

Linear Algebra

Vectors

Interpretations of Vectors	3
Vector Addition and Subtraction	3
Vector Scalar Multiplication.....	4
The Dot Product	4
Properties of the Dot Product.....	4
Vector Length.....	5
Geometric Interpretation of the Dot Product	5
Other Properties of Vectors	6
Vector Hadamard Multiplication	6
Outer Product.....	6
Cross Product	7
Primer on Complex Numbers.....	7
Conjugate Transpose	8
Unit Vectors	8
Linear Combinations	9
Subspace.....	9
Subsets.....	9
Span.....	10
Linear Independence	10
Basis	11

Matrices

Matrix Terminology	12
Square and Rectangular Matrices	13
Symmetric and Skew-Symmetric Matrices	13
Identity and Zero Matrices	13
Diagonal and Triangular Matrices.....	14
Augmented and Complex Matrices	14
Basic Matrix Operations	15
Matrix Addition and Subtraction	15
Matrix Scalar Multiplication.....	15
Transposition.....	15
Diagonal and Trace	16
Broadcasting	16

Matrix Multiplication

Standard Matrix Multiplication	17
Standard Matrix Multiplication Perspectives	17
Various Notes on Matrix Multiplication	19
Diagonal Matrix Multiplication	19

Order of Operations	19
Matrix Vector Multiplication	19
Additive and Multiplicative Matrices	20
Creating Symmetric Matrices	20
Hadamard Multiplication	21
Matrix Norms	22
Frobenius Norm	22
Induced p-Norm	23
Schatten p-Norm	23
Matrix Rank	
Rank Terminology	24
Maximum Rank	24
Computing Rank	24
Rank After Addition and Multiplication	24
Rank of $A^T A$ and AA^T	25
Full Rank via “Shifting”	25
Matrix Spaces	
Systems of Linear Equations	
Linear Equations	27
Solutions	27
Elementary Operations	28
Gaussian Elimination	29
Row Echelon form	29
The Gaussian Algorithm	30
Homogeneous Equations	31
Homogeneous Solution Sets	31
The Determinant	
Matrix Inverse	
—To Be Defined—	

Vectors

Interpretations of Vectors

- **Algebraic vectors** (\mathbf{v} , \vec{v}): an ordered list of numbers, e.g.,

$$\mathbf{v} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

- Vectors can be written as **rows** (seen above) or **columns** (seen below), but differ only at the level of notation, some notation is more useful with certain conventions.
- The order of elements in a vector matters:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

- **Dimensionality**: the number of elements in a vector, where each new element provides new information, or geometrically, a new direction.
- **Euclidean (geometric, spatial) vectors**: a line in geometric space that indicates the **magnitude** and **direction** from a starting point (tail) to an end point (head).
 - Geometric vectors can start at any point in space, but often represented as starting from the **origin**—such vectors are in **standard position**.
 - Coordinates are not the same as vectors, but they do indicate where the head of a vector will land if it is in standard position.

Vector Addition and Subtraction

- Algebraically, **dimensionality** of vectors **must be equal**. When they are, then addition or subtraction vectors is done on the corresponding elements of each vector, e.g.,

$$\begin{bmatrix} 1 \\ 0 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ -6 \\ 11 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ -2 \\ 16 \end{bmatrix}$$

- Geometrically, addition can be thought of translating the tail of one vector to the head of the other—resulting in a new vector.
- Geometric interpretations of subtraction can be thought of in two ways:
 1. Multiplying one vector by -1, then applying vector addition method above.
 2. Placing both vectors in standard position, with the resulting vector between the two heads being the answer.

Vector Scalar Multiplication

- **Scalar**: typically denoted with lower case Greek letters (e.g., α , λ) indicating an element of a field (typically real numbers) used in scalar multiplication of vectors.
- Algebraically, scalar multiplication is the multiplication of each element of a vector by a particular scalar, e.g.,

$$\lambda \mathbf{v} \rightarrow 7 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -7 \\ 0 \\ 7 \end{bmatrix}$$

- Geometrically, scalar-multiplication can be thought of as the **extension** ($\lambda > 1$) or **compression** ($\lambda \in (0, 1)$) of a vector.
 - When $\lambda < 0$, then it can be thought of inverting its direction with respect to the origin.

The Dot Product

- **Dot (scalar, inner) product**: an algebraic operation that takes two **equal-length** sequences of numbers (usually coordinate vectors), and returns a **single number**.
 - The result of a dot product is a **scalar**[↑], so often it is represented as such.
 - It can also be represented as multiplication between two vectors ($\mathbf{a} \cdot \mathbf{b}$).
 - However, it is commonly represented as $\mathbf{a}^T \mathbf{b}$ – **transpose**[↓] will be explained in more detail when dealing with **matrix multiplication**[↓].
 - Algebraically: $\sum_{i=1}^n \mathbf{a}_i \mathbf{b}_i$ – where Σ denotes summation and n is the dimension of the vector space, e.g.,

$$\begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = (1 \cdot 4) + (3 \cdot -2) + (-5 \cdot -1) = 3$$

Properties of the Dot Product

- Note: the following properties hold as long as \mathbf{a} , \mathbf{b} , and \mathbf{c} are real vectors.
- ✓ **Distributive**: $\mathbf{a}^T (\mathbf{b} + \mathbf{c}) = \mathbf{a}^T \mathbf{b} + \mathbf{a}^T \mathbf{c}$ – vector multiplication distributes over vector addition.
- ✗ **Associative**: $\mathbf{a}^T (\mathbf{b}^T \mathbf{c}) \neq (\mathbf{a}^T \mathbf{b}) \mathbf{c}$ – in general the associative property does not hold, as the dot product would most likely produce different scalars.
 - Additionally, \mathbf{a} could have a different dimensionality than \mathbf{b} and \mathbf{c} . I.e., even if \mathbf{b} and \mathbf{c} had the same dimensionality ($\mathbf{a}^T (\mathbf{b}^T \mathbf{c})$ would be valid vector scalar multiplication) then $\mathbf{a}^T \mathbf{b}$ would be invalid.
- ✓ **Commutative**: $\mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$ – the order of the vectors does not matter.

Vector Length

- **Vector norm (magnitude, length):** denoted with double vertical bars $\|\mathbf{v}\|$, indicating length of a vector in euclidean space. Not to be confused with absolute value $|x|$ of a scalar's "norm." However, sometimes the notation $|\mathbf{v}|$ is used.
- Calculating $\|\mathbf{v}\|$ is done using the Euclidean norm:

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + v_3^2}$$

- This is a consequence of the Pythagorean theorem, since the **basis vectors** \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 are **orthogonal** **unit vectors**.
- Thus, the **norm** can easily be found by taking the square root of the dot product of the vector with itself:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \mathbf{v}}$$

Geometric Interpretation of the Dot Product

- The dot product of two **Euclidean vectors** \mathbf{a} and \mathbf{b} is defined by:

$$\lambda = \mathbf{a}^T \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos(\theta)$$

where θ is the angle between \mathbf{a} and \mathbf{b} .

- **Features based on θ :**
 - When $\cos \theta > 0$ ($\theta < 90^\circ$ – **acute**) then $\lambda > 0$ (+)
 - When $\cos \theta < 0$ ($\theta > 90^\circ$ – **obtuse**) then $\lambda < 0$ (–)
 - When $\cos \theta = 0$ ($\theta = 90^\circ$ – perpendicular) then $\lambda = 0$
 - This represents a special case where the vectors are said to be **orthogonal**.
 - **Orthogonality:** the generalization of the notion of **perpendicularity** to the linear algebra of bilinear forms.
 - When $\cos \theta = 1$ then the vectors are **codirectional**:

$$\mathbf{a}^T \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\|$$

- Thus, the dot product with a vector \mathbf{v} with itself is

$$\mathbf{v}^T \mathbf{v} = \|\mathbf{v}\|^2$$

- Which gives us the **norm** as defined above, i.e., $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \mathbf{v}}$
- If $\cos \theta = -1$, then really vectors are still codirectional, but point in opposite directions with respect to the origin.

Other Properties of Vectors

- **Matrices**↓ are properly defined later. However, vectors are technically single row or column matrices, so many of the following operations also work on vectors, thus use of matrices often appears in this section.

Vector Hadamard Multiplication

- **Hadamard (element-wise) product**: a binary operation (only takes two operands) that matrices of the same dimensions and produces another matrix of the same dimension as the operands, e.g., vector Hadamard multiplication:

$$\begin{bmatrix} 1 \\ 0 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ -6 \\ 11 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -24 \\ 55 \end{bmatrix}$$

Outer Product

- Recall that the **dot (scalar, inner) product**↑ produces a 1×1 matrix, or rather a scalar, hence “scalar” product.
 - Note: the typical notation used for dot products is $\mathbf{v}^T \mathbf{w}$ — part of reasoning for the “inner” product.
- **Outer product**: an $n \times m$ matrix that results from the product of two vectors with dimensions n and m .

$$\mathbf{v} \mathbf{w}^T = n \times m$$

- The subtle change in notation matters in contrast to the dot product, as both represent distinct operations (assuming they are column vectors).
- The outer product allows for the multiplication of vectors with different dimensionality
- Can be thought of in two different ways:
 - The **row** perspective:

$$\begin{bmatrix} 1 \\ 0 \\ 4 \\ 2 \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} 1a & 1b & 1c \\ 0a & 0b & 0c \\ 4a & 4b & 4c \\ 2a & 2b & 2c \end{bmatrix}$$

- The **column** perspective:

$$\begin{bmatrix} 1 \\ 0 \\ 4 \\ 2 \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} a1 & b1 & c1 \\ a0 & b0 & c0 \\ a4 & b4 & c4 \\ a2 & b2 & c2 \end{bmatrix}$$

Cross Product

- **Cross (vector, directed area) product:** denoted by the symbol \times , indicating a binary operation on two vectors in **three-dimensional space** \mathbb{R}^3 .
 - Given two **linearly independent**[↓] vectors \mathbf{a} and \mathbf{b} , then the cross product $\mathbf{a} \times \mathbf{b}$ produces a new vector that is **orthogonal** to both \mathbf{a} and \mathbf{b} , or **normal** to the plane containing them.
 - The **direction** of the vector is given by the **right-hand rule** (\mathbf{a} = pointer, \mathbf{b} = index, thumb = direction).
 - The **magnitude** of the vector represents the **area of the parallelogram** that the vectors span.
- The cross product can be defined by the formula:

$$\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \sin(\theta) \mathbf{v}$$

- Notice this is similar to the **geometric interpretations of the dot product**[↑] – the contrast between the two leads to an intuitive interpretation:
 - $\cos(\theta)$ in the **dot product** is used to measure how “parallel” the two vectors are, i.e., they are **codirectional** when $\theta = 1$, allowing for calculation of the **norm**[↑].
 - $\sin(\theta)$ in the **cross product** is used to measure how “perpendicular” two vectors are, i.e., they are **orthogonal** when $\theta = 1$. There are multiple directions of the orthogonal vector, so calculation of the signed area returns vector \mathbf{v} that describes both magnitude and direction as described above.
 - The intuition described here will be more clear when the **determinant**[↓] is discussed in more detail.
- An algebraic example of vector e.g.,

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2c - 3b \\ 3a - 1c \\ 1b - 2a \end{bmatrix}$$

Primer on Complex Numbers

- **Complex number:** a number that can be expressed in the form $a + bi$, where a and b are **real numbers**, and i (j) is a symbol called the **imaginary unit** that satisfies the equation $i = \sqrt{-1}$.
- This imaginary unit allows for the **imaginary set** ($a + bi \in \mathbb{C}$) to be described.
- The combination of a real part and an imaginary part ($a + bi$) gives imaginary numbers both a direction and magnitude on the 2-D plane created between the two parts—thus they can be described as 2-D vectors in this space.

- Multiplication of complex numbers can be done by factoring the imaginary unit, e.g.,

$$z = a + bj$$

$$w = c + dj$$

$$\begin{aligned} zw &= (a + bj)(c + dj) \\ &= ac + adj + cbj + bdj^2 \\ &= ac + adj + cbj - bd \quad (j^2 = -1) \end{aligned}$$

- Computing the dot product with complex vectors is the same, just including the factoring mentioned above when necessary, e.g.,

$$\begin{aligned} &\begin{bmatrix} 1 + 3j \\ -2j \\ 4 \\ 5 \end{bmatrix}^T \begin{bmatrix} 6 + 2j \\ 8 \\ 3j \\ -5 \end{bmatrix} \\ &= (1 + 3j)(6 + 2j) + -16j + 12j + 25 \\ &= 6 + 2j + 18j - 6 - 16j + 12j + 25 \\ &= 25 + 16j \end{aligned}$$

Conjugate Transpose

- **Conjugate (Hermitian) transpose** M^H , M^* : the n -by- m matrix obtained by taking the **transpose**[↓] and then taking the complex conjugate of each entry.
- **Complex conjugate**: the number with both equal real and imaginary parts equal in magnitude but **opposite in sign**.

$$a + bi = a - bi \iff a - bi = a + bi$$

- Multiplying a vector by the conjugate transpose allows us to calculate the **magnitude** of the vector containing imaginary numbers by using the complex conjugate to “canceling out” all the imaginary units and give a **real number answer** (or rather, $a + 0i$).
- Matrices will return a matrix of real numbers—this is part of what makes using imaginary numbers useful for future computations.

Unit Vectors

- **Unit vectors** $\hat{i}, \hat{j}, \hat{k}$: a vector scaled such that $(:)$ the length of the vector **equals 1** in a normed vector space.

$$\hat{i} = \lambda \mathbf{v} : \|\lambda \mathbf{v}\| = \frac{1}{\|\mathbf{v}\|} \|\mathbf{v}\| = 1$$

- **Normed vector space**: a vector space, over the real or complex numbers, on which a **norm**[↑] is defined.

Linear Combinations

- **Linear combination:** an expression constructed from a set of terms by multiplying each term by a scalar and adding the results, i.e.,

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + a_n \mathbf{v}_n$$

- For example, in a 3-D vector space \mathbb{R}^3 , then **any vector** in the space is can be made by a linear combination of the following three vectors \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 (unit vectors) multiplied by some scalar:

$$\lambda \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Subspace

- **Linear (vector) subspace:** a vector space that is a subset of some larger vector space.
- **Algebraically**, a subspace is the set of all vectors that can be created by taking a linear combination of some vector or a set of vectors.
- A vector subspace must be **closed under addition** and **scalar multiplication** while also **containing the zero vector**, i.e.,

$$\forall \mathbf{v}, \mathbf{w} \in L; \quad \forall \lambda, \alpha \in \mathbb{R}; \quad \lambda \mathbf{v} + \alpha \mathbf{w} \in L$$

- For all (\forall) vectors \mathbf{v} , \mathbf{w} in (\in) the linear subspace (V), and for all scalars λ , α in the set of real numbers (\mathbb{R}), then any linear combination of the two are in the same subspace.
- The above equation also implies the inclusion of zero vector, which is a trivial subspace of the vector space.
- The **geometric** interpretation is best described through some examples:
 - A \mathbb{R}^1 vector and all the scaled possibilities added together describes a line stretches infinity in both directions, while a \mathbb{R}^2 vector would create a 2-D plane.
 - Both of the previous subspaces exist in the higher dimensional vector spaces, i.e., a 2-D plane and a 1-D line both exist in the 3-D vector space.
 - All subspaces also pass through the origin, including the 0-D subspace, which is just the origin.

Subsets

- **Subset** \subseteq : a set A is a subset of a set B if all elements of A are also elements of B ; B is then a **superset** \supseteq of A .
- Not all subsets of vector spaces are subspaces; subsets don't need to include the origin, don't need to be closed, or could have boundaries.

Span

- **Linear span (hull)** $\text{span}(S)$: a set S of vectors (from a vector space) that is the smallest linear subspace that contains the set.
- The span is typically infinite, but it also can be defined as the set of all finite **linear combinations**[↑] of vectors of S given an arbitrary field K :

$$\text{span}(S) = \left\{ \sum_{i=1}^k \lambda_i \mathbf{v}_i \mid k \in \mathbb{N}, \mathbf{v}_i \in S, \lambda_i \in K \right\}$$

- A frequent question is asked whether a vector is in a span or not, e.g., are vectors $\mathbf{v}, \mathbf{w} \in \text{span}(S)$:

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \quad S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix} \right\}$$

- $\mathbf{v} \in \text{span}(S)$ since both vectors in S can be scaled then combined in some way to form \mathbf{v} , i.e.,

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \frac{5}{6} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{6} \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix}$$

- Determining the weights will be discussed later using **matrix computations**[↓].
- However, it's clear that $\mathbf{w} \notin \text{span}(S)$ since 0 cannot be scaled to equal 1.
- Geometrically, a useful intuitive example is a 2-D span in 3-D space; the 2-D plane can be moved anywhere in the 3-D space as long as the two vectors describing the 2-D plane are **linearly independent**[↓].

Linear Independence

- **Linearly dependent**: a set of V vectors where at least one vector in the set can be defined as a **linear combination**[↑] of the others.
- **Linearly independent**: when no vector in the set can be written in the above way.
- Algebraically, a sequence of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ that a vector space V are linearly independent if there exist scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ (not all zero) such that they form the zero vector $\mathbf{0}$, i.e.,

$$\lambda_1 \mathbf{v}_1, \lambda_2 \mathbf{v}_2, \dots, \lambda_k \mathbf{v}_k = \mathbf{0} \quad \lambda \in \mathbb{R}$$

- Geometrically, a set of V vectors in independent if each vectors points in a geometric dimension not reachable using other vectors in the set.
 - E.g., if two vectors are lie along the same line, then they are really just the same vector; same concept with a 2-D plane containing three vectors in \mathbb{R}^3 .

Basis

- **Basis**: the combination of **span**[↑] and **linear independence**[↑], i.e., the set B of vectors in vector space V if every element of V may be written as a **unique** finite **linear combination**[↑] of elements of B .
 - More concisely, a basis is simply a **linearly independent spanning set**.
 - **Components (coordinates)**: the coefficients of the linear combination and are referred to as of the vector with respect to B , or **basis vectors**.
- Examples of standard basis vectors:

$$\mathbb{R}^2 \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad \mathbb{R}^3 \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

- The standard basis vectors of \mathbb{R}^3 was used in the example of **linear combinations**[↑] to demonstrate how these unit vectors could all be scaled so that they could describe any vector in the space.
- **Basis vectors can be changed**[↓], so it's better to think of them as the “rulers” used to describe any other vector in the space.

Matrices

Matrix Terminology

- **Matrix $A_{r,c}$:** a rectangular array of elements arranged in rows \leftrightarrow and columns \updownarrow , e.g.,

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 5 & 4 & 2 \\ 7 & 6 & 9 \end{bmatrix} \quad A_{3,2} = 6$$

- **Block (partitioned) matrix:** a matrix that is interpreted as having been broken into sections called blocks or submatrices, e.g.,

$$A = \begin{bmatrix} D & N \\ Y & D \end{bmatrix} = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 6 & 9 & 0 & 0 \\ 1 & 1 & 4 & 2 \\ 1 & 1 & 6 & 9 \end{bmatrix}$$

$$D = \begin{bmatrix} 4 & 2 \\ 6 & 9 \end{bmatrix} \quad N = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- Can be used for large matrices with high level structure, offering convenient notation, and sometimes providing computational benefits.
- **Diagonal:** the elements of matrix starting from the top left \searrow lower right.
 - **Off-diagonal:** elements not along the diagonal (0s and 1s in example below)
 - Works for both square and rectangular matrices \downarrow , e.g.,

$$\begin{bmatrix} 4 & 1 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 6 & 0 \\ 1 & 1 & 0 & 9 \end{bmatrix} \quad \begin{bmatrix} 4 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 6 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 9 & 1 & 0 & 1 \end{bmatrix}$$

- **Matrix size:** a matrix with m rows and n columns is called an $m \times n$ matrix, or m -by- n matrix, while m and n are the dimensions.
 - Order matters—the convention is rows then columns, i.e., $m \times n \neq n \times m$.
 - **MR. NiCe** guy: a useful mnemonic to remember typical conventions.
- **Matrix dimensionality:**
 - \mathbb{R}^{mn} : describes the total number of elements, here the multiplication of the dimensions is commutative (order doesn't matter).
 - $\mathbb{R}^{m \times n}$: the specific matrix size using rows and columns as described above.
 - $C(M) \in \mathbb{R}^m$: a collection of column vectors, i.e., a matrix spanned by set column vectors with m elements.
 - $R(M) \in \mathbb{R}^n$: a collection of row vectors, the inverse of above.

Square and Rectangular Matrices

- **Square matrix:** a matrix with the same number of rows and columns.
 - An $n \times n$ matrix is known as a square matrix of order n .
 - Any two square matrices of the same order can be added and multiplied.
- **Rectangular matrix:** a matrix with an unequal number of rows and columns, i.e., $m \neq n$.
- Both square and rectangular matrices have a **diagonal**[†], as described above.

Symmetric and Skew-Symmetric Matrices

- **Symmetric matrix:** a square matrix that can be mirrored across the diagonal, e.g.,

$$\begin{bmatrix} 4 & -6 & -1 \\ -6 & 2 & 9 \\ -1 & 9 & 0 \end{bmatrix}$$

- Algebraically, it's a square matrix \mathbf{A} that is equal to its **transpose**[‡], i.e., $\mathbf{A} = \mathbf{A}^T$.
- It does not matter what is on the diagonal, as any number is equal to itself.
- **Skew-symmetric matrix:** a square matrix that is still symmetric, but all elements mirrored across the diagonal and inverted, e.g.,

$$\begin{bmatrix} 0 & +6 & +1 \\ -6 & 0 & -9 \\ -1 & 9 & 0 \end{bmatrix}$$

- Algebraically, it's a square matrix \mathbf{A} that is equal to its negative transpose, i.e., $\mathbf{A} = -\mathbf{A}^T$
- Here all elements on the diagonal must be zero, as zero is the only number that can be equal its inverse.

Identity and Zero Matrices

- **Identity matrix I_n :** a matrix with size $n \times n$ with **all elements along the diagonal = 1** and **all other elements = 0**, e.g., I_3 and I_n (\dots indicate continuation of pattern):

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

- Essentially, the identity matrix is the equivalent of the number 1 in linear algebra.
- **Zero matrix $\mathbf{0}$:** a matrix of **all zeros**.

Diagonal and Triangular Matrices

- **Diagonal matrix:** when all elements **outside the main diagonal** are zero, i.e.,

$$\begin{bmatrix} