MTH202 Linear Algebra for Engineers

Parin Chaipunya, Ph.D.

1. What is Linear Algebra all about?

Linear Algebra is a spin-off subject from the study of Linear Systems (or System of Linear Equations) through generalizations and re-interpretations.

$$\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n & = & b_2 \\ & \vdots & \\ a_{mn}x_1 + \cdots + a_{mn}x_n & = & b_m \end{array} \iff \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{mn} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \iff Ax = b.$$

In the above figure, we may observe a differnt view from a linear system into a simple matrix equation. This gives a new interpretation; instead of finding x that solves an equation, we find a vector x that is transformed (by a transformation A) into a new vector b.

In the first half of the course, we focus on the matrix equation $Ax = b$ and go through different approaches to characterize its solution. The topics include

- Solving $Ax = b$ by Gauss-Jordan method.
- Solutions of $Ax = b$ characterized by $rank(A)$.
- Determinant and its use in the equation $Ax = b$.
- Vector spaces, subspaces, and dimension. This part is mainly used to describe the solution set of $Ax = b$.
- Linear transformation. This part gives a new perspective to the matrix equation $Ax = b$.

The second half focuses on more advanced topics that revolves around representation of matrices in different bases. One of the most applied basis is the eigenbasis, which allows useful decomposition for faster matrix computation in modern GPUs. Moreover, we speak of geometric sides of linear algebra as well as some of its applications. The second half would cover

- Canonical forms. This allows representation of a matrix directly in the chosen basis.
- Eigenvalues and associated eigenvectors/eigenspace. Eigenbasis.
- Diagonalization. This is one of the decomposition methods that allows for fast computation in modern computers.

- Similar matrices.
- Norms. Orthogonality. Quadratic forms. `We explore geometric aspects of linear algebra here.`
- Some applications.

1.1 Matrix algebra: a quick review

Let us give a quick summary of elementary matrix algebra. Recall first that a matrix is an array of numbers listed as a tableau as follows

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

This matrix is said to have dimension of $m \times n$, i.e. having m rows and n columns. The i^{th} row of this matrix is the array

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix},$$

while the j^{th} column is the array

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

In this course, matrices are usually presented with capital roman alphabets $A, B, C, \cdots$, etc.

The anotomy of a matrix.
The following illustration depicts the i^{th} row and j^{th} column of a matrix:

$$i^{\text{th}} \text{ row} \rightarrow \begin{bmatrix} \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdots & \cdot \end{bmatrix} \qquad \begin{bmatrix} \cdot & \cdots & a_{1j} & \cdots & \cdot \\ \cdot & \cdots & a_{2j} & \cdots & \cdot \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \cdot & \cdot & a_{mj} & \cdot & \cdot \end{bmatrix}$$

$$\uparrow$$
$$j^{\text{th}} \text{ column}$$

Let us consider some special classes of matrices now.

Vectors.
A matrix that has a single column (i.e. of dimension $n \times 1$) is called a column vector. Likewise, a matrix that has a single row (i.e. of dimension $1 \times n$) is called a row vector. In this course, we use the simple term vector to refer to a column vector. Vectors in this course are usually presented with lower-case letters like $a, b, c, \cdots$ or more often with $x, y, u, v, \cdots$,etc.

To save up spaces, a column vector x (with n entries) is usually written also as $x = (x_1, \cdots, x_n)$. Hence, writting

$$x = (x_1, \cdots, x_n)$$

is the same as

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Square matrices.
A square matrix is the one having equal rows and columns. Let $A = [a_{ij}]_{n\times n}$ be an $n \times n$ square matrix. The diagonal of A refers to the elements $a_{11}, a_{22}, \cdots, a_{nn}$, as shown on the following figure:

$$\begin{bmatrix} a_{11} & \cdot & \cdots & \cdot \\ \cdot & a_{22} & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdots & a_{nn} \end{bmatrix}$$

