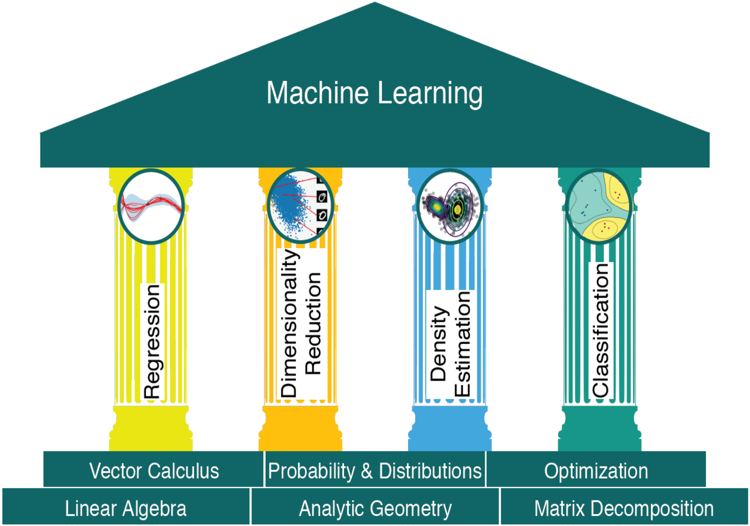
Course Overview & Linear Algebra 1/2

The foundational components of machine learning



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

The foundational components of machine learning

Each of the listed areas are relatively broad,

- Linear Algebra:
- Analytic Geometry
- Matrix Decompositions
- Probability and Distributions
- Continuous Optimization

But large components of each of the above areas are vital to machine learning and data science.

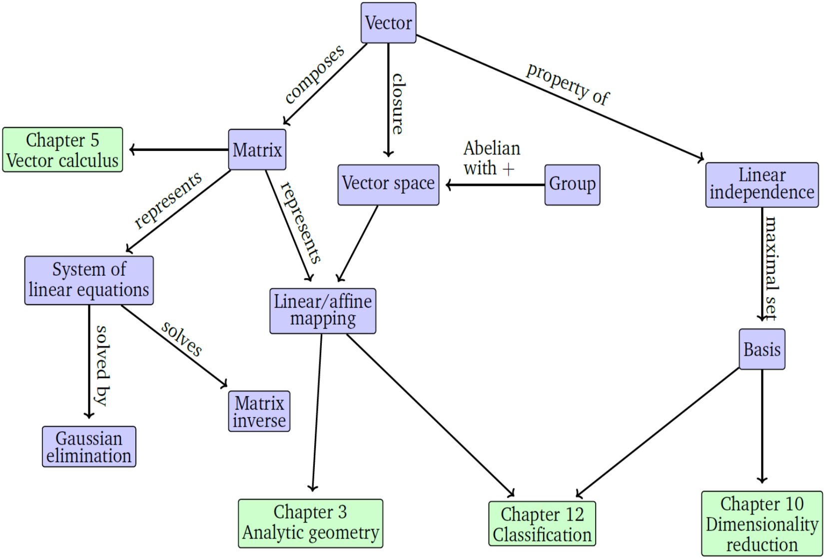
- Virtually all existing approaches use multiple areas.

Why is the underlying mathematics important to you?

In the modern era there is a plethora of tools and libraries that do the heavy lifting for you.

- However, a tool in and of itself, does not tell you
  - why a technique worked or did not;
  - what the technique actually does;
  - if a technique is likely to be effective for your problem instance;
  - the underlying assumptions that the technique uses;
- Importantly, most off the shelf approaches are not state of the art
  - If you want to truly push the boundaries you need to innovate, but without really understanding of the foundational components it is nearly impossible.
    - No foundational knowledge $\implies$ having to always work from first principles
  - Big advancements in the understanding of machine learning require rigorous mathematical analysis.

