.** *Let A be an invertible matrix (Definition 35) of order m, then the following properties hold:*

1. $AB_1 = AB_2 \implies B_1 = B_2$.

2. $C_1A = C_2A \implies C_1 = C_2$.

This does not hold for matrix A when it is non-singular.

Proof. For first the part,

$$AB_1 = AB_2 \implies AB_1 - AB_2 = 0 \implies A(B_1 - B_2) = 0. \tag{81}$$

Then since A is invertible, the HLS has only trivial solution by Theorem 11, so $B_1 - B_2 = 0$ and it follows that $B_1 = B_2$. For part 2, write

$$(C_1 - C_2)A = 0 \implies (C_1 - C_2)AA^{-1} = 0 \implies (C_1 - C_2)\mathbb{1} = 0, \tag{82}$$

and the result follows. □

We may use the discussions in Theorem 11 to compute the matrix inverse. For A satisfying $E_k \cdots E_1 A = \mathbb{1}$, see that $E_k \cdots E_1 = A^{-1}$ by the post multiplication of A^{-1} to both the RHS and LHS. Recall this is valid, since we are guaranteed the invertibility of A. Furthermore, this is unique (Theorem 9). Consider the $n \times 2n$ matrix $(A|\mathbb{1}_n)$. Then

$$\begin{aligned} E_k \cdots E_1(A|\mathbb{1}) &= (E_k \cdots E_1 A | E_k \cdots E_1 \mathbb{1}) && (83) \\ &= (\mathbb{1}|A^{-1}). && (84) \end{aligned}$$

That is, to the augmented matrix $(A|\mathbb{1})$, if we perform Gauss-Jordan elimination (see Theorem 5) and get RREF $\mathbb{1}$ on the LHS of $|$, the RHS is A^{-1}. Otherwise, A is singular and does not have an inverse. The following theorem shows us that given square matrices A, B - when we are to verify $A^{-1} = B$, we are only required to check one of $AB = \mathbb{1}$ or $BA = \mathbb{1}$.

Theorem 13. *Let A, B be square matrix order n. If $AB = \mathbb{1}$, then A, B are both invertible and*

$$A^{-1} = B, \qquad B^{-1} = A, \qquad BA = \mathbb{1}. \tag{85}$$

Proof. Consider HLS (Definition 18) $Bx = 0$. If $Bu = 0$, then

$$ABu = \mathbb{I}u \implies A0 = u \implies 0 = u. \tag{86}$$

Then $Bx = 0$ only has the trivial solution. By Theorem 11, B is invertible. Since B is invertible:

$$AB = \mathbb{1} \implies ABB^{-1} = \mathbb{1}B^{-1} \implies A\mathbb{1} = B^{-1} \implies A = B^{-1}. \tag{87}$$

So A is invertible by Theorem 11 and $A^{-1} = (B^{-1})^{-1} = B$, $BA = BB^{-1} = \mathbb{1}$. □

Exercise 15. *For square matrix A, given $A^2 - 3A - 6\mathbb{1} = 0$, show that A is invertible.*

Proof. Since we may write

$$A(A - 3\mathbb{1}) = A^2 - 3A\mathbb{1} = A^2 - 3A = 6\mathbb{1}, \tag{88}$$

then $A\left[\frac{1}{6}(A - 3\mathbb{1})\right] = \mathbb{1}$, and it follows that A is invertible from Theorem 13. □

Theorem 14 (Singularity of Matrix Products)**.** *Let A, B be two square matrices of order n. Then if A is singular, AB, BA are both singular (see Definition 14).*

Proof. Suppose not. Then AB is invertible, and let $C = (AB)^{-1}$. Then we may write

$$ABC = \mathbb{1}, \tag{89}$$

then A is invertible since $A^{-1} = BC$ by Theorem 13. This is contradiction. □

Theorem 15 (Elementary Column Operations)**.** *See from Lemma 3 that the pre-multiplication of an elementary matrix to matrix A is equivalent to doing an ERO on $A_{p\times m}$ matrix. Let E_k, E_s, E_l be elementary matrices corresponding to $kR_i, R_i \leftrightarrow R_j, R_i + kR_j$ respectively (see Definition 10). Then, the post multiplication of the matrices E_k, E_s, E_l correspond to*

1. *Multiplying the i-th column of A by k.*
2. *Swap columns i, j in A.*
3. *Add k times j-th column of A to i-th column of A*

respectively and let these be known collectively as elementary column operations (ECOs). They shall be denoted $kC_i, C_i \leftrightarrow C_j, C_i + kC_j$.

3.1.2.4 Matrix Determinants

It turns out that whether a square matrix is invertible (Definition 35) depends on a quantity of the matrix known as the determinant. We define this recursively.

Definition 39 (Determinants and Cofactors)**.** *For square matrix A order n, let M_{ij} be an $(n-1)\times(n-1)$ square matrix obtained from A by deleting the i-th and j-th column. Then the determinant of A is defined as*

$$det(A) = \begin{cases} a_{11} & \text{if } n = 1, \\ a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} & \text{if } n > 1, \end{cases} \tag{90}$$

where $A_{ij} = (-1)^{i+j} det(M_{ij})$. The number A_{ij} is known as the ij-cofactor of A. This method of recursively computing matrix determinants are known as cofactor expansion. Often, we adopt the equivalent notations for determinant of A:

$$det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{91}$$

Exercise 16 (Cofactor Expansion Examples)**.** *Here we show some instances of co-factor expansion. When the matrix is 2×2, then we have a general form*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{92}$$

Then see that the determinant by cofactor expansion

$$a \cdot (-1)^{1+1} det(d) + b \cdot (-1)^{1+2} det(c) = ad - bc. \tag{93}$$

Then for larger matrices, we may use these sub-results. For instance, the determinant for $B = \begin{pmatrix} -3 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 2 & 4 \end{pmatrix}$ via cofactor expansion is obtained

$$det(B) = (-3) \begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} - (-2) \begin{vmatrix} 4 & 1 \\ 0 & 4 \end{vmatrix} + 4 \begin{vmatrix} 4 & 3 \\ 0 & 2 \end{vmatrix} = -3(3 \cdot 4 - 1 \cdot 2) + 2(4 \cdot 4 - 1 \cdot 0) + 4(4 \cdot 2 - 3 \cdot 0) = 34. \tag{94}$$

Result 3 (Cofactor Expansion Invariance)**.** *For square matrix A order n, $det(A)$ (Definition 39) may be found via cofactor expansion along any row or any column.*

Theorem 16 (Cofactor Expansion of Triangular Matrices)**.** *For triangular matrix A, the determinant A is equal to the product of diagonal entries of A.*

Proof. By definition of triangular matrices (Definition 28), both the upper triangular and lower triangular has a row that is all zeros except for possibly a singly entry (the diagonal itself). That is, an upper triangular takes general form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \tag{95}$$

Additionally, since matrix is square, cofactor expansion along the last row, last entry has the term $(-1)^{i+i} = 1$. By Result 3, see that if we apply recursively the cofactor expansion along the last row, we obtain just the product of the diagonal entries. A similar reasoning is applied if the matrix is lower triangular. □

See that the determinant of $\mathbb{1}$ is one by Theorem 16.

Theorem 17 (Determinant of Matrix Transpose)**.** *For square matrix A of order n, $det(A) = det(A')$.*

Proof. We prove by induction. The base case is for a matrix containing a single scalar value. This is trivially true, since the transpose of a matrix 1×1 is itself. Next, assume $det(A) = det(A')$ for any square matrix A order k. We show this holds for $(k+1) \times (k+1)$ matrix. In particular, by cofactor expansion along the first row of A, obtain

$$det(A) = \sum_{i}^{n} (-1)^{1+i} a_{1i} det(M_{1i}). \tag{96}$$

Next perform, cofactor expansion along the first column of A', then

$$det(A') = \sum_{i}^{n} (-1)^{1+i} a_{1i} det(M'_{1i}). \tag{97}$$

By induction, $det(A) = det(A')$ since $det(M_{ij}) = det(M'_{ij})$. □

Theorem 18 (Determinant of Repeated Row/Column Matrix)**.** *The determinant of a square matrix with two identical rows is zero. The determinant of a square matrix with two identical columns is zero.*

Proof. We prove by induction. The base case is for matrix A size 2×2. For matrix $A = \begin{pmatrix} a & b \\ a & b \end{pmatrix}$, by Exercise 16 we have $det(A) = ab - ab = 0$. Assume that for $k < n$, $det(A)$ size $k \times k$ with repeated row is zero. Then consider a $(k+1) \times (k+1)$ matrix with row i equivalent to row j, $i \neq j$. Then by cofactor expansion along some row m that is neither i nor j, we have

$$det(A) = a_{m1} A_{m1} + \cdots + a_{m,k+1} A_{m,k+1} \tag{98}$$

A_{mr} is the cofactor $(-1)^{m+r} det(M_{mr})$, which has identical rows and by inductive assumption has determinant zero. Then $det(A) = 0$ and we are done. Since $det(A) = det(A')$, a square matrix with two identical columns has transpose with two identical rows and the result follows. □

Theorem 19. *Recall the notations for EROs (Definition 10) and correspondence to their elementary matrices (Lemma 3). Let A be square matrix, and*

(i) B be a square matrix obtained by some ERO kR_i. Then, $det(B) = kdet(A)$.

(ii) B be a square matrix obtained by some ERO $R_i \leftrightarrow R_j$. Then, $det(B) = -det(A)$.

(iii) B be a square matrix obtained by some ERO $R_i + kR_j$. Then, $det(B) = det(A)$.

(iv) E be some elementary matrix with size $nrows(A) \times nrows(A)$. Then $det(EA) = det(E)det(A)$.

It turns out that this is quite useful because the determinants of elementary matrices are fairly easy to compute. Only the elementary matrix corresponding to the swap operation is a non-triangular matrix (Definition 28), but even the swap operation has corresponding elementary matrix where each sub-square matrix has row/column with only a single scalar entry of one and the rest zero.

Proof. We do not prove this theorem but this may be obtained via the rather mechanical cofactor expansion and definition of matrix determinants (Definition 39). □

Theorem 20. *Recall the notations for CROs (Definition 15) and correspondence to their elementary matrices. Let A be square matrix, and*

(i) B be a square matrix obtained by some CRO kC_i. Then, $det(B) = kdet(A)$.

(ii) B be a square matrix obtained by some CRO $C_i \leftrightarrow C_j$. Then, $det(B) = -det(A)$.

(iii) B be a square matrix obtained by some CRO $C_i + kC_j$. Then, $det(B) = det(A)$.

(iv) E be some elementary matrix with size $nrows(A) \times nrows(A)$. Then $det(AE) = det(E)det(A)$.

Theorem 21 (Determinants and Invertibility)**.** *Square matrix A is invertible iff $det(A) \neq 0$.*

Proof. For square matrix A we may write $B = E_k \cdots E_1 A$, where each E_i is elementary matrix and B is RREF. By Theorem 19, $det(B) = det(A)\Pi_{i=1}^{k} det(E_i)$. By Theorem 11, $B = \mathbb{1}$, and $det(B) = 1$. Then $det(A) \neq 0$ since $\nexists i$ s.t. $det(E_i) = 0$. If A is singular, then B has zero row (Definition 13). By cofactor expansion (Theorem 3) along the zero row, $det(B) = 0$, then $det(A) = 0$ since again, $\nexists i$ s.t. $det(E_i) = 0$. □

Theorem 22. *For square matrix A, B order n and $c \in \mathbb{R}$, the following hold:*

1. $det(cA) = c^n det(A)$,

2. $det(AB) = det(A)det(B)$,

3. If A is invertible, then $det(A^{-1}) = \frac{1}{det(A)}$.

Proof. -

1. This follows from Theorem 19 and seeing that cA is multiplying each of the n rows by c.

2. If A is singular, then AB is singular by Theorem 14. Then $det(AB) = det(A)det(B) = 0$. Otherwise, matrix A may be represented by product of elementary matrices s.t.

$$det(AB) = det(E_1 \cdots E_k B) = det(B)\Pi_{i=1}^{k} det(E_i) = det(B)det(A). \tag{99}$$

3. Follows since $det(A)det(A^{-1}) = det(AA^{-1}) = det(\mathbb{1}) = 1$. The first equality follows from part 2.

□

Definition 40 (Classical Adjoint)**.** *Let A be square matrix order n. Then the (classical) adjoint of A is $n \times n$ matrix*

$$adj(A) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}', \tag{100}$$

where A_{ij} is (i,j) cofactor of A (Definition 39).

Theorem 23 (Inverse by Adjoint). *Let A be square matrix, then if A is invertible, we have*

$$A^{-1} = \frac{1}{det(A)} adj(A). \tag{101}$$

Proof. Let $B = A \cdot adj(A)$, then

$$\begin{aligned} b_{ij} &= a_{i1}A'_{1j} + a_{i2}A'_{2j} + \cdots a_{in}A'_{nj} \quad (102) \\ &= a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots a_{in}A_{jn}. \quad (103) \end{aligned}$$

By definition of cofactor expansion (see Definition 39 and Theorem 3), see that

$$det(A) = b_{ii}. \tag{104}$$

By Equation 103, see that when $i \neq j$, then b_{ij} is the cofactor expansion along the row j of matrix A where the entries of row i, j are both $a_{i1}, a_{i2}, \cdots a_{in}$. Then by Theorem 18, $b_{ij} = 0$ if $i \neq j$. Then

$$A \cdot adj(A) = det(A)\mathbb{1} \implies \frac{1}{det(A)} A \cdot adj(A) = \mathbb{1}. \tag{105}$$

Then the result follows. □

Theorem 24 (Cramer's Rule). *Suppose $Ax = b$ is linear system (Definition 5), where A is square matrix order n. Then if A_i is the matrix obtained from replacing i-th column of A by b, and if A is invertible, then the system has unique solution*

$$x = \frac{1}{det(A)} \begin{pmatrix} det(A_1) \\ det(A_2) \\ \cdots \\ det(A_n) \end{pmatrix}. \tag{106}$$

Since

$$Ax = b \leftrightarrow x = A^{-1}b = \frac{1}{det(A)} adj(A) \cdot b, \tag{107}$$

then

$$x_i = \frac{b_1A_{1i} + b_2A_{2i} + \cdots + b_nA_{ni}}{det(A)} = \frac{det(A_i)}{det(A)}. \tag{108}$$

Exercise 17. *For $A_{m\times n}, B_{n\times p}$ matrices, if $Bx = 0$ has infinitely many solutions, how many solutions does $ABx = 0$ have? What about if $Bx = 0$ has only the trivial solution?*

Proof. Suppose $Bx = 0$ has infinitely many solutions, and let this solution space be S. See that $\forall s \in S$, $ABs = A0 = 0$. There are at least as many solutions as Bx, and this is in fact infinitely many. On other hand, we cannot make comments about the solutions to $ABx = 0$ when $Bx = 0$ only has trivial solution. For instance, if $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the cases for matrix $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ give rise to a linear system with trivial solution and infinitely many solutions respectively. □

Definition 41 (Trace). *For square matrix A of order n, the matrix trace denoted $tr(A)$ is the sum of entries along the diagonals of A. For A, B square matrix both of order n, $C_{m\times n}, D_{n\times m}$, we have*

1. *that*

$$tr(A + B) = tr(A) + tr(B). \tag{109}$$

2. *that* $tr(cA) = ctr(A)$.

3. *that* $tr(CD) = tr(DC)$.

4. *that* $\nexists A, B$ *s.t.* $AB - BA = \mathbb{1}$.

Proof. The first two properties are easy to proof by definitions of trace and matrix. For the third statement, see that

$$(CD)_{ii} = \sum_j^n c_{ij}d_{ji}, \tag{110}$$

$$tr(CD) = \sum_i^m \sum_j^n c_{ij}d_{ji} \tag{111}$$

$$= \sum_j^n \sum_i^m d_{ji}c_{ij}. \tag{112}$$

See that the RHS is precisely $tr(DC)$. Lastly, since $tr(AB - BA) = tr(AB) - tr(BA) = tr(AB) - tr(AB) = 0$ by the earlier parts and $tr(\mathbb{1}_n) = n$, it cannot be that $AB - BA = \mathbb{1}$. □

Exercise 18 (Orthogonal Matrices)**.** *A square matrix is an orthogonal matrix if*

$$AA' = \mathbb{1} = A'A. \tag{113}$$

Suppose A, B *is square matrix order* n *and orthogonal, then show* AB *is orthogonal.*

Proof. See that (by Theorem 6)

$$AB(AB)' = ABB'A' = A\mathbb{1}A' = AA' = \mathbb{1}, \tag{114}$$

and that

$$(AB)'AB = B'A'AB = B'\mathbb{1}B = B'B = \mathbb{1}. \tag{115}$$

□

Orthogonal matrices are treated in Section 3.1.5.3.

Exercise 19 (Nilponent Matrices)**.** *A square matrix is a nilpotent matrix if* $\exists k \in \mathbb{Z}^+$ *s.t.* $A^k = 0$. *Let* A, B *be square matrices order* n*, and that* $AB = BA$ *with nilpotent matrix* A*. Show that* AB *is nilpotent. Show that we require the condition* $AB \neq BA$.

Proof. Let k be some constant s.t. $A^k = 0$. Then by Exercise 9 we have

$$(AB)^k = A^k B^k \implies 0B^k = 0, \tag{116}$$

so AB is nilpotent. No - we may prove by simple counterexample, say $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. □

Exercise 20. *Show that for diagonal matrix* A*, the power of the diagonal matrix* A^k *is diagonal matrix with entry* a_{ii}^k*, for* $i \in [nrows(A)]$.

Proof. Obtain this by simply writing out the mathematical induction proof. □

Exercise 21. *Prove or disprove the following:*

1. *If A, B diagonal matrices of same size, $BA = BA$.*
2. *If A is square matrix, and $A^2 = 0$, then $A = 0$.*
3. *If A is matrix s.t. $AA' = 0$, $A = 0$.*
4. *A, B invertible $\implies$ $A + B$ invertible.*
5. *A, B singular $\implies$ $A + B$ singular.*

Proof. -

1. This statement is true. See that $AB_{ij} = a_{ii}b_{ii}$ and $BA_{ij} = b_{ii}a_{ii}$.
2. This statement is false by counterexample $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.
3. This statement is true. For matrix A size $m \times n$, AA' is square matrix $m \times m$. $AA'_{ii} = \sum_j^n a_{ij}a'_{ji} = \sum_j^n a_{ij}^2$ and this implies that if $AA' = 0$, $a_{ij} = 0$ for all values i, j. A must be zero matrix.
4. This statement is false by counterexample:
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{117}$$
5. This statement is false by counterexample:
$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{118}$$

□

Exercise 22. *Let A be square matrix. Then*

1. *Show that if $A^2 = 0$, then $\mathbb{1} - A$ is invertible. Find the inverse.*
2. *Show that if $A^3 = 0$, then $\mathbb{1} - A$ is invertible. Find the inverse.*
3. *Find the relation at higher order powers.*

Proof. -

1. Since
$$(\mathbb{1} - A)(\mathbb{1} + A) = \mathbb{1} - A^2 = \mathbb{1}, \tag{119}$$
then $\mathbb{1} - A$ is invertible with inverse $\mathbb{1} + A$.
2. See that
$$(\mathbb{1} - A)(\mathbb{1} + A + A^2) = \mathbb{1} - A^3 = \mathbb{1}, \tag{120}$$
so the inverse of $\mathbb{1} - A$ is $\mathbb{1} + A + A^2$.
3. As in previous parts, the general form matrix inverse of $\mathbb{1} - A$ where $A^n = 0$ is
$$\sum_{j=0}^{n-1} A^j. \tag{121}$$

□

Exercise 23. *Suppose A, B is invertible square matrix order n, and that $A+B$ is invertible. Then show that $A^{-1}+B^{-1}$ is invertible and find $(A+B)^{-1}$.*

Proof. If $A+B$ is invertible, then the matrix $(A(A+B)^{-1}B)$ must be invertible. Consider the inverse of this matrix, by Theorem 10 we have

$$(A(A+B)^{-1}B)^{-1} = B^{-1}(A+B)A^{-1} = (B^{-1}A + \mathbb{1})A^{-1} = B^{-1} + A^{-1}. \tag{122}$$

We have effectively shown that the inverse of $A^{-1}+B^{-1}$ exists and is $(A(A+B)^{-1}B)$. Then we may write

$$\begin{aligned} A(A+B)^{-1}B &= (A^{-1}+B^{-1})^{-1} & (123)\\ A^{-1}A(A+B)^{-1}BB^{-1} &= A^{-1}(A^{-1}+B^{-1})^{-1}B^{-1} = (A+B)^{-1} & (124) \end{aligned}$$

and we are done. □

Exercise 24. *Let A, P, D be square matrices s.t.*

$$A = PDP^{-1}. \tag{125}$$

Show that $A^k = PD^kP^{-1}$ for all $k \in \mathtt{Z}^+$.

Proof. See that $A^k = PDP^{-1}\underbrace{PDP^{-1}\cdots PDP^{-1}}_{\text{k times}}$. Then all the adjacent $P^{-1}P$ is identity and we arrive at PD^kP^{-1}. □

Exercise 25. *Show that for matrix $A_{m\times n}, B_{n\times m}$, and $A \overset{\mathcal{R}}{\equiv} REF(A)$ with $REF(A)$ having some zero row, show that AB is singular.*

Proof. If $A \overset{\mathcal{R}}{\equiv} REF(A)$ with $REF(A)$ having a zero row, then $A = E_k \cdots E_1 REF(A)$ for elementary matrices $E_i, i \in [k]$, and $AB = E_k \cdots E_1 REF(A)B$. It follows that $AB \overset{\mathcal{R}}{\equiv} REF(A)B$ and since $REF(A)$ has zero row, by the block matrix multiplication (Exercise 11) AB has $REF(AB)$ where $REF(AB)$ has zero row. This can never be row equivalent to $\mathbb{1}$, and by Theorem 11, AB is singular. □

Exercise 26. *For matrix $A_{m\times n}$ and $m > n$, see if is possible for AB to be invertible where B is matrix size $n \times m$.*

Proof. AB will always be singular. The REF of A has at most n non-zero rows, and since $m > n$, REF form of A has zero row. Then by the proof in Exercise 25, AB must be singular. □

Exercise 27. *Let A be some 2×2 orthogonal matrix (Definition 18). Prove that*

1. *$det(A) = \pm 1$,*
2. *$A = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}$ for some $\theta \in \mathbb{R}$ if $det(A) = 1$,*
3. *and otherwise $A = \begin{pmatrix} cos(\theta) & sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}$.*

Proof. -

1. $det(\mathbb{1}) = det(AA') = det(A)det(A') = det(A)^2 = 1$.

2. For matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, if A is orthogonal, $A^{-1} = A'$. Then using invertibility by adjoint (Theorem 23), we can write

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \frac{1}{det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{126}$$

So $a = d, b = -c$ and by assumption $a^2 + c^2 = ad - bc$. Let $a = cos(\theta), c = sin(\theta)$, the result follows.

3. Follow part 2. with $a \to -d, b \to c$.

□

Exercise 28. *Let A be invertible square matrix order n. Then*

1. Show that $adj(A)$ is invertible.

2. Find $det(adj(A)), adj(A)^{-1}$.

3. Show $det(A) = 1 \implies adj(adj(A)) = A$.

Proof. -

1. By Theorem 23, we have

$$A \left[\frac{1}{det(A)} adj(A) \right] = \mathbb{1} \implies \left[\frac{1}{det(A)} A \right] adj(A) = \mathbb{1} \tag{127}$$

by Theorem 13.