↖ diagonal of A

A diagonal matrix is a square matrix whose nonzero entries are only on its diagonal. If the elements along the diagonal are $d_1, d_2, \cdots, d_n$, then we write $diag(d_1, \cdots, d_n)$ to denote such a diagonal matrix, that is

$$diag(d_1, \cdots, d_n) = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{bmatrix}.$$

The identity matrix of dimension $n \times n$ is the matrix I given by

$$I = diag(\underbrace{1, \cdots, 1}_{n \text{ times}}) = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{n \times n}$$

A matrix A is said to be an upper triangular matrix if all the entries below its diagonal are all 0. Likewise, it is said to be a lower triangular matrix if all the entries above its diagonal are all 0. The following figure illustrates, respectively, an upper triangular matrix and a lower triangular matrix:

$$\begin{bmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix} \qquad \begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ * & & a_{nn} \end{bmatrix}$$

We simply say that A is a triangular matrix if it is either an upper or a lower triangular matrix.

Let us now turn to algebraic operations of matricecs.

Transposition.
Suppose that A is an $m \times n$ matrix given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]_{m\times n}.$$

The transpose of A, written as A^T, is an $n \times m$ matrix defined by

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix} = [a_{ji}]_{n\times m}.$$

It is clear that $(A^T)^T = A$.

A matrix A is said to be symmetric if $A^T = A$. Of course, every symmetric matrix is a square matrix.

Scalar multiplication.
We can multiply a scalar (a constant) to a matrix in a componentwise fashion. Let $c \in \mathbb{R}$ be a scalar and $A = [a_{ij}]_{m\times n}$ a matrix. Then the multiplication of c with A, written cA, is defined by

$$cA = [ca_{ij}]_{m\times n} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}.$$

If $c = -1$, then we write $-A = (-1)A$.

Matrix addition and subtraction.
Adding/subtracting two or more matrices, all of them need to have the same dimension. Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$ be any two matrices. Then $A \pm B$ is defined componentwise, that is

$$A \pm B = [a_{ij} \pm b_{ij}]_{m\times n} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{22} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

For matrices A, B, C with the same dimension, it is clear that

- $A - B = A + (-B)$,
- and $A \pm B = B \pm A$,
- $(A \pm B)^T = A^T \pm B^T$,
- $(A \pm B) \pm C = A \pm (B \pm C) = A \pm B \pm C$,
- if c is a scalar, then $c(A \pm B) = cA \pm cB$.

Matrix multiplication.
Now, we consider multiplying two matrices A and B, denoted AB. For this, we require the number of columns of A and the number of rows of B to be the same. Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{n\times r}$, then $AB = [c_{ij}]$ is the matrix of dimension $m \times r$ where the entry c_{ij} is defined by

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nr}.$$

The following illustration helps understanding the calculation of c_{ij}:

$$\begin{bmatrix} \cdot & \cdot & \cdots & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdots & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdots & b_{1j} & \cdots & \cdot \\ \cdot & \cdots & b_{2j} & \cdots & \cdot \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \cdot & \cdot & b_{nj} & \cdot & \cdot \end{bmatrix} = \begin{bmatrix} & \cdot & \\ \cdot & c_{ij} & \cdots \\ & \vdots & \end{bmatrix}.$$

While m and r are not necessarily equal, BA is not even well-defined. For this reason alone, there is no reason that AB and BA are the same. In general, even for square matrices A and B, AB and BA are not equal.

Let A, B, C be two matrices. Then we have

- $A(B \pm C) = AB \pm AC$,
- $(A \pm B)C = AC \pm BC$,
- $(AB)C = A(BC) = ABC$,
- $(AB)^T = B^T A^T$,
- $IA = A = AI$, (The two I's can be of different dimension!)
- if c is a scalar, then $c(AB) = (cA)B = A(cB)$,

whenever the above products are well-defined.

Let us practice a little bit to conclude this review session.

