



# Lecture 1

Math Foundations Team



**BITS Pilani**

Pilani | Dubai | Goa | Hyderabad

July 26, 2025

# What is linear algebra?



- ▶ Linear algebra is the study of vectors and rules to manipulate vectors.
- ▶ Vectors are not only the familiar geometric vectors from high school (points in 2D/3D space) but any special objects which can be added together and multiplied by scalar values to produce another object of the same kind. For example, polynomials can also be treated as vectors.
- ▶ We shall deal with vectors in the space  $R^n$



- ▶ Let's say we have a bunch of mathematical objects and we perform some operations on them. Do we get back similar objects?
- ▶ This leads to the idea of a vector space which underlies much of machine learning.



- ▶ Systems of linear equations form a central part of linear algebra.
- ▶ Many problems can be formulated as systems of linear equations.
- ▶ Tools of linear algebra can be used to solve such problems.

# Scenario: Manufacturing Problem



A furniture company produces three types of kits: Basic Chair Kits, Table Kits, and Cabinet Kits. Each of these kits requires specific amounts of resources, including wood, labor hours, machine time, and shipping effort. The production of each kit consumes different amounts of these resources. The company has a limited supply of each resource. All these constraints must be met exactly. The objective is to determine how many units of each furniture kit—Basic Chair, Table, and Cabinet—should be produced to fully utilize the available resources while satisfying all the given limitations.

# Interpretation of the Matrix; Available Resources & Constraints



Resource	Basic Chair Kit ( $x_1$ )	Table Kit ( $x_2$ )	Cabinet Kit ( $x_3$ )	Total Available
Wood	1 unit	2 units	1 unit	9 units
Labor Hours	2 hours	1 hour	1 hour	8 hours
Machine Time	1 hour	1 hour	2 hours	7 hours
Shipping Effort	3 effort units	1 effort unit	2 effort units	11 effort units

**Objective:** Determine how many units of each furniture kit to produce.

**Products:**

- ▶  $x_1$ : Basic Chair Kit
- ▶  $x_2$ : Table Kit
- ▶  $x_3$ : Cabinet Kit

## Resources and Constraints:

- ▶ Wood used:  $x_1 + 2x_2 + x_3 = 9$
- ▶ Labor hours:  $2x_1 + x_2 + x_3 = 8$
- ▶ Machine time:  $x_1 + x_2 + 2x_3 = 7$
- ▶ Shipping effort:  $3x_1 + x_2 + 2x_3 = 11$

## Solution:

$$x_1 = 2, \quad x_2 = 3, \quad x_3 = 1$$

## Interpretation:

- ▶ Produce 2 Chair Kits
- ▶ Produce 3 Table Kits
- ▶ Produce 1 Cabinet Kit

# Does a linear system always have solutions?



- ▶ A linear system has zero, one or infinitely many solutions
- ▶ Linear regression, a Machine Learning technique, provides an approximate solution to an overconstrained linear system, i.e one with no solution



Consider the following system of linear equations

$$x_1 + x_2 + x_3 = 3$$

$$x_1 - x_2 + 2x_3 = 2$$

$$2x_1 + 3x_3 = 1$$

Adding the first and second equations gives  $2x_1 + 3x_3 = 5$  which contradicts the third equation. Thus there is no set of values for the variables  $x_1, x_2, x_3$  such that the equations above are simultaneously satisfied.



Consider a slightly modified example

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

In this case we can see from the first and third equations that  $x_1 = 1$ . Substituting this value of  $x_1$  into equation (2), we get  $-x_2 + 2x_3 = 1$ . Adding this equation to equation (3), we get  $3x_3 = 3$  which means  $x_3 = 1$ . Substituting  $x_3 = 1$  into equation (3) shows  $x_2 = 1$ , so the overall solution is  $x_1 = x_2 = x_3 = 1$ . This is the unique solution to the problem



Now consider another modification to the original set of equations

$$x_1 + x_2 + x_3 = 3$$

$$x_1 - x_2 + 2x_3 = 2$$

$$2x_1 + 3x_3 = 5$$

Adding the first and second equations gives  $2x_1 + 3x_3 = 5$  which is the same as the third equation. Thus the solution to the three equations is any tuple  $x_1, x_2, x_3$  which satisfies  $2x_1 + 3x_3 = 5$ , and there are infinite solutions. We now express these solutions in a way whose motivation will become clear later: adding equations (1) and (2) above we get  $2x_1 = 5 - 3x_3$ .



- ▶ Subtracting equation (2) from (1) we get  $2x_2 - x_3 = 1$ , so we can write

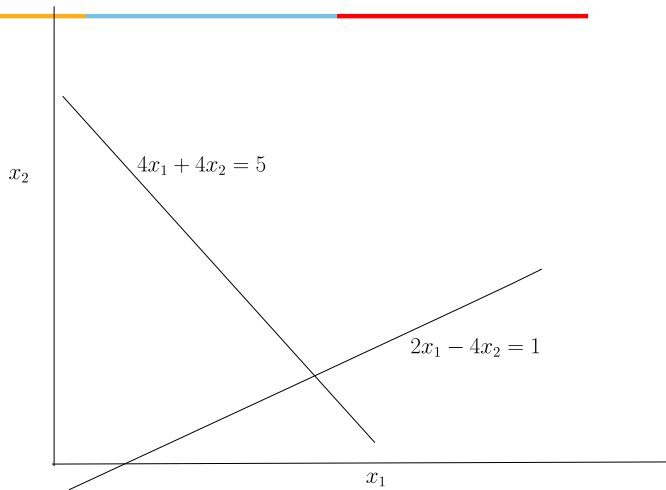
$$\begin{aligned}x_1 &= \frac{5}{2} - \frac{3}{2}x_3 \\x_2 &= \frac{1}{2} + \frac{x_3}{2}\end{aligned}$$

- ▶ For the previous problem we can express the set of infinite solutions in terms of the free variable  $x_3$ .
- ▶ Once  $x_3$  is fixed, the other two variables have to take on specific values - they are known as pivot variables.
- ▶ We will show later how to identify pivot and free variables using Gaussian Elimination

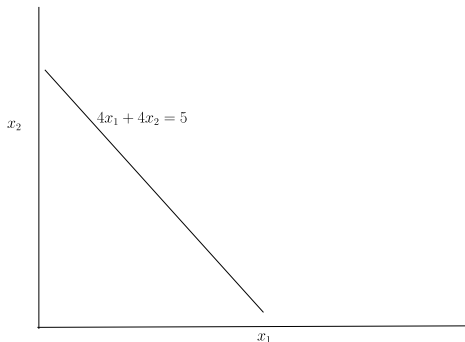


The solution to linear equations can be given a geometrical interpretation. When equations are given in terms of two variables, the solution to two equations in two variables could be a point (unique solution), a line (infinite solutions) or no solution (parallel lines). The first case is shown in the next slide.

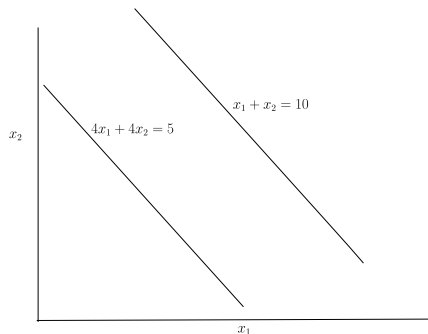
# Geometrical Interpretation



In the second case both constraints are the same, so there are an infinite number of solutions:



In the third case the constraints are mutually incompatible, so there is no assignment to  $x_1, x_2$  which satisfies both constraints. The graph of both constraints shows a pair of parallel lines:







- ▶ In 3D each constraint is a plane.
- ▶ The intersection of two planes is a line.
- ▶ The intersection of the third plane with the first two planes will be a point on the line in case of a unique solution, or it may lead to pairs of parallel lines (constraint 1 intersection constraint 2 gives one line, constraint 1 intersection constraint 3 gives parallel line, constraint 2 intersection constraint gives parallel line) which means there is no solution.
- ▶ All three constraints or planes may intersect in the same line which means infinite solutions.