2. By Theorem 22, since

$$det(\mathbb{1}) = \left(\frac{1}{det(A)} \right)^n det(adj(A)) \det(A) = 1, \tag{128}$$

then $det(adj(A)) = det(A)^{n-1}$ and $adj(A)^{-1} = \frac{1}{det(A)} A$.

3. From the general form $A \left[\frac{1}{det(A)} adj(A) \right] = \mathbb{1}$, we can write

$$adj(A) \left[\frac{1}{det(adj(A))} adj(adj(A)) \right] = \mathbb{1}. \tag{129}$$

Then by part 2, we have

$$adj(adj(A)) = det(adj(A))adj(A)^{-1} = det(adj(A)) \frac{1}{det(A)} A = det(A)^{n-1} det(A)^{-1} A = det(A)^{n-2} A.$$

If $det(A) = 1$, then it follows that

$$adj(adj(A)) = A. \tag{130}$$

□

Exercise 29. *Prove or disprove the following statements.*

1. *A, B square matrices of order n satisfies $det(A + B) = det(A) + det(B)$.*

2. *If A is square matrix, $det(A + \mathbb{1}) = det(A' + \mathbb{1})$.*

3. *A, B square matrices of order n and $A = PBP^{-1}$ for some invertible P satisfies $det(A) = det(B)$.*

4. *A, B, C square matrices of order and $det(A) = det(B)$ satisfies $det(A + C) = det(B + C)$.*

Proof. -

1. This is false by counterexample:

$$A = \mathbb{1}_2, \qquad B = -\mathbb{1}_2. \tag{131}$$

2. This is true, since $det(A + \mathbb{1}) = det((A + \mathbb{1})') = det(A' + \mathbb{1})$.

3. This is true, since

$$det(A) = det(PBP^{-1}) = det(P)det(B)det(P^{-1}) = det(B)det(P)det(P^{-1}) = det(B) \cdot 1. \tag{132}$$

4. This is false by counterexample:

$$A = -\mathbb{1}_2, \qquad B = \mathbb{1}_2, \qquad C = \mathbb{1}_2. \tag{133}$$

□

3.1.3 Vector Spaces

3.1.3.1 Finite Euclidean Spaces

A vector may be specified by the direction of the arrow, and its length specified by its magnitude. Two vectors are equal if the share direction and magnitude. If we denote a length of the vector u by $\|u\|$, then clearly the length of a scaled vector cu must be $c\|u\|$. The geometrical interpretations for vectors are somewhat elusive past three dimensional spaces, however, it should be noted that the theorems constructed in spaces of dimensions lower than three may be extended to higher finite dimensions, even if it may not be visualized.

Definition 42 (Vector and Coordinates)**.** *A n-vector or ordered n-tuple of real numbers takes form*

$$(u_1, u_2, \cdots, u_n) \tag{134}$$

where $u_i \in \mathbb{R}, i \in [n]$. The i-th component or coordinate of a vector is the entry u_i.

Definition 43 (Vector Terminologies)**.** *Two n-vectors u, v are equal if $\forall i \in [n]$, $u_i = v_i$. The vector $w = u + v$ is s.t $\forall i \in [n], w_i = u_i + v_i$. Scalar multiple of vector is the operation for some $c \in \mathbb{R}, w = cu$ s.t. $\forall i \in [n], w_i = cu_i$. The negative of vector u is the scalar multiple of vector where $c = -1$. The subtraction of vector v from u is the addition of vector u to negative of vector v. A zero vector is one in which $\forall i \in [n], u_i = 0$.*

See that we may identify vectors as special cases of matrices, that is either the row vector or column vector (Definition 21).

Theorem 25 (Vector Operations)**.** *For n-vector u, v, w, the following hold:*

1. $u + v = v + u$,
2. $u + (v + w) = (u + v) + w$,
3. $u + 0 = u = 0 + u$,
4. $u + (-u) = 0$,
5. $c(du) = (cd)u$,
6. $c(u + v) = cu + cv$,
7. $(c + d)u = cu + du$,
8. $1u = u$.

Proof. These properties follow from their definitions. Otherwise, see that vectors are matrices, and use the same result on matrices (i.e. Theorem 7, Definition 29 and Definition 31). □

We give formal definitions for Euclidean spaces.

Definition 44 (Euclidean Space)**.** *A Euclidean space is the set of all n-vectors of real numbers. This is denoted $\mathbb{R}^n$. When $n = 1$, we usually just write $\mathbb{R}$. For any element $u \in \mathbb{R}^n$, u is n-vector.*

See that the solution set of a linear system (Definition 5) must be a subset of the Euclidean space.

Exercise 30 (Expressions for Geometric Objects in the Euclidean Space)**.** *We show implicit and explicit expressions for objects in low dimensional spaces.*

1. *See that a line in $\mathbb{R}^2$ may be represented (implicitly) by the set notation*

$$\{(x, y)|ax + by = c\}, \tag{135}$$

where $a, b, c \in \mathbb{R}$, and it is not the case that both a, b are zero. This may (explicitly) also be written as

$$\left\{\left(\frac{c - bt}{a}, t\right) |t \in \mathbb{R}\right\} \qquad \text{if } a \neq 0, \text{ or equivalently} \tag{136}$$

$$\left\{\left(t, \frac{c - at}{b}\right) |t \in \mathbb{R}\right\} \qquad \text{if } b \neq 0. \tag{137}$$

2. *A plane in $\mathbb{R}^3$ may be expressed*

$$\{(x, y, z)|ax + by + cz = d\} \tag{138}$$

where $a, b, c \in \mathbb{R}$ not all zero and $d \in \mathbb{R}$. We may also write explicitly as any of the equivalent forms

$$\left\{\left(\frac{d - bs - ct}{a}, s, t\right) |s, t \in \mathbb{R}\right\} \qquad a \neq 0, \tag{139}$$

$$\left\{\left(s, \frac{d - as - ct}{b}, t\right) |s, t \in \mathbb{R}\right\} \qquad b \neq 0, \tag{140}$$

$$\left\{\left(s, t, \frac{d - as - bt}{c}\right) |s, t \in \mathbb{R}\right\} \qquad c \neq 0. \tag{141}$$

3. A line in R^3 may be represented by the explicit set notation

$$\{(a_0 + at, b_0 + bt, c_0 + ct | t \in \mathbb{R}\} = \{(a_0, b_0, c_0) + t(a, b, c) | t \in \mathbb{R}\}, \tag{142}$$

where $a, b, c, a_0, b_0, c_0 \in \mathbb{R}$, and not all a, b, c are zero.

Definition 45 (Set Cardinality)**.** *For finite set S, the number of elements in the set (cardinality) is denoted $|S|$.*

3.1.3.2 Linear Spans

Definition 46 (Linear Combination)**.** *Let $u_i, i \in [k]$ be vectors in $\mathbb{R}^n$, then $\forall c_i \in \mathbb{R}, i \in [k]$, the vector*

$$\sum_i^k c_i u_i \tag{143}$$

is said to be linear combination of the vectors $u_i, i \in [k]$.

Definition 47 (e_i)**.** *Denote vectors $e_i \in \mathbb{R}^n$, as the vectors with 1 in the i-th entry and zero everywhere else. That is*

$$e_i = (0 \cdots 0 \underbrace{1}_{i\text{-}th} 0 \cdots 0). \tag{144}$$

See that for $u \in \mathbb{R}^n$, we can write $u = \sum_i^n u_i e_i$.

Definition 48 (Linear Span)**.** *Let $S = \{u_i, i \in [k]\}$ be set of vectors in $\mathbb{R}^n$, then the set of all linear combinations of $u_i, i \in [k]$, that is*

$$\left\{\sum_i^k c_i u_i \mid \forall i \in [k], c_i \in \mathbb{R}\right\} \tag{145}$$

is called the linear span of set S and is denoted as $span(S)$ or $span\{u_1, \cdots u_k\}$.

See that we may express spans in different ways. For instance, a set $V = \{(2a+b, a, 3b-a) \mid a, b \in \mathbb{R}\}$ can be written as $span\{(2, 1, -1), (1, 0, 3)\}$.

Exercise 31. *Show that*

$$V = span\{(1, 0, 1), (1, 1, 0), (0, 1, 1)\} = \mathbb{R}^3. \tag{146}$$

Proof. $V = \mathbb{R}^3$ if we may write arbitrary vector (x, y, z) as a linear combination of elements in the spanning set of V (we formally define this later, but treat this for now to be the three vectors given). That is, $\exists a, b, c$ s.t.

$$a(1, 0, 1) + b(1, 1, 0) + c(0, 1, 1) = (x, y, z), \tag{147}$$

and this corresponds to augmented matrix system

$$\left[\begin{array}{ccc:c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 1 & 0 & 1 & z \end{array}\right] \overset{\text{GE (Def. 5)}}{\rightarrow} \left[\begin{array}{ccc:c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 0 & 0 & 2 & z - x + y \end{array}\right]. \tag{148}$$

This system is consistent regardless of the values of x, y, z. On the other hand, supposed we performed Gaussian Elimination and obtain zero row on the LHS, that is the coefficient matrix. Then, it is possible for the last column to be a pivot column and for the system to be inconsistent (Result 2). ☐

We may generalize Exercise 31 to a more general question of whether a set of vectors span the entire Euclidean space $\mathbb{R}^n$.

Corollary 2. *For set $S = \{u_i, i \in [k]\} \in \mathbb{R}^n$, S spans $\mathbb{R}^n$ iff for arbitrary vector $v \in \mathbb{R}^n$, the linear system represented by the augmented matrix (Definition 9) is consistent, where $(A|v)$ and A is coefficient matrix created from horizontally stacking the column vectors $u_i, i \in [k]$. This is immediately made obvious if we consider the discussion inside the matrix representation for linear systems in Definition 33. By Theorem 2, if $REF(A)$ has no zero row, then the linear system is always consistent. Otherwise, the system is not always consistent and $span(S) \neq \mathbb{R}^n$.*

Theorem 26 (Cardinality of a Set and Its Spanning Limitations)**.** *For set $S = \{u_i, i \in [k]\}$ be set of vectors in $\mathbb{R}^n$, if $k < n$, then S cannot span $\mathbb{R}^n$.*

Proof. Since the coefficient matrix obtained from stacking k columns is size $n \times k$, then the result follows directly from Theorem 26. □

Theorem 27 (Zero Vector and Span Closure)**.** *Let $S = \{u_i, i \in [k]\} \subseteq \mathbb{R}^n$. Then,*

1. *$0 \in span(S)$.*
2. *For any $v_i \in span(S)$ and $c_i \in \mathbb{R}, i \in [r]$, $\sum_i^r c_i v_i \in span(S)$.*

Proof. -

1. See that $0 = \sum_i 0u_i \in span(S)$.
2. For each $v \in span(S)$, they are linear combination of $u_i, i \in [k]$. Then we may express

$$\begin{aligned} v_1 &= a_{11}u_1 + \cdots + a_{1k}u_k, && (149)\\ v_2 &= a_{21}u_1 + \cdots + a_{2k}u_k, && (150)\\ \cdots & && (151)\\ v_r &= a_{r1}u_1 + \cdots + a_{rk}u_k, && (152)\\ & && (153) \end{aligned}$$

so that for linear combination

$$\begin{aligned} c_1v_1 + \cdots + c_rv_r &= (c_1a_{11} + c_2a_{21} + \cdots + c_ra_{r1})u_1 && (154)\\ &\quad +(c_1a_{12} + c_2a_{22} + \cdots + c_ra_{r2})u_2 && (155)\\ &\quad +\cdots && (156)\\ &\quad +(c_1a_{1k} + c_2a_{2k} + \cdots + c_ra_{rk})u_k. && (157) \end{aligned}$$

See this is in $span(S)$.

□

Theorem 28 (Spanning Set of a Set Span)**.** *For $S_1 = \{u_i, i \in [k]\}, S_2 = \{v_j, j \in [m]\} \subseteq \mathbb{R}^n$, $span(S_1) \subseteq span(S_2)$ iff for all $i \in [k]$, u_i is a linear combination of $v_j, j \in [m]$.*

Proof. $\rightarrow$: Assume $span(S_1) \subseteq span(S_2)$, then since $S_1 \subseteq span(S_1) \subseteq span(S_2)$, each u_i is linear combination of v's.
$\leftarrow$: Assume $\forall i \in [k], u_i$ is linear combination of v's. Then, $u_i \in span(S_2), \forall i \in [k]$. By Theorem 27, any w that is linear combination of these u's can rewritten as linear combination of the v's, which is itself in $span(S_2)$. Then we are done. □

Exercise 32. *Discuss how one may approach to see if for some set S_1, S_2, whether $span(S_1) \subseteq span(S_2)$.*

Proof. Let the vectors in S_1 be denoted $u_i, i \in [n]$ and in S_2 be denoted $v_j, j \in [m]$. Then in order to see if each u_i may be represented as a linear combination of the v_j's, we may simultaneously solve for multiple linear systems. These linear systems may be represented by an augmented matrix $(V|u_1|u_2 \cdots |u_k)$, and by Gaussian Elimination we are able to check if any of the systems $(V|u_i), i \in [n]$ are not consistent. V here is obtained by horizontally stacking the column vectors for v_i. This follows from the discussion made in Definition 33 on constant matrix as linear combinations of the columns in the coefficient matrix. ☐

Theorem 29 (Redundant Vectors)**.** *Let $S = \{u_i, i \in [k]\} \subseteq \mathbb{R}^n$, and if $\exists j \in [k]$ s.t. u_j is linear combination of vectors in $S \backslash u_j$, then $span(S) = span(S \backslash u_j)$.*

Proof. The proof follows directly from applying Theorem 28. ☐

Let u, v be two nonzero vectors. Then $span\{u, v\} = su + tv$, $\forall s, t \in \mathbb{R}$. If it is not the case that $u//v$, then $span\{u, v\}$ is a plane containing origin. In $\mathbb{R}^2$ space, the span is just the entire space. In $\mathbb{R}^3$, the span can be written

$$span\{u, v\} = \{su + tv | s, t, \in \mathbb{R}\} = \{(x, y, z) \mid ax + by + cz = 0\}, \tag{158}$$

where (a, b, c) is solution to the system of two linear equations $u_1 a + u_2 b + u_3 c = 0, v_1 a + v_2 b + v_3 c = 0$ for $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3)$.

For a line in $\mathbb{R}^2, \mathbb{R}^3$, see that any point on the line may be represented by a point x plus some vector u that is scaled. That is, the line may be written by some

$$L = \{x + tu \mid t \in \mathbb{R}\} \tag{159}$$

$$= \{x + w \mid w \in span(u)\}. \tag{160}$$

On the other hand, for some plane in $\mathbb{R}^3$, and u non-parallel to v, we may represent plane

$$P = \{x + su + tv \mid s, t \in \mathbb{R}\} \tag{161}$$

$$= \{x + w \mid w \in span\{u, v\}\}. \tag{162}$$

A generalization of this statement can be made in $\mathbb{R}^n$. That is,

1. for $x, u \in \mathbb{R}^n$, $u \neq 0$, the set

$$L = \{x + w \mid w \in span\{u\}\} \tag{163}$$

   is a line in $\mathbb{R}^n$.

2. For $x, u, v \in \mathbb{R}^n$, $u.v \neq 0$, and $u \neq kv$ for some $k \in \mathbb{R}$, then the set

$$P = \{x + w \mid w \in span\{u, v\}\} \tag{164}$$

   is plane in $\mathbb{R}^n$.

3. Take $x, u_1, u_2, \cdots u_r, \in \mathbb{R}^n$ the set

$$Q = \{x + w \mid w \in span\{u_1, \cdots, u_r\}\} \tag{165}$$

   is a k-plane in $\mathbb{R}^n$ where k is the dimension of the span$\{u_1, \cdots, u_r\}$. Dimensions of vector spaces are introduced in Section 3.1.3.6.

3.1.3.3 Subspaces

Definition 49 (Subspace)**.** *For $V \subseteq \mathbb{R}^n$, V is subspace of $\mathbb{R}^n$ if $V = span(S), S = \{u_1, \cdots, u_k\}$ for some vectors $u_{i\in[k]} \in \mathbb{R}^n$. We say that V is the subspace spanned by S. We say that S spans V. We say that $u_1, u_2, \cdots, u_k$ span V. We say that S is the spanning set for V.*

Definition 50 (Zero Space)**.** *From Definition 49 and Theorem 27, see that $0 \in \mathbb{R}^n$ spans the subspace that contains itself, that is $span\{0\} = \{0\}$. This is known as the zero space.*

Recall the vectors e_i's defined as in (Definition 47). For vectors $e_i, i \in [n] \in \mathbb{R}^n$, see that for all $u = (u_1, \cdots, u_n) \in \mathbb{R}^n$, we may write $u = \sum_i^n u_i e_i$, so it follows that $\mathbb{R}^n = span(\{e_1, \cdots, e_n\})$. Trivially, $\mathbb{R}^n$ is subspace of itself. In abstract linear algebra texts, the definition of subspace is relaxed to permit abstract objects and are usually provided as follows (see that Theorem 27 holds under this definition):

Definition 51 (Subspace)**.** *Let V be non-empty subset of $\mathbb{R}^n$. Then V is subspace of $\mathbb{R}^n$ iff*

$$\forall u, v \in V, \forall c, d \in \mathbb{R}, \qquad cu + dv \in V. \tag{166}$$

Theorem 30 (HLS Solution Space)**.** *The solution set of a HLS (Definition 18) in n variables is subspace of $\mathbb{R}^n$. We call this the solution space of the HLS.*

Proof. Let the matrix representation of the HLS be $Ax = 0$. If the HLS only has trivial solution, then the solution space is spanned by the trivial solution and is the zero space. Next, if it has non-trivial solution, then it has infinitely many solutions (see Lemma 2). Then by Definition 33, we may let solutions $x = \sum_i^{ncols(A)} a_i$ where a_i is column vector of the coefficient matrix A. That is, the solution space is spanned by the columns of A, and is therefore subspace of $\mathbb{R}^n$. □

If we solve some linear system and arrive at the general solution, it is easy to find the spanning vectors. For instance, let the general solution be

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2s - 3t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}. \tag{167}$$

The solution space is therefore $\{(2s - 3t, s, t) \mid s, t \in \mathbb{R}\} = span\{(2, 1, 0), (-3, 0, 1)\}$.

3.1.3.4 Linear Independence

We saw the concept of vector redundancy in a spanning set in Theorem 29. Here, we give formal treatment to such vectors with the concept of linear independence.

Definition 52 (Linear (In)Dependence)**.** *For set $S = \{u_i, i \in [k]\} \in \mathbb{R}^n$, consider $\sum_i^k c_i u_i = 0$, for $c_i \in \mathbb{R}, i \in [k]$. This has a HLS representation (Definition 18) where the coefficient matrix U is obtained from stacking the vectors horizontally, s.t $U = \begin{pmatrix} u_1 & \cdots & u_k \end{pmatrix}$ and $c = \begin{pmatrix} c_1 \\ \cdots \\ c_k \end{pmatrix}$ is the variable matrix. Then see that the zero solution satisfies the system always. The set S is said to be linearly independent and $u_1, \cdots, u_k$ are said to be linearly independent if the HLS only has the trivial solution. Otherwise, $\exists a_{i\in[k]} \neq 0$ and $\sum_i^k a_i u_i = 0$; a non-trivial solution exists. Then S is a linearly dependent set and $u_1, \cdots u_k$ are said to be linearly dependent vectors. For brevity, we use the notations*

$$LIND(S) = LIND\{u_1, u_2, \cdots, u_k\} \tag{168}$$

to indicate linear independence and

$$\neg LIND(S) = \neg LIND\{u_1, u_2, \cdots, u_k\} \tag{169}$$

to indicate linear dependence.

Let $S = \{u\}$ be a subset of $\mathbb{R}^n$, then S is linearly dependent iff $u = 0$. For $S = \{u, v\} \subset \mathbb{R}^n$, S is linearly dependent iff $u = av$, for some $a \in \mathbb{R}$. If $0 \in S$ for arbitrary $S \in \mathbb{R}^n$, it must be linearly dependent.

Theorem 31 (No Redundancy of Linearly Independent Set)**.** *Let $S = \{u_i, i \in [k]\} \subset \mathbb{R}^n$ where $k \geq 2$. Then, S is linearly dependent iff $\exists i \in [k]$ s.t. u_i is a linear combination of vectors in $S \backslash u_i$. Equivalent statement by the iff condition is that S is linearly independent iff no vector in S may be written as linear combination of the other vectors.*

Proof. $\rightarrow$: If $LIND(S)$, then $\sum_i^n a_i u_i = 0$ has non-trivial solution by Definition 52. Without loss of generality, let $a_i \neq 0$, then

$$u_i = -\frac{a_1}{a_i} - \frac{a_2}{a_i}u_2 - \cdots - \frac{a_{i-1}}{a_i}u_{i-1} - \frac{a_{i+1}}{a_i}u_{i+1} - \cdots - \frac{a_k}{a_i}u_k. \tag{170}$$

We have showed directly that u_i is l.c of the other vectors. $\leftarrow$: If $\exists u_i = \sum_{j\neq i}^k a_j u_j$, for some real numbers $a_{j\in[k],j\neq i}$. Then let $a_i = -1$, for which we have

$$a_1u_1 + \cdots + a_{i-1}u_{i-1} + a_iu_i + a_{i+1}u_{i+1} + \cdots + a_ku_k \tag{171}$$

$$= u_i - u_i \tag{172}$$

$$= 0. \tag{173}$$

So we have found some non-zero solution, and hence by definition, S must be linearly dependent. $\square$

Recall Theorem 26 on the minimum size of a spanning set required for $\mathbb{R}^n$. Here we give statements that allow us to determine the maximum size of the spanning set for $\mathbb{R}^n$ that is linearly independent.

Theorem 32. *Let $S = \{u_i, i \in [k]\} \in \mathbb{R}^n$. If $k > n$, then S is linearly dependent.*

Proof. The proof follows immediately by seeing that the HLS representation by stacking columns of u has non-trivial solutions by Lemma 2. S is linearly dependent by Definition 52 as a result. $\square$

Theorem 33 (No Redundancy of Non-Linearly Combinable Element)**.** *Let $u_i, i \in [k]$ be linearly independent vectors in $\mathbb{R}^n$. If $u_{k+1} \in \mathbb{R}^n$, and it is not l.c. of $u_i, i \in [k]$, then $\{u_i, i \in [k]\} \cup \{u_{k+1}\}$ is linearly independent.*

Proof. We show that the vector equation

$$\sum_i^{k+1} c_i u_i = 0 \tag{174}$$

has only trivial solution. See that $c_i, i \in [k]$ must be zero by itself in the HLS in k variables by assumption and definition for linear independence (Definition 52). We just need to show that $c_{k+1} = 0$. Suppose not, then we may write

$$u_{k+1} = -\sum_{i=1}^k \frac{c_i}{c_{k+1}} u_i \tag{175}$$

and this is a contradiction since we assumed no linear combination is possible. So, c_{k+1} must be zero. Therefore, the HLS represented for u_i, $i \in [k+1]$ must have only the trivial solution. $\square$

3.1.3.5 Bases

Definition 53 (Vector Spaces and Subspaces of Vector Space)**.** *A set V is vector space if either $V = \mathbb{R}^n$ or V is subspace (Definition 49, 51) of $\mathbb{R}^n$ for some $n \in \mathtt{Z}^+$. For some vector space W, the set S is subspace of W if S is a vector space contained inside W.*

We may be interested in finding the smallest set possible s.t. all vector in some vector space V may be represented as a linear combination of the elements in the set.

Definition 54 (Basis)**.** *Let $S = \{u_1, u_2, \cdots, u_k\}$ be subset of a vector space V (Definition 53). Then we say that S is a basis for V if (i) S is linearly independent (Definition 52) and (ii) S spans V (Definition 48). When $V = \{0\}$, the zero space, set $\emptyset$ to be the basis.*

That is, a basis for a vector space V must contain the smallest possible number of elements that can span V, since it must have no redundant vectors. Recall from Theorem 28 that for vector space V spanned by some set S, if all elements in S may be represented by some linear combination of vectors in $\tilde{S}$, and $\tilde{S}$ is linearly independent, then $\tilde{S}$ must be basis for $span(S) = V$ by definition of basis (Definition 54).

Theorem 34 (Unique Representation of Elements on Basis)**.** *If $S = \{u_i, i \in [k]\}$ is basis for vector space V, then $\forall v \in V$, v has unique representation $v = \sum_i^k c_i u_i$.*

Proof. Suppose $\exists c_{i\in[k]}, d_{j\in[k]}$ s.t. $v = \sum_{i=1}^k c_i u_i = \sum_{j=1}^k d_j u_j$, then by subtracting the two equations, get

$$(c_1 - d_1)u_1 + (c_2 - d_2)u_2 + \cdots + (c_k - d_k)u_k = 0. \tag{176}$$

But since S is linearly independent (it is basis), the only solution is the trivial solution, so $\forall i \in [k], c_i = d_i$. ☐

By Theorem 34, we should be able to specify an arbitrary vector in some vector space w.r.t to the coefficients of the l.c. on its basis.

Definition 55 (Basis Coordinates)**.** *Let $S = \{u_i, i \in [k]\}$ be basis for a vector space V and $v \in V$, then since v may uniquely expressed by some $c_i, i \in [k]$ (by Theorem 34) as $v = \sum_i^k c_i u_i$, we say that the coefficients c_i are coordinates of v relative to basis S and call the vector $(v)_S = (c_1, c_2, \cdots, c_k) \in \mathbb{R}^k$ the coordinate vector of v relative to basis S.*

To find the coordinate vector of some v relative to some basis S, we may simply solve for the linear system $\tilde{S}x = v$, where $\tilde{S}$ is coefficient matrix obtained by stacking the column vectors of elements of S. We give formal definition for a collection of vectors that we denoted e_i (Definition 47).

Definition 56 (Standard Basis)**.** *Let $E = \{e_i, i \in [n]\}$ where e_i is the vector of all zeros, except for a single entry of one in the ith-coordinate. Then it is easy to see that E spans $\mathbb{R}^n$, and that $LIND(E)$. E is basis for $\mathbb{R}^n$. In particular, we call this the standard basis, and see that*

$$(u)_E = (u_1, \cdots, u_n) = u. \tag{177}$$

Corollary 3. *By Definition 55, for basis S of V, $\forall u, v \in V$, $u = v$ iff $(u)_S = (v)_S$. Additionally, by Definition 55, $\forall v_{i\in[r]} \in V$, $c_{i\in[r]} \in \mathbb{R}$, see that*

$$(c_1 v_1 + c_2 v_2 + \cdots + c_r v_r)_S = c_1(v_1)_S + c_2(v_2)_S + \cdots + c_r(v_r)_S. \tag{178}$$

Theorem 35 (Linear Dependence Duality). *Let S be basis for vector space V (Definition 54, 53), and $|S| = k$. Let $v_i \in V, i \in [r]$, then*

1. $LIND(\{v_i, i \in [r]\}) \leftrightarrow LIND(\{(v_i)_S, i \in [r]\})$ *for vectors* $(v_i)_S \in \mathbb{R}^k$.
2. $span\{v_i, i \in [r]\} = V$ *iff* $span\{(v_i)_S, i \in [r]\} = \mathbb{R}^k$.

Proof. -

1. By Corollary 3, we can write $\sum_i^r c_i v_i = 0 \leftrightarrow (\sum_i^r c_i v_i)_S = (0)_S \leftrightarrow \sum_i^r c_i (v_i)_S = (0)_S$, where $(0)_S \in \mathbb{R}^k$. The first equality has non-trivial solution iff the last equality has the non-trivial solution and we are done.

2. Assume $S = \{u_i, i \in [k]\}$. $\rightarrow$: Assume $span\{v_i, i \in [r]\} = V$. Then by closure (Theorem 27) and basis definitions (Definition 54), we may write

$$\forall a = (a_1, \cdots, a_k) \in \mathbb{R}^k, \qquad w := \sum_i^k a_i u_i \in V = \sum_j^r c_j v_j \tag{179}$$

for some constants $c_j, j \in [r]$. By basis coordinate (Definition 55) and Corollary 3, we may write

$$a = (w)_S = (c_1 v_1 + \cdots + c_r v_r)_S = c_1 (v_1)_S + \cdots + c_r (v_r)_S. \tag{180}$$

Then it follows that $(v_i)_S, i \in [r]$ spans $\mathbb{R}^k$. $\leftarrow$: On the other hand, suppose $span\{(v_i)_S, i \in [r]\} = \mathbb{R}^k$. See that $\forall w \in V, (w)_S \in \mathbb{R}^k$ so $\exists c_i, i \in [r]$ s.t.

$$(w)_S = \sum_i^r c_i (v_i)_S = (\sum_i^r c_i v_i)_S, \tag{181}$$

and therefore $w = \sum_i^r c_i v_i$ by Corollary 3. Since we picked arbitrary w, we are done.

$\square$

3.1.3.6 Dimensions

Theorems 26 and 32 give statements of the number of elements required for a basis for a vector space that is $\mathbb{R}^k$ - here we use the duality given by Theorem 35 to make comments on arbitrary real vector space V.

Theorem 36 (Vector space has fixed size basis). *Let V be vector space with basis S, $|S| = k$. Then*

1. *Any subset of V with more than k vectors is always linearly dependent, and*
2. *Any subset of V with less than k vectors cannot span V.*

Proof. -

1. Let $T = \{v_i, i \in [r]\} \subset V$, and $r > k$. Then their coordinate vectors $(v_i)_S$ are set of r vectors in $\mathbb{R}^k$, and since $r > k$, by Theorem 32, $(v_i)_S, i \in [r]$ is linearly dependent, then by duality (Theorem 35) it follows that $\neg LIND(T)$.

2. Let $Q = \{v_i, i \in [t]\}, \subset V$ and $t < k$, then $(v_i)_S, i \in [t]$ may not span $\mathbb{R}^k$ (Theorem 26) and Q cannot span V by duality (Theorem 35).

□

Theorem 36 gives us a metric for the 'size' of a vector space. We formalize this with dimensions.

Definition 57 (Dimensions, *dim*)**.** *The dimension of a vector space V, denoted $dim(V)$ is the number of vectors in any basis for V. Since zero space has basis $\emptyset$ (Definition 54), $dim(0) = 0$.*

We can see that the dimension of a vector space denote the concept of degrees of freedom. Consider the subspace $W = \{(x, y, z)|y = z\}$. We may write $\forall w \in W, w := (x, y, y) = x(1, 0, 0) + y(0, 1, 1)$, s.t. $W = span\{(1, 0, 0), (0, 1, 1)\}$. Additionally, $(1, 0, 0), (0, 1, 1)$ are linearly independent and so they form basis. $dim(W) = 2$.