Linear algebra Mindmap

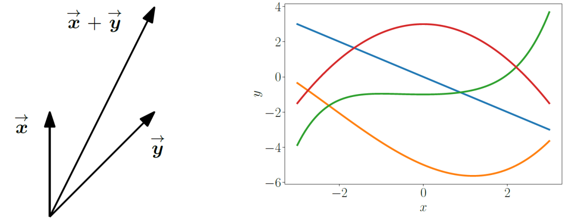


Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

The Vector

The vector is a fundamental abstraction

- The most familiar use is a geometric vector, generally in 2 and 3 dimensional space, and denoted using arrow notation $\vec{\mathbf{x}}$
- Many mathematically entities can be fully represented using a tuple of scalars. If all these elements meet criteria of a vector space, then these elements are vectors.
  - For example all polynomial can be represented as a vectors
- The most common vector we will work is a tuple of n scalars, specifically $\mathbf{x} \in \mathbb{R}^n$



(a) Geometric vectors. (b) Polynomials.

Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

Systems of Linear Equations: Simple Optimization Task

Consider the following simple problem.

- You are a factory that produces n different products: $(N_1, \cdots, N_n)$
- You have m different raw materials: $(R_1, \cdots, R_m)$
  - You have b_i of each raw materials available
  - Product N_j requires a_{ij} of raw materiel R_i to make
- The question is how many of each product should you make to, ideally, use all your raw material?

Systems of Linear Equations: Simple Optimization Task

We can write this problem as the following system

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

where x_j corresponds to the amount of product N_j.

- Can we find the tuple $(x_1, \ldots, x_n) \in \mathbb{R}^n$ to solve this?
- Is there one unique solution or multiple?
- If there are no solutions, how "close" can I get to solving the system?

Systems of Linear Equations: Simple Optimization Task

In principle we can try directly answer, at least the first two questions, using simple algebra.

- Using a matrix approach is easier and more scalable (particularly to automation)
- See Example 2.2

Systems of Linear Equations: Simple Optimization Task

The system we have set up can be rewritten as

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, and $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$

- What is great is $\mathbf{A}$ in and of itself holds the answer to our previous questions!

Matrix Multiplication Reminder

The compact representation relies on matrix multiplication, specifically

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^{n} a_{1l} x_l \\ \vdots \\ \sum_{l=1}^{n} a_{ml} x_l \end{bmatrix}$$

where $b_i = \sum_{l=1}^{n} a_{il} x_l$ for $i = 1, \cdots, m$

$$\begin{bmatrix} | & | & | \\ a_1 & a_2 & a_3 \\ | & | & | \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{matrix} | \\ a_1 \\ | \end{matrix} x_1 + \begin{matrix} | \\ a_2 \\ | \end{matrix} x_2 + \begin{matrix} | \\ a_3 \\ | \end{matrix} x_3 = \sum_{i=1}^{3} \begin{matrix} | \\ a_i \\ | \end{matrix} x_i$$

Matrix Operator Reminder

Addition

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times n}$ then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Multiplication

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times k}$ then

$$\mathbf{AB} = \mathbf{C} \in \mathbb{R}^{m \times k}$$

where $\boxed{c_{i,j} = \sum_{l=1}^{n} a_{il} b_{lj}}$

Matrix Operator Reminder

It is important to note that not all matrices can be multiplied together, specifically

- Matrices can only be multiplied if their "neighboring" dimensions match. For instance, an $n \times k$-matrix $\mathbf{A}$ can be multiplied with a $k \times m$-matrix $\mathbf{B}$, but only from the left side:

$$\underbrace{\mathbf{A}}_{n \times k} \underbrace{\mathbf{B}}_{k \times m} = \underbrace{\mathbf{C}}_{n \times m}$$

- Why would $\mathbf{BA}$ not be defined?

Note: matrix multiplication is **not** commutative, but it is associative and distributive. Additionally we do have a left and right multiplicative identity (which is the same when we have a square matrix).

Note: Cij = AijBij is known as the hadamard product (not matrix multiplication)

Inverse and Transpose

Inverse

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. The matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$, that has the property

$$\mathbf{AB} = \mathbf{I}_n = \mathbf{BA}, \tag{2}$$

is called the *inverse* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

As we alluded to in the first system example not every matrix has an inverse

- If an inverse **exists** we call $\mathbf{A}$ *regular*, *invertible*, or *nonsingular*.
- If an inverse **does not exists** we call $\mathbf{A}$ *singular* or *noninvertible*.

Inverse and Transpose

Transpose

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$. The matrix $\mathbf{B} \in \mathbb{R}^{n \times m}$ is called the transpose of $\mathbf{A}$ if $b_{ij} = a_{ji}$. The transpose is denoted as $\mathbf{A}^T$.

Example:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}^T = \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}$$

Symmetric Matrix

The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *symmetric* if $\mathbf{A}^T = \mathbf{A}$.

Inverse and Transpose

Basics properties

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n = \mathbf{A}^{-1}\mathbf{A} \tag{3}$$

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{4}$$

$$(\mathbf{A}+\mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1} \tag{5}$$

$$(\mathbf{A}^T)^T = \mathbf{A} \tag{6}$$

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \tag{7}$$

$$(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T\mathbf{A}^T \tag{8}$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1} = \mathbf{A}^{-T} \tag{9}$$

Multiplication by a Scalar

It is worth being explicit with how scalars interact with matrices. Specifically, if $\lambda \in \mathbb{R}$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$ then

$$\lambda \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \lambda$$

This gives us a number of useful properties for $\lambda, \psi \in \mathbb{R}$

$$(\lambda\psi)\mathbf{C} = \lambda(\psi\mathbf{C}), \qquad \mathbf{C} \in \mathbb{R}^{m\times n} \tag{10}$$

$$\lambda(\mathbf{BC}) = (\lambda\mathbf{B})\mathbf{C} = \mathbf{B}(\lambda\mathbf{C}) = (\mathbf{BC})\lambda, \qquad \mathbf{B} \in \mathbb{R}^{m\times n}, \mathbf{C} \in \mathbb{R}^{n\times k} \tag{11}$$

$$(\lambda\mathbf{C})^T = \mathbf{C}^T\lambda^T = \mathbf{C}^T\lambda = \lambda\mathbf{C}^T, \qquad \mathbf{C} \in \mathbb{R}^{m\times n} \tag{12}$$

$$(\lambda+\psi)\mathbf{C} = \lambda\mathbf{C} + \psi\mathbf{C}, \qquad \mathbf{C} \in \mathbb{R}^{m\times n} \tag{13}$$

$$\lambda(\mathbf{B}+\mathbf{C}) = \lambda\mathbf{B} + \lambda\mathbf{C}, \qquad \mathbf{B}, \mathbf{C} \in \mathbb{R}^{m\times n} \tag{14}$$

Particular and General Solution

Before we describe a more mechanistic approach to solving system of linear equation,let us build a bit more intuition. Consider

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

- Do you think there will be a unique solution?

Particular and General Solution

Before we describe a more mechanistic approach to solving system of linear equation,let us build a bit more intuition. Consider

$$\begin{array}{c} \begin{bmatrix} 1 & 0 & 8 & 4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \\ \begin{array}{cccc} c_1 & c_2 & c_3 & c_4 \end{array} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

- Do you think there will be a unique solution?
- Simply the fact that there are 4 unknowns and only 2 equations tells us we will, generally, have infinity many solutions. (Expect when?)

$$\begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \quad c_1 x_1 + c_2 x_2 + c_3 x_3$$

Particular and General Solution

We can get at least one possible solution by considering the underlying problem:

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 4 \\ 12 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix} \tag{15}$$

If we only use x_1, x_2 we can get a **particular** solution be setting them to 42 and 8 respectively while setting $x_3 = x_4 = 0$. Namely $[42, 8, 0, 0]^T$.

- Is this actually the only possible **particular** solution?

Particular and General Solution

We can get at least one possible solution by considering the underlying problem:

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 4 \\ 12 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix} \tag{15}$$

If we only use x_1, x_2 we can get a **particular** solution be setting them to 42 and 8 respectively while setting $x_3 = x_4 = 0$. Namely $[42, 8, 0, 0]^T$.

- Is this actually the only possible **particular** solution?
    - $[0, -2\frac{1}{2}, 5\frac{1}{4}, 0]^T$.
- We could use either of these to derive the general solution, but the first is easier to use.

Particular and General Solution

In order to obtain the general solution we need to engineer $\mathbf{0}$ solutions using the previously unused columns.

- This way if we add this newly constructed solution to the found particular solution we still actually have a solution to original equation (15)
- Let's start by building the 3rd column vector, $\mathbf{c}_3$, from $\mathbf{c_1}$ and $\mathbf{c_2}$. Specifically:

$$\begin{bmatrix} 8 \\ 2 \end{bmatrix} = 8 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{16}$$

it follows that $\mathbf{0} = 8 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 8 \\ 2 \end{bmatrix}$. This correspond to the vector $(x_1, x_2, x_3, x_4) = (8, 2, -1, 0)$ .Furthermore for any $\lambda_1 \in \mathbb{R}$ scalar applied to the vector yields $\mathbf{0}$

Particular and General Solution

- We can follow the same logic to build c_4, namely:

$$\begin{bmatrix}-4\\12\end{bmatrix} = -4\begin{bmatrix}1\\0\end{bmatrix} + 12\begin{bmatrix}0\\1\end{bmatrix} \tag{17}$$

  Which means that $(-4, 12, 0, -1)$ also produces $\mathbf{0}$. Furthermore for any $\lambda_2 \in \mathbb{R}$ scalar applied to the vector yields $\mathbf{0}$

- Putting together our engineered **0**s with our **particular solution** we obtain the **general solution** as the set

$$\left\{ \mathbf{x} \in \mathbb{R}^4 \;:\; \mathbf{x} = \begin{bmatrix}42\\8\\0\\0\end{bmatrix} + \lambda_1\begin{bmatrix}8\\2\\-1\\0\end{bmatrix} + \lambda_2\begin{bmatrix}-4\\12\\0\\-1\end{bmatrix}\;,\; \lambda_1, \lambda_2 \in \mathbb{R} \right\} \tag{18}$$

Particular and General Solution

The general approach is as follows

1. Find a particular solution to $\mathbf{Ax} = \mathbf{b}$.
2. Find all solutions to $\mathbf{Ax} = \mathbf{0}$.
3. Combine the solutions from steps 1. and 2. to create the general solution.

Homework:

- Repeat the process using the other particular solution mentioned

Solving Linear System

It is at this point worth introducing an even more convenient representation for a system of linear equations.

- Namely the *augmented matrix* form of $\mathbf{Ax} = \mathbf{b}$, as $[\mathbf{A}|\mathbf{b}]$.
- For example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix},$$

  becomes

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 3 & 4 & 6 \end{array}\right].$$

Solving Linear System

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 3 & 4 & 6 \end{array}\right].$$

In this form it becomes easy to use *elementary transformations* to change the matrix into **row-echelon form** followed by **reduced row-echelon form**.

- By *elementary transformations* we mean
  - Exchange of two equations (rows in the matrix representing the system of equations)
  - Multiplication of an equation (row) with a non-zero real valued scalar
  - Addition of two equations (rows)

Solving Linear System

Row-Echelon Form

A matrix is in *row-echelon form* if

- All rows that contain only zeros are at the bottom of the matrix
- Looking at nonzero rows only, the first nonzero number from the left (also called the pivot or the leading coefficient) is always strictly to the right of the pivot of the row above it.

For example:

$$\begin{bmatrix} 1 & 2 & -1 & 5 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Solving Linear System

Reduced Row Echelon Form

A matrix is in *reduced row-echelon form* if

- It is in row-echelon form.
- Every pivot is 1.
- The pivot is the only nonzero entry in its column.

For example:

$$\begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Solving Linear System

Now we can transform our system in the *augmented matrix* form, $[\mathbf{A}|\mathbf{b}]$ into **reduced row echelon form** using the elementary transformations.

1. If $\mathbf{A}$ was invertible and we applied the elementary transforms to $\mathbf{b}$ as we did the transformation the resulting right hand side of the augmented matrix is in fact our solution. $\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 1 \end{array}\right]$
2. If $\mathbf{A}$ was not invertible we either arrive at a contradiction (which means there are no solutions) or we are in a good position now find a particular and the general solution to the system.

We will consider and example of both cases.

- It is worth mentioning that generally on a PC the transformation is done using an algorithmic approach called Gaussian Elimination (Or rather a more numerically stable version of it).

Solving Linear System: Case 1

$$\left[\begin{array}{ccc|c} 2 & 5 & -1 & 15 \\ 1 & 0 & -5 & -2 \\ -3 & 8 & 1 & 8 \end{array}\right] \quad \text{swap } \mathbf{r}_1 \text{ with } \mathbf{r}_2$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 2 & 5 & -1 & 15 \\ -3 & 8 & 1 & 8 \end{array}\right] \begin{array}{l} \\ -2\mathbf{r}_1 \\ +3\mathbf{r}_1 \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 5 & 9 & 19 \\ 0 & 8 & -14 & 2 \end{array}\right] \begin{array}{l} \\ \leftrightarrow \frac{\mathbf{r}_2}{5} \text{ (Step 2)} \\ -\frac{8\mathbf{r}_2}{5} \text{ (Step 1)} \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 1 & \frac{9}{5} & \frac{19}{5} \\ 0 & 0 & -\frac{142}{5} & -\frac{142}{5} \end{array}\right] \begin{array}{l} \\ \\ \times -\frac{5}{142} \end{array}$$