Exercise 1.1. Define

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -1 & -2 \\ -1 & 0 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \text{ and } C = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix}.$$

Find

(a) AB (b) BC (c) ABC (d) AC

(e) BA (f) $B^T A$ (g) $C^T B A^T$ (h) $AA^T + C^T C$

Lab session 1

This lab session is an introduction to the MATLAB application and elementary syntax. The following topics will be covered:

- defining variables, matrices and vectors,
- accessing variables, entries of a matrix / vector, rows and columns,
- elementary matrix operations and broadcasting.

Assignment of the week.

Let $A = \begin{bmatrix} 1 & 3 & 0 & 0 & -1 \\ 2 & 2 & 3 & -2 & 2 \\ 3 & 4 & 9 & -2 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 2 & 2 \\ 3 & -5 & -1 \\ 0 & -2 & 1 \\ -1 & -1 & 0 \end{bmatrix}$. Find the following:

(a) The first and last rows of AB.

(b) The first and last columns of BA.

2. Linear systems and matrix equations

2.1 Transforming a linear system into a matrix equation

Recall that a single-variable equation is said to be linear if it is in the form

$$ax = b.$$

This can be easily solved if $a \neq 0$ achieving $x = \frac{b}{a}$.

Having two variables, say x and y, we need two equations to fix the solution. In this way, we cannot just choose any x and y independently because they are linked by the given equations. This system of two equations are said to be a linear system if it is in the form

$$\begin{aligned} a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2. \end{aligned}$$

To solve this system, we first eliminate a variable and fix the solution of the other then substitute back into one of the original equations.

The same principle applies to the case of linear systems with three variables, x, y and z, as the following form

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3. \end{aligned}$$

To find a solution, we eleminate and substitute variables one by one.

We may now see the pattern: A linear system with n variables and n equations, namely $x_1, \cdots, x_n$, takes the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad\qquad & \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

Again, the same elimination-and-substitution strategy still works here, but with a lot more effort.

Question. Is it necessary to have equal number of variables and equations?

The simple answer to the above question is "no." We may have n variables that are linked by m linear equations:

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \vdots \\ \underbrace{a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n}_{n \text{ variables}} & = & b_m. \end{array}\right\} m \text{ equations}$$

The array of values $u_1, \cdots, u_n$ is called a solution of the above linear system if subsituting $x_1 = u_1, \cdots, x_n = u_n$ into the LHS gives RHS.

To review and familiarize linear systems, let us do some preliminary exercises.

Exercise 2.1. Verify whether or not $x_1 = 3, x_2 = -1$ a solution of the following linear system

$$\begin{aligned} 2x_1 - 3x_2 &= 9 \\ x_1 + x_2 &= 2. \end{aligned}$$

Exercise 2.2. Verify whether or not $x_1 = 2, x_2 = 5$ a solution of the following linear system

$$\begin{aligned} x_2 - x_1 &= 3 \\ 2x_1 - 3x_2 &= -11. \end{aligned}$$

On the other hand, how do we find a solution of this system?

Exercise 2.3. Find a solution to the following linear system

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

Exercise 2.4. Find a solution to the following linear system

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

Solving the previous systems, we may have observed that our calculation is dependent on the given equations. This is difficult to proceed when the number of variables and equations grow larger. Hence we would need a more disciplined strategy, and this involves changing the system into a matrix equation.

Turning a linear system into a matrix equation.
Given a linear system

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

we would define

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We call the A the coefficient matrix, b the target vector, and x an unknown vector. Then we have

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

and so the matrix equation

$$Ax = b$$

is equivalent to the linear system in question.

Augmented matrix for $Ax = b$.
To solve $Ax = b$ by hand, it is often helpful to simplify it as an augmented matrix $[A \,|\, b]$, that is

$$[A \,|\, b] = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right].$$

Row operations.
Recall the operations that we did when solving linear systems in the previous exercises:

- adding an equation into another,
- multiplying a constant to an equation,
- swapping order of equations.