Exercise 33 (Finding the Nullity and Basis of a HLS Solution Space)**.** *By considering the (R)REF of an HLS (Definition 18), it is easy to see that the dimension of the solution space is the number of non-pivot columns (Definition 16) in the (R)REF form. To see this, suppose that the RREF representation of some HLS in variables (v, w, x, y, z) may be written to be*

$$\left[\begin{array}{ccccc:c} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \tag{182}$$

then by back substitution (Exercise 4), see that the linear system may have general solution

$$\begin{pmatrix} v \\ w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -s-t \\ s \\ -t \\ 0 \\ t \end{pmatrix} = s \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \tag{183}$$

for $s, t \in \mathbb{R}$. Then see that the dimension of the solution space is 2, and in fact we found the basis for the solution space $\{(-1, 1, 0, 0, 0), (-1, 0, -1, 0, 1)\}$. This solution space is known as the nullspace, and we have found the basis of the nullspace. The cardinality of this basis is known as the nullity. The nullspace, basis, and nullity are discussed later in Definition 64, Definition 54 and Definition 65 respectively.

Theorem 37. *Let V be vector space, dimension k (Definition 57) and $S \subset V$. The statements are equivalent for:*

1. *S is basis for V.*
2. *$LIND(S) \wedge |S| = k$.*
3. *S spans V and $|S| = k$.*

That is, if we know $|S| = k$, we only need to check if $span(S) = V$ or $LIND(S)$ to show it is basis for V.

Proof. The statements for $1 \to 2, 1 \to 3$ follow from Theorem 36. Additionally, to show $2 \to 1$, assume S is linearly independent and $|S| = k$. Suppose it is not basis for V, then take the vector $u \in V \wedge u \notin span(S)$. Then by Theorem 31, $S' = S \cup \{u\}$ is set of $k + 1$ linearly independent vectors, and Theorem 36 asserts the contradiction. To show $3 \to 1$, assume S spans V, $|S| = k$ and suppose S is not basis. Then $\exists v \in S$ s.t. $v = \sum_{s_i \in S \backslash v} c_i s_i$ for some constants $c_i \in \mathbb{R}$, and $\tilde{S} := S \backslash v$ is set of $k - 1$ vectors where $span(\tilde{S}) = span(S) = V$ by Theorem 29. Theorem 36 asserts the contradiction. □

Theorem 38 (Dimension of a Subspace)**.** *Let U be subspace (Definition 49) of vector space V. Then $dim(U) \leq dim(V)$. In particular, $U \neq V \implies dim(U) < dim(V)$.*

Proof. Let S be basis for U, so $S \subseteq U \subseteq V$ and since it is basis, S is linearly independent subset of V. By part 1, Theorem 36, since S is linearly independent, it must not have more than $k = dim(V)$ vectors, that is $dim(U) = |S| \leq dim(V)$. On the other hand, assume $|S| = dim(U) = dim(V)$, then Theorem 37 asserts that the linear independence of S and set cardinality makes $V = span(S) = U$. So we have shown that

$$dim(U) = dim(V) \implies U = V \tag{184}$$

Since $(dim(U) \leq dim(V)) \wedge (dim(U) \geq dim(V)) \leftrightarrow dim(U) = dim(V)$, we have effectively showed the contrapositive of the statement, and by logical equivalency we are done. □

Theorem 39 (Invertibility of Square Matrices, 2)**.** *If A is square matrix order n, then the following statements are equivalent:*

1. *A is invertible.*
2. *$Ax = 0$ has only the trivial solution.*
3. *RREF of A is identity $\mathbb{1}$ matrix.*
4. *A can be expressed as $\Pi_i^n E_i$, where E_i are elementary matrices.*
5. *$det(A) \neq 0$.*
6. *Rows of A form basis for $\mathbb{R}^n$.*
7. *Columns of A form basis for $\mathbb{R}^n$.*

Proof. See proof in Theorem 11 for the iff conditions for statement $1 \leftrightarrow 4$. $1 \leftrightarrow 5$ is proved by Theorem 21. $6 \leftrightarrow 7$ by Theorem 10 - rows of A are columns of A' and A invertible iff A' is invertible. We are done if we show any $i \in [5] \leftrightarrow 7$. We show $2 \leftrightarrow 7$. If $Ax = 0$ only has trivial solution, then the columns are linearly independent by the statements given in Definition 52. There are n columns. Then by Theorem 37, $\{a_1, a_2, \cdots a_n\}$ where a_i is i-th column of A is basis of $\mathbb{R}^n$. □

3.1.3.7 Transition Matrices

Definition 58 (Row/Column Vector Representation of Basis Coordinates)**.** *Recall that for basis $S = \{u_i, i \in [k]\}$ of vector space V and $v \in V$, v has unique coordinate vector representation (Definition 55, Theorem 34) written*

$$(v)_S = (c_1, \cdots, c_k) \tag{185}$$

and we write also write this as a column vector

$$[v]_S = \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_k \end{pmatrix}. \tag{186}$$

It is trivial that bases are not unique. For two bases S, T spanning vector space V, we may be interested in the relation $[w]_S \sim [w]_T$. This relation is captured by the transition matrix. In particular, let $S = \{u_i, i \in [k]\}, T = \{v_i, i \in [k]\}$ and some $w \in V$ be written $w = \sum c_i u_i$ s.t. $[w]_S = \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_k \end{pmatrix}$, then since each u_i's may be represented by the vectors in T, suppose

$$\forall i \in [k], \qquad u_i = a_{1i}v_1 + a_{2i}v_2 + \cdots + a_{ki}v_k. \tag{187}$$

That is, each $u_i \in [k]$ has T-basis coordinate representation $[u_i]_T = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \cdots \\ a_{ki} \end{pmatrix}$ and see that

$$w = \sum_j^k (c_1 a_{j1} + c_2 a_{j2} + \cdots + c_k a_{jk}) v_j. \tag{188}$$

That is,

$$[w]_T = \begin{pmatrix} c_1 a_{11} + c_2 a_{12} + \cdots c_k a_{1k} \\ c_1 a_{21} + c_2 a_{22} + \cdots c_k a_{2k} \\ \cdots \\ c_1 a_{k1} + c_2 a_{k2} + \cdots c_k a_{kk} \end{pmatrix} = \begin{pmatrix} [u_1]_T & [u_2]_T & \cdots & [u_k]_T \end{pmatrix} [w]_S. \tag{189}$$

Define $P = \begin{pmatrix} [u_1]_T & [u_2]_T & \cdots & [u_k]_T \end{pmatrix}$, then $[w]_T = P[w]_S$ for all $w \in V$ and we call P the transition matrix.

Definition 59 (Transition Matrix)**.** *Let $S = \{u_1, \cdots, u_k\}$ and T be two bases for vector space. Then $P = ([u_1]_T \cdots [u_k]_T)$ is said to be transition matrix from S to T, and $[w]_T = P[w]_S$ holds for all $w \in V$.*

We may find the transition matrix by the Gaussian Elimination (or Gauss Jordan) algorithm discussed in Theorem 5 and using the interpretations for linear systems as in Definition 33. For two bases $S = \{u_i, i \in [k]\}, T = \{v_i, i \in [k]\}$ respectively, we solve for the system with augmented matrix representation (Definition 9) $(T|u_1|u_2 \cdots |u_k)$, where T is coefficient matrix obtained from stacking column vectors v_i, $i \in [k]$. Then the column vectors on the RHS of the RREF augmented matrix are the weights for the linearly combined columns of T. In fact, the RHS of the augmented matrix from the first | onwards is precisely the transition matrix $P : [w]_S \to [w]_T$.

Theorem 40 (Properties of the Transition Matrix)**.** *Let S, T be two bases of vector space V and P be transition matrix from $S \to T$, then*

1. *P is invertible and*
2. *P^{-1} is the transition matrix from $T \to S$.*

Proof. It is easy to both logicize this argument and to prove it. Note that for $S = \{u_i, i \in [k]\}$, the vectors $[u_i]_S, i \in [k]$ is standard basis (Definition 56) in $\mathbb{R}^k$. Let Q be transition matrix from T to S. Then see that for $i \in [k]$, the i-th column of QP is written $QP[u_i]_S = Q[u_i]_T = [u_i]_S$. Then stacking the columns $[u_i]_S$, $i \in [k]$ gives us $\mathbb{1}_k$. □