Solving Linear System:
Case 1

$$[A|b] \rightsquigarrow A^{-1}[A|b] \rightsquigarrow [I|A^{-1}b]$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 1 & \frac{9}{5} & \frac{19}{5} \\ 0 & 0 & 1 & 1 \end{array}\right] \begin{array}{l} +5\mathbf{r}_3 \\ -\frac{9}{5}\mathbf{r}_3 \\ \end{array}$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

since the left side is $\mathbf{I}$, $(3, 2, 1)^T$ is the unique solution to the linear system.

Solving Linear System: Case 2

Now consider the new system

$$\left[\begin{array}{ccc|c} 2 & 0 & -10 & -4 \\ 1 & 0 & -5 & -2 \\ -3 & 8 & 1 & 8 \end{array}\right]$$

$$\rightsquigarrow \left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 1 & -\frac{7}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \end{array}\right] \quad \text{verify this at home}$$

Since we have the left side is not $\mathbf{I}$ and we have not encountered a contradiction, we have infinitely many solutions. Obtaining the particular and general solution can be done as before, but the reduced row-echelon form makes this easy.

Solving Linear System: Case 2, particular solution

$$\left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 1 & -\frac{7}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

To obtain a particular solution is we express the right-hand side of the equation system using the pivot columns, such that $\mathbf{b} = \sum_{i=1}^{P} \lambda_i \mathbf{p}_i$ where $\mathbf{p}_i$, $i = 1, \ldots, P$ are the pivot columns.

- This can be done in row-echelon form
- Much easier in reduced row-echelon form.

For the current system $\lambda_1 = -2$ and $\lambda_2 = \frac{1}{4}$, leading to a particular solution $(-2, \frac{1}{4}, 0)^T$.

Minus 1 trick

Obtaining the general solution can be done using the following practical trick.

- Assume $\mathbf{A} \in \mathbb{R}^{k \times n}, k < n$ and is in REF.
- Assume there are no rows with only zero elements.
- We extend this matrix to an $n \times n$ matrix $\hat{\mathbf{A}}$ by adding $n - k$ rows of the form

$$\begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{bmatrix}$$

  so that the diagonal of the augmented matrix $\hat{\mathbf{A}}$ contains either 1 or -1.
- Then, the **columns** of $\hat{\mathbf{A}}$ that contain the -1 as pivots are solutions of the homogeneous equation system $\mathbf{Ax} = 0$.

Solving Linear System: Case 2

Moving back to our system

$$\left[\begin{array}{ccc|c} 1 & 0 & -5 & -2 \\ 0 & 1 & -\frac{7}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \end{array}\right]$$

We use the minus-1 trick using the matrix

$$\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & -\frac{7}{4} \end{bmatrix} \text{ which we augmented to } \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & -\frac{7}{4} \\ 0 & 0 & -1 \end{bmatrix}$$

Then our general solution is just

$$\left\{ \mathbf{x} \in \mathbb{R}^3 \ : \ \mathbf{x} = \begin{bmatrix} -2 \\ \frac{1}{4} \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} -5 \\ -\frac{7}{4} \\ -1 \end{bmatrix}, \ \lambda_1 \in \mathbb{R} \right\}$$

Calculating the Inverse

In order to find the inverse of a matrix we can use an extension of the approach used for linear systems.

- Let $\mathbf{A} \in \mathbb{R}^{n\times n}$
- The inverse, $\mathbf{A}^{-1}$, is the matrix, $\mathbf{X}$ that satisfies $\mathbf{AX} = \mathbf{I}$.
- This is fundamentally a set of systems of linear equations where $\mathbf{X} = [\mathbf{x}_1|\cdots|\mathbf{x}_n]$.
- We use the augmented matrix notation for a compact representation of this set of systems of linear equations and obtain

$$[\mathbf{A}|\mathbf{I}] \rightsquigarrow [\mathbf{I}|\mathbf{A}^{-1}]$$

Approximate Solution

What do we do when our system, $\mathbf{Ax} = \mathbf{b}$, does not have any exact solutions?

- We can still under mild assumptions (all $\mathbf{A}$ columns are linearly independent) find a $\mathbf{x}$ 'close' to solving the system.
- Namely we use the following transformation

$$\mathbf{Ax} = \mathbf{b} \iff \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b} \iff \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

where $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the *Moore-Penrose pseudo-inverse* and $\mathbf{x}$ results in the smallest $\|\mathbf{Ax} - \mathbf{b}\|_2$

$$Ax = b \Rightarrow \underbrace{(A^TA)^{-1}A^TA}_{(A^TA)^{-1}(A^TA)=I}x = (A^TA)^{-1}Ab \quad \therefore x = (A^TA)^{-1}Ab$$

Getting Abstract: Groups

Group

Let $\mathcal{G}$ be a set and $\otimes$ the operator $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is a *group* if the following hold:

1. Closure of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. Associativity: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. Neutral element: $\exists e \in \mathcal{G}\ \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. Inverse element: $\forall x \in \mathcal{G}\ \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$
    - y is often denoted x^{-1}

Getting Abstract: Groups

Group

Let $\mathcal{G}$ be a set and $\otimes$ the operator $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$. Then $G := (\mathcal{G}, \otimes)$ is a *group* if the following hold:

1. Closure of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. Associativity: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. Neutral element: $\exists e \in \mathcal{G}\ \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. Inverse element: $\forall x \in \mathcal{G}\ \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$
    - y is often denoted x^{-1}

Abelian Group

The group $G = (\mathcal{G}, \otimes)$ is an *Abelian Group* if the following holds:

- Commutativity: $\forall x, y \in \mathcal{G} : x \otimes y = y \otimes x$

Getting Abstract: Groups

Some examples:

- $(\mathbb{Z}, +)$ is an Abelian group.
- $(\mathbb{N} \cup \{0\}, +)$ is not a group.
  - Why?
- $(\mathbb{R}^{n \times n}, \cdot)$ where $\cdot$ is normal matrix multiplication as defined earlier is **not** an group
  - Why?

Getting Abstract: Groups

Some examples:

- $(\mathbb{Z}, +)$ is an Abelian group.
- $(\mathbb{N} \cup \{0\}, +)$ is not a group.
  - Why?
- $(\mathbb{R}^{n\times n}, \cdot)$ where $\cdot$ is normal matrix multiplication as defined earlier is **not** an group
  - Why?
  - The set of all inevitable/regular matrices, with normal matrix multiplication is a Group.
  - Often refereed to as the *General Linear Group*, and denoted as $GL(n, \mathbb{R})$
  - General Linear Group it is not an *Abelian* group.
    - Why?

Getting Abstract: Vector Spaces

Vector Space

A real-valued *vector space* $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$$
\begin{aligned}
+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V} \quad &\text{(Inner operator)} \\
\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V} \quad &\text{(Outer operator)}
\end{aligned}
$$

where

1. $(\mathcal{V}, +)$ is an Abelian group + closure w.r.t. scalar product (outer opt)
2. Distributivity:
   1. $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V}: \ \lambda \cdot (\mathbf{x} + \mathbf{y}) = \lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}$
   2. $\forall \lambda, \psi \in \mathbb{R}: \ (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$
3. Associativity (outer operation):
   $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda \cdot \psi) \cdot \mathbf{x}$
4. Neutral element with respect to the outer operation:
   $\forall \mathbf{x} \in \mathcal{V}: \ 1 \cdot \mathbf{x} = \mathbf{x}$

Some examples:

- $\mathcal{V} = \mathbb{R}^n$ with vector addition and multiplication by scalars, is the most commonly encountered vector space.
- $\mathcal{V} = \mathbb{R}^{m \times n}, m, n \in \mathbb{N}$ is a vector space with
  - Elementwise addition '+': $\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$ for all $\mathbf{A}, \mathbf{B} \in \mathcal{V}$
  - Multiplication by scalars '$\cdot$': $\lambda \cdot \mathbf{A} = \begin{bmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{bmatrix}$ for all $\mathbf{A} \in \mathcal{V}, \lambda \in \mathbb{R}$
- We will also consider a $\mathcal{V}$ that has functions as elements later in the course