These operations transfer to the matrix equation as (elementary) row operations, which include

- adding a row into another,
- multiplying a constant to a row,
- swapping rows.

2.2 Analyzing the equation $Ax = b$

We are not only interested in solving $Ax = b$, but also to understand it.

The first and foremost behavior we would like to observe about the equation $Ax = b$ is the number of solutions it has. We may very roughly speak of consistent systems (those having a solution) and inconsistent systems (those without a solution), but we want a more detailed description in this course. This leads us to consider the three cases:

- (well-posed) $Ax = b$ has a unique solution,
- (under-determined) $Ax = b$ has multiple solutions,
- (over-determined) $Ax = b$ has no solution.

To convince that all the three cases could happen, let us do the following exercise.

Exercise 2.5. Let us observe that:

(a) Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$.
Then $Ax = b$ has a unique solution, which is $x^* = (1, 1, 1)$.

(b) Let $A = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$ and $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
Then $Ax = b$ has multiple solutions. For example, $x^* = (0, 1, 0)$ and $x^{**} = (1, 2, 1)$ are both solutions for this system.

(c) Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ -1 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.
Then $Ax = b$ has no solution.

It is *casually understood* that

- a well-posed system has equal number of variables and equations,
- an under-determined system has less equations than variables,
- an over-determined system has more equations than variables.

However, to *really* use this verdict, we need to be sure that the systems is reduced into its simplest form, i.e. the row echelon form.

Let us do one more exercise to persuade ourselves that we need to reduce a system before judging it.

Exercise 2.6. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$ and $b = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}$.
Then we have 3 variables and 3 equations, but multiple solutions. For instance, $x^* = (2, 0, 1)$ and $x^{**} = (3, 1, 0)$ are both solutions of this system. We will come back to explain this exercise again.

2.2.1. Echelon forms

Row echelon forms.
Consider a matrix $M = [m_{ij}]_{m \times n}$. A row of M is called a zero row if it only consists of 0. A row that is not a zero row is called non-zero row. In a non-zero row, the first non-zero column in that row is called the leading entry.

Looking at the following matrix

$$P = \begin{bmatrix} 0 & 0 & 0 & \circledast & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & \circledast & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix},$$

we conclude that the rows #2 and #4 are zero rows of this matrix, while he circled entries ($*$'s are any nonzero numbers) are the leading entries of their rows.

This matrix M is in a row echelon form if the following two conditions are satisfied:

(1) All zero rows are stacked at the bottom.

(2) In each non-zero row, its leading entry is on the left of all the leading entries below it.

One may simply observe this visually by stacking zero rows at the bottom and see that the leading entries are arranged in a staircase manner (see an example below with $\hat{P}$).

Notice that the above matrix P is not in an echelon form (why?). However, we may use row operations (swapping rows) to acheive an equivalent matrix in the following echelon form

$$\hat{P} = \begin{bmatrix} 0 & \circledast & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \circledast & \cdots \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

Reduced row echelon form.
A matrix M is in a reduced row echelon form if it is in a row echelon form and the following additional conditions are satisfied:

(3) all leading entries are 1,

(4) all entries above a leading entry are 0.

Exercise 2.7. Determine whether or not the matrices given in the following are in row echelon and reduced row echelon form.

(a) $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix}$

(b) $B = \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 9 \\ 0 & 0 & -1 & 2 & -2 & 5 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

(c) $C = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -2 & 5 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(d) $D = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Let us note with the following facts.

Fact.

- Every matrix can be turned into a row echelon form using elementary row operations.
- A single matrix may be turned into several row echelon forms.
- every matrix can be reduced into a unique *reduced* row echelon form

Exercise 2.8. Use elementary row operations to find row echelon forms of the matrices from Exercise 2.7.