An  $(m, n)$  matrix  $A$  is a  $mn$  tuple of elements  $a_{ij}$  arranged in  $m$  rows and  $n$  columns as below:

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

Note that row and column vectors are also matrices - a row vector is a  $(1, n)$  matrix and a column vector is a  $(m, 1)$  matrix.  $R^{m \times n}$  is the set of all  $(m, n)$  matrices consisting of real numbers as elements



The sum of two matrices  $A \in R^{m \times n}$ ,  $B \in R^{m \times n}$  is defined as an element-wise sum of the elements of the two matrices:

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & & & \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$



- ▶ The product of two matrices  $A \in R^{m \times k}$ ,  $B \in R^{k \times n}$  is defined as the  $m \times n$  matrix  $C = AB$  where the elements  $c_{ij}$  are calculated as follows:  $c_{ij} = \sum_{l=1}^{l=k} a_{il}b_{lj}$
- ▶ Essentially we are multiplying the elements of the  $i$ th row of  $A$  with the  $j$ th column of  $B$ .
- ▶ The product  $BA$  is not defined if  $n \neq m$



- ▶ Associativity:  
 $\forall A \in R^{m \times n}, B \in R^{n \times p}, C \in R^{p \times k} (AB)C = A(BC)$
- ▶ Distributivity:  $\forall A, B \in R^{m \times n},$   
 $\forall C, D \in R^{n \times p}, (A + B)C = AC + BC, A(C + D) = AC + AD$
- ▶ Multiplication with the identity matrix:  $\forall A \in R^{m \times n},$   
 $I_m A = A I_n = A$  where  $I_m$  is the  $m \times m$  identity matrix and  $I_n$  is the  $n \times n$  matrix.



Consider a square matrix  $A \in R^{n \times n}$ . Let matrix  $B \in R^{n \times n}$  exist such that  $AB = I_n$ . Does every matrix  $A \in R^{n \times n}$  possess an inverse?

Let us take a simple  $2 \times 2$  example  $\rightarrow$  under what circumstances does it possess an inverse?

## $2 \times 2$ case



Define a  $2 \times 2$  matrix  $A$  as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Define matrix  $B$  to be

$$B = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

The product  $AB$  is

$$AB = \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & a_{12}a_{11} - a_{11}a_{12} \\ -a_{21}a_{22} - a_{22}a_{21} & -a_{21}a_{12} + a_{11}a_{22} \end{bmatrix}$$



$$AB = \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & 0 \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix} = (a_{11}a_{22} - a_{12}a_{21})I_2$$

$AB = (a_{11}a_{22} - a_{12}a_{21})I_2$ . We can define

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

whenever  $a_{11}a_{22} - a_{12}a_{21} \neq 0$ .

Bottomline is that the inverse of a square matrix exists if and only if its determinant is non-zero.



# Inverse and Transpose Properties



For  $A \in R^{m \times n}$ , the matrix  $B \in R^{n \times m}$  with  $b_{ij} = a_{ji}$  is called the transpose of  $A$ . We say  $B = A^T$ . Transposition means writing the rows of one matrix as the columns of the other. The following properties can be shown:

$$\begin{aligned}AA^{-1} &= A^{-1}A = I \\(AB)^{-1} &= B^{-1}A^{-1} \\(A + B)^{-1} &\neq B^{-1} + A^{-1} \\(A^T)^T &= A \\(A + B)^T &= A^T + B^T \\(AB)^T &= B^T A^T\end{aligned}$$

# Proof of $(AB)^T = B^T A^T$



- ▶ How do we show these properties? Let us look at the last one for example ...
- ▶ The  $(i, j)$ th entry of  $AB$  is obtained by taking the inner product of the  $i$ th row of  $A$  with the  $j$ th column of  $B$ .
- ▶ But the  $i$ th row of  $A$  is the  $i$ th column of  $A^T$ , and the  $j$ th column of  $B$  is the  $j$ th row of  $B^T$ .
- ▶ Thus we can take the inner product of the  $j$ th row of  $B^T$  with the  $i$ th column of  $A^T$  to get the same value for the  $(i, j)$ th entry of  $AB$ . Thus when we compute  $C = B^T A^T$ , we find that  $C_{ji} = (AB)_{ij}$ , so  $C = (AB)^T$ .

- ▶ Let  $A \in R^{m \times n}$  and  $\lambda$  be a scalar. The  $\lambda A = B$  such that  $B_{ij} = \lambda A_{ij}$ , i.e every element in  $A$  is scaled by  $\lambda$  to get the corresponding element in the scaled matrix  $B$ .
- ▶ For  $\lambda, \psi \in R$  we have
  - ▶ associativity:  $(\lambda\psi)A = \lambda(\psi A)$ ,  $A \in R^{m \times n}$  and  $\lambda(AB) = (\lambda A)B = A(\lambda B)$  where  $A \in R^{m \times n}$  and  $B \in R^{n \times k}$ .
- ▶ distributivity:
  - ▶  $(\lambda A)^T = A^T \lambda^T = A^T \lambda = \lambda A^T$  since  $\lambda = \lambda^T$  for all  $\lambda \in R$ .
  - ▶  $(\lambda + \psi)A = \lambda A + \psi A$
  - ▶  $\lambda(A + B) = \lambda A + \lambda B$



Consider the following system of equations:

$$2x_1 + 3x_2 + 5x_3 = 1$$

$$4x_1 - 2x_2 - 7x_3 = 8$$

$$9x_1 + 5x_2 - 3x_3 = 2$$

In matrix terms we can write this set of equations as

$$\begin{bmatrix} 2 & 3 & 5 \\ 4 & -2 & -7 \\ 9 & 5 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 2 \end{bmatrix}$$

# Solving a system of equations



- ▶ From the previous slide we see that a system of equations can be expressed as  $Ax = b$  where  $A \in R^{m \times n}$ ,  $x$  is a  $n \times 1$  matrix and  $b$  is a  $m \times 1$  matrix.
- ▶ We now look at how to obtain a particular and general solution for a system of equations
- ▶ Let us first look at an example.

# Example of system of equations



$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

- ▶ This system has two equations and four unknowns, so it is underconstrained. We expect an infinity of solutions.
- ▶ Is there a special way in which to express the solutions to this system?
- ▶ Let us examine the structure of the given problem matrix.



- ▶ Looking at the previous slide we can see that a linear combination of columns of the matrix will give the right hand side.
- ▶ The  $i$ th column vector in the matrix appears in the linear combination, scaled by the corresponding  $x_i$  as below.

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} -4 \\ 12 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

## Particular solution to the example



- ▶ A closer look at the linear combination to give the right hand side shows that we can take  $x_1 = 42$ ,  $x_2 = 8$ ,  $x_3 = 0$ ,  $x_4 = 0$  since the first two columns are  $(1, 0)^T$  and  $(0, 1)^T$  respectively.
- ▶ Therefore a solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix}$$

- ▶ This solution is called the particular solution



# Any other solutions possible?



- ▶ We can generate other solutions than the particular solution, by adding the vector  $\mathbf{0}$  to the particular solution
- ▶ But isn't this the same as the particular solution as any vector  $+\mathbf{0}$  is that vector itself?
- ▶ The trick is to express  $\mathbf{0}$  in terms of the linear combination of some vectors.
- ▶ Describing  $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4$  as the four column vectors associated with the given matrix in the example we can see that  $8\mathbf{c}_1 + 2\mathbf{c}_2 - 1\mathbf{c}_3 + 0\mathbf{c}_4 = \mathbf{0}$ .