2.2.2. Gauss elimination

The aim of the Gauss elimination is to give a systematic procedure to simplify any matrix into a row echelon form. To do this, we eliminate as many nonzero entries as possible by the following the steps below:

1. Start with the first nonzero column. Pick a nonzero entry, swap that row to the top. Set $i = 1$.
2. Eliminate nonzero entries below the chosen entry from the previous step.
3. Find the next column with nonzero entry after the row i. Pick that nonzero entry and swap to the row $i + 1$.
4. Eliminate all the nonzero entries below the chosen entry from the previous step.
5. Set $i \leftarrow i + 1$ and repeat Steps 3–5 to complete all columns/rows.

Exercise 2.9. Use Gauss elimination to turn the following matrices into a row echelon form..

(a) $A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix}$

(b) $B = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 2 & 1 & 0 \\ 2 & 4 & 1 & 1 \\ 3 & 6 & 2 & 1 \end{bmatrix}$

(c) $C = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(d) $D = \begin{bmatrix} 2 & 0 & 0 & 1 & 2 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$

An observation.

- Every square matrix's row echelon form is a triangular matrix.

2.2.3. Gauss-Jordan elimination

The process of Gauss elimination reduces any given matrix into a row echelon form. However, if we would like to go further in obtaining a *reduced* row echelon form, we need a small extra computation in the Gauss elimination steps. This modified algorithm is known as the Gauss-Jordan elimination whose steps are listed as follows:

1. Start with the first nonzero column. Pick a nonzero entry, *make it into* 1 and swap that row to the top. Set $i = 1$.
2. Eliminate *all nonzero entries in that column* except the 1 from the previous step.
3. Find the next column with nonzero entry after the row i. Pick that nonzero entry, *make it into* 1 and swap to the row $i + 1$.
4. Eliminate *all the nonzero entries in that column* except the 1 from the previous step.
5. Set $i \leftarrow i + 1$ and repeat Steps 3–5 to complete all columns/rows.

Exercise 2.10. Use Gauss-Jordan elimination to obtain the reduced row echelon form of the following matrices.

(a) $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 1 & 0 & -1 & 2 & 3 \end{bmatrix}$ (b) $B = \begin{bmatrix} -2 & 1 & 3 & 0 \\ 2 & 1 & 3 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 1 & 3 \end{bmatrix}$

2.2.4. Rank and solutions of $Ax = b$

We regard each row of a matrix as an information. A matrix is then a collection of information. Some of these information may be obtained from others and hence does not genuinely a new information. A *row echelon form* (or *reduced row echelon form*) of a matrix is viewed as a collection of *genuine* information in the sense that each row cannot be deduced from the remaining rows, hence no rows can be removed. It is natural that each matrix contains a certain amount of information, and this amount is fixed.

Nonzero rows of row echelon matrices.
Let A be a given matrix. Then

- all row echelon forms of A has the same number of nonzero rows,
- the reduced row echelon form of A is unique.

With the consistence of nonzero rows in the above observation, one may define a rank of a matrix to be the number of nonzero rows in a row echelon form. That is, the rank is the number of genuine information that a matrix carries.

Rank.
The rank of a matrix A is defined to be the number of nonzero rows of any row echelon form of A. The rank of A is denoted by $rank(A)$.

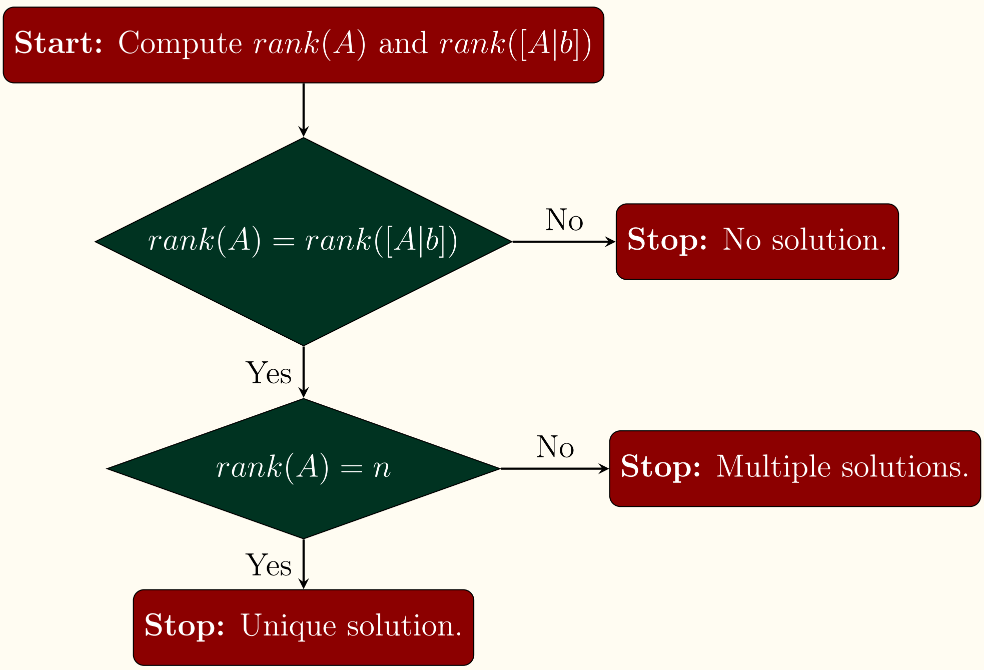
Exercise 2.11. Find the rank of the matrices from the previous exercises.

Recall that a row echelon form of a matrix reveals the minimal true information that it carries. Therefore, the number of solutions of a matrix equation $Ax = b$ can be observed from the row echelon form of the augmented matrix $[A|b]$. More precisely, the number of solutions of $Ax = b$ can be concluded from the $rank([A|b])$.

Number of solutions through the rank.
Consider the equation $Ax = b$ that corresponds to a system of n variables and m equations. Then the following conclusion is drawn:

- $rank(A) < rank([A|b]) \iff Ax = b$ has no solution;
- $rank(A) = rank([A|b]) = n \iff Ax = b$ has a unique solution;
- $rank(A) = rank([A|b]) < n \iff Ax = b$ has multiple solutions.

The following flow chart helps the explaining the use of rank to determine the number of solutions of $Ax = b$.



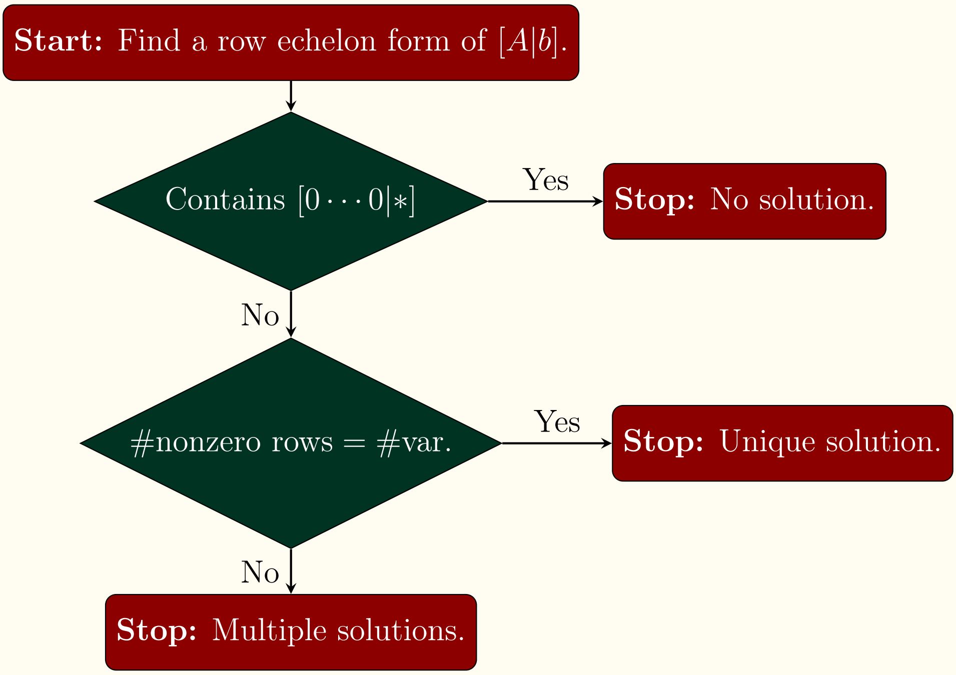
Practical aspects.
Consider again the system represented by $Ax = b$ with n variables and m equations. In practice, to see if $rank(A) < rank([A|b])$ happens, we do not

need to compute $rank(A)$ and $rank([A|b])$ separately.

Consider directly the matrix $[A|b]$. When reducing $[A|b]$ to one of its row echelon form $[\hat{A}|\hat{b}]$, the $\hat{A}$ is always a row echelon form of A. Therefore, we also obtain a row echelon form of A in the process of obtaining a row echelon form of $[A|b]$. Then the following hold:

> If a row echelon form of $[A|b]$ consists of a row of the form $[0 \cdots 0|*]$ (where $* \neq 0$), then $rank(A) < rank([A|b])$ and so $Ax = b$ has no solution.

Let us rewrite the above flow chart in a more practical perspective.



Exercise 2.12. Determine the number of solutions of the following system:

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

Exercise 2.13. Determine the number of solutions of the following system:

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

Exercise 2.14. Determine the number of solutions of the following system:

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

Exercise 2.15. Determine the number of solutions of the following system:

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

2.3 Finding solutions of $Ax = b$.

Now, after being able to determine the number of solutions of $Ax = b$, it is time to actually find its solutions (if one exists). The strategy is to simplify the augmented matrix $[A|b]$ using either Gauss elimination or Gauss-Jordan elimination, and start analyzing from there.

Exercise 2.16. Solve the following linear system

$$\begin{aligned} x_1 + x_2 - x_3 &= 6 \\ x_1 - x_2 + x_3 &= -2 \\ 2x_1 - 2x_2 + 3x_3 &= -5 \end{aligned}$$

Exercise 2.17. Solve the following linear system

$$\begin{aligned} x_1 + x_2 - x_3 - x_4 &= 0 \\ x_1 - x_2 + x_3 + x_4 &= 2 \\ x_2 - 2x_3 &= -1 \\ x_1 + x_2 + 3x_3 + 2x_4 &= 7 \end{aligned}$$

Exercise 2.18. Solve the following linear system

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

Exercise 2.19. Solve the following linear system

$$\begin{aligned} x_1 + 2x_2 - x_3 + x_4 &= 3 \\ 2x_1 - 2x_2 + 2x_3 - x_5 &= 2 \\ 2x_1 - 3x_3 &= -1 \end{aligned}$$

3. Vector spaces

Let us motivate the study of a vector space, again, from the equation $Ax = b$ that corresponds to a system of n variables and m equations. Let us begin with the special case where $b = 0$.

Multiple solutions... Yes, but how many?
It turns out that when $Ax = 0$ has more than one solution, it has infinitely many of them. Let us demonstrate this fact. Let x^* and x^{**} both be solutions of $Ax = 0$. Take any numbers α and β. Then we get

$$A(\alpha x^* + \beta x^{**}) = \alpha A x^* + \beta A x^{**} = 0.$$

This means from two different solutions, we can generate infinitely many more by choosing different values of α and β.

Let $\mathcal{S}$ be the set that contains all the solutions of $Ax = 0$. It turns out that $\mathcal{S}$ has a particular shape to it, e.g. a single dot, a straight line, a flat plane, a space, etc. We actually observe that all of them are *linear shapes*. In fact, we shall see subsequently that a solution set of the equation $Ax = 0$, if not empty, is a linear space (or a vector space).