# Any other solutions possible?



- ▶ Writing the linear combination in terms of a matrix-vector product we have

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- ▶ Any vector  $\lambda(8, 2, -1, 0)^T$ ,  $\lambda \in R$  will also produce the **0** vector

# Any other solutions possible?



- ▶ We can add the vector  $(8, 2, -1, 0)^T$  to the original particular solution  $(42, 8, 0, 0)^T$  to get another solution since

$$A\left(\begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 42 \\ 8 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

# Any other solutions possible?



- ▶ Following the same line of reasoning as before, we can create the **0** vector by expressing the fourth column of the matrix **A** in terms of the first two columns - note that the first two columns appear capable of generating any two-dimensional vector!
- ▶ We can see that  $-4\mathbf{c}_1 + 12\mathbf{c}_2 + 0\mathbf{c}_3 - 1\mathbf{c}_4 = \mathbf{0}$ .
- ▶ Thus we have

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} (\lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



- ▶ We obtain the following general solution as the sum of the particular solution and a linear combination of solutions to the equation  $\mathbf{Ax} = \mathbf{0}$  as follows:

$$\{\mathbf{x} \in \mathbf{R}^4 : \mathbf{x} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}\}$$

- ▶ The general approach consisted of finding a particular solution to  $\mathbf{Ax} = \mathbf{b}$ , finding all solutions to  $\mathbf{Ax} = \mathbf{0}$  and combining the particular and general solutions.
- ▶ Neither the particular nor general solutions are unique  $\rightarrow$  why?

# Algorithmic way of solving equations



- ▶ The system of equations in our example was easy to solve because of the special structure of the matrix - we could guess the solution without much difficulty.
- ▶ Can we develop an algorithmic way of solving a general system of equations?
- ▶ The answer is yes → we call the procedure Gaussian elimination



- ▶ The key idea is to take a complex looking matrix and transform it using elementary row operations to a simple looking matrix like the one we just handled, for which solutions could be obtained essentially by inspection.
- ▶ To make this work we need to preserve solutions of the original system of equations, i.e ensure that elementary transformations of the original matrix do not change its solutions.
- ▶ Do such elementary transformations exist?

# What are the elementary operations?



- ▶ Exchange of rows
- ▶ Multiplying a row by a constant  $\lambda \in R \setminus \{0\}$
- ▶ Adding a row to another row
- ▶ Question  $\rightarrow$  why must any multiplier to a row be non-zero?



# Example to illustrate elementary operations



Consider the following system where we seek all solutions for some  $a \in R$ .

$$-2x_1 + 4x_2 - 2x_3 - x_4 + 4x_5 = -3$$

$$4x_1 - 8x_2 + 3x_3 - 3x_4 + x_5 = 2$$

$$x_1 - 2x_2 + x_3 - x_4 + x_5 = 0$$

$$x_1 - 2x_2 - 3x_4 + 4x_5 = a$$

Let us take the preceding equations and express them compactly in matrix form:

$$\left[ \begin{array}{ccccc|c} -2 & 4 & -2 & -1 & 4 & -3 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ 1 & -2 & 1 & -1 & 1 & 0 \\ 1 & -2 & 0 & -3 & 4 & a \end{array} \right]$$

This matrix is called the augmented matrix. It is on this matrix that we will perform the elementary row operations.

Now swap rows 1 and 3 ( $R_1 \leftrightarrow R_3$ ) in the augmented matrix to get

$$\left[ \begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ -2 & 4 & -2 & -1 & 4 & -3 \\ 1 & -2 & 0 & -3 & 4 & a \end{array} \right]$$

Does this change the system of equations? No, because we are swapping **both left and right hand sides of the equality sign**, so we are still dealing with the same set of equations.

# Subtract rows



$$\left[ \begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ -2 & 4 & -2 & -1 & 4 & -3 \\ 1 & -2 & 0 & -3 & 4 & a \end{array} \right] \quad \begin{array}{l} R_2 \leftarrow R_2 - 4R_1 \\ R_3 \leftarrow R_3 + 2R_1 \\ R_4 \leftarrow R_4 - R_1 \end{array}$$

The notation above is used to convey that we could like to add  $-4\times$  first row to the second row,  $2\times$  the first row to the third row, and  $1\times$  the first row to the fourth row to get a new augmented matrix.

# New Augmented Matrix



$$\left[ \begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -3 & 2 \\ 0 & 0 & 0 & -3 & 6 & -3 \\ 0 & 0 & -1 & -2 & 3 & a \end{array} \right] \quad R_4 \leftarrow R_4 - R_2 - R_3$$

Note that the augmented matrix shown is obtained by performing the operations shown on the previous slide. To get the next augmented matrix we subtract the second and third rows of this augmented matrix from the last row.

# New Augmented Matrix



$$\left[ \begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & -3 & 2 \\ 0 & 0 & 0 & -3 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 & a+1 \end{array} \right] \quad \begin{array}{l} R_2 \leftarrow -1 * R_2 \\ R_3 \leftarrow -1/3 * R_3 \end{array}$$

Now multiply the second row by -1 and the third row by  $-\frac{1}{3}$  to get the augmented matrix in its final form, known as the row-echelon form.

$$\left[ \begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 3 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & a+1 \end{array} \right]$$

We can revert to the set of equations represented by the augmented matrix as follows. These equations are equivalent to the original set of equations:

$$x_1 - 2x_2 + x_3 - x_4 + x_5 = 0$$

$$x_3 - x_4 + 3x_5 = -2$$

$$x_4 - 2x_5 = 1$$

$$0 = a + 1$$

# Existence of solution and particular solution



- ▶ The preceding set of equations cannot be solved when  $a \neq -1$ .
- ▶ The last equation is consistent only for  $a = -1$ .
- ▶ A particular solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$





$$x \in R^5 : x = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}, \lambda_1, \lambda_2 \in R$$

# Row-echelon form definition



- ▶ All rows that contain only zeros are at the bottom of the matrix.
- ▶ All rows that contain at least one nonzero element are on top of rows that contain only zeros.
- ▶ Considering only the non-zero rows, the first non-zero element in a given row is called the pivot and is always to the right of the pivot in the row above it.
- ▶ The positions of the pivots in the non-zero rows give rise to a staircase pattern.
- ▶ The variables corresponding to the pivot variables are called basic variables and those corresponding to the non-pivot positions correspond to the free variables.

# Finding the particular solution



- ▶ The row echelon form makes finding a particular solution easy
- ▶ Remember that the idea is that a linear combination of the pivot columns must give the right hand side.
- ▶ In the example above this means that

$$\lambda_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

- ▶ This looks like any regular linear combination for which we need to find the coefficients  $\lambda_1, \lambda_2, \lambda_3$ , so how is this really different from the original problem  $\mathbf{Ax} = \mathbf{b}$ ?

# Finding a particular solution



- ▶ The linear combination from the previous slide is easily solved.
- ▶ Start with finding the value of  $\lambda_3$ . We can see that the third equation establishes  $\lambda_3 = 1$ .
- ▶ The second equation involves only  $\lambda_2$  and  $\lambda_3$ . Plugging the just discovered value of  $\lambda_3$  into the second equation, we can find  $\lambda_2 = -1$ .
- ▶ Now we can plug the values of  $\lambda_2, \lambda_3$  into the first equation to get  $\lambda_1 = 2$



- ▶ We can convert the row-echelon form into a simpler form called the reduced row-echelon form.
- ▶ In reduced row-echelon form, every pivot is equal to 1.
- ▶ The pivot is the only non-zero entry in its column
- ▶ Therefore the pivot columns look like canonical basis vectors of  $R^m$  where the original given matrix  $A$  is a  $R^{m \times n}$  matrix.

- ▶ Consider the following matrix in reduced row-echelon form.

$$A = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}$$

- ▶ To find solutions for  $\mathbf{Ax} = \mathbf{0}$  we need to look at non-pivot columns and note that the pivot columns are "strong enough" to generate the non-pivot columns.
- ▶ Our strategy to find solutions to  $\mathbf{Ax} = \mathbf{0}$  is to find linear combinations of the pivot columns to the left of a non-pivot column to cancel out the non-pivot column, while setting all other coefficients to zero.



- ▶ Thus we note that the second column is a non-pivot column which can be expressed as a multiple of the first column such that 3 times the first column + -1 \* second column is equal to zero. This gives us our first solution.

$$\begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- ▶ Similarly we note that  $3 \times$  the first column  $+ 9 \times$  the third column  $+ -4 \times$  the fourth column  $+ -1 \times$  the fifth column is equal to zero. This gives us our second solution:

$$\begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}$$





- ▶ If  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are solutions to  $\mathbf{Ax} = \mathbf{0}$ , then any linear combination  $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$ ,  $\lambda_1, \lambda_2 \in R$  is also a solution
- ▶ Thus the general solution to the problem is

$$x \in R^5 : x = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \lambda_1, \lambda_2 \in R$$



- ▶ Consider the system of equations  $\mathbf{Ax} = \mathbf{b}$  where  $A$  is a  $n \times n$  matrix.
- ▶ If  $A$  is invertible, it means that  $A^{-1}$  exists such that  $AA^{-1} = A^{-1}A = I_n$
- ▶ In such a case the row-reduced echelon form of  $A$  is  $I_n$ , i.e every column is a pivot column where the pivot is 1.
- ▶ The process of converting  $A$  to  $I_n$  that we have discussed above is called Gaussian Elimination



- ▶ In Gaussian Elimination we use multiples of the first row to eliminate the entries in the first column below the first row.
- ▶ Then we use multiples of the second row to eliminate entries in the second column below the second row and so on until we get an upper-triangular matrix.
- ▶ This process is shown diagrammatically in the next slide.
- ▶ Then we take multiples of the last row to eliminate non-zero entries in the last column above the last entry, followed by multiples of the last but one row to eliminate non-zero entries in the last but one column and so on. This gives us a diagonal matrix.

# Gaussian elimination diagram



$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ 0 & e' & f' \\ 0 & h' & i' \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ 0 & e' & f' \\ 0 & 0 & i'' \end{bmatrix} \rightarrow \begin{bmatrix} a & 0 & c' \\ 0 & e' & f' \\ 0 & 0 & i'' \end{bmatrix}$$
  
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \leftarrow \begin{bmatrix} a & 0 & 0 \\ 0 & e' & 0 \\ 0 & 0 & i'' \end{bmatrix}$$



- ▶ Can the Gaussian elimination procedure calculate the inverse of a matrix?
- ▶ For example, let  $A$  be a  $n \times n$  matrix whose inverse  $A^{-1}$  exists. We would like to compute its inverse using Gaussian elimination. Is this possible?
- ▶ Yes we can compute the inverse in the following way: we simply set up  $n$  linear systems of the form  $\mathbf{Ax} = \mathbf{e}_i$ ,  $1 \leq i \leq n$  where  $\mathbf{e}_i$  is the  $i$ th canonical basis vector and find their solutions  $\mathbf{x}$ . Each solution vector constitutes a column in  $A^{-1}$ . Why is this true?



- ▶ Consider the linear system  $\mathbf{Ax} = \mathbf{e}_i$ .
- ▶ Gaussian elimination will convert this system to the equivalent system  $\mathbf{I}_n \mathbf{x} = \mathbf{c}_i$  whose solution is  $\mathbf{x} = \mathbf{c}_i$ .
- ▶ On the other hand, the solution to  $\mathbf{Ax} = \mathbf{e}_i$  is  $\mathbf{x} = \mathbf{A}^{-1} \mathbf{e}_i$ .
- ▶ Since the two systems are equivalent they have the same solution, so  $\mathbf{x} = \mathbf{c}_i = \mathbf{A}^{-1} \mathbf{e}_i$  which means  $\mathbf{c}_i$  is the  $i$ th column of  $\mathbf{A}^{-1}$ .
- ▶ Thus when we create the augmented matrix  $[\mathbf{A} \mathbf{e}_i]$ , Gaussian elimination will convert it into  $[\mathbf{I}_n \mathbf{c}_i]$ .
- ▶ We can solve  $n$  linear systems at once by letting the augmented matrix be  $[\mathbf{A} \mid \mathbf{I}_n]$  which will become  $[\mathbf{I}_n \mathbf{A}^{-1}]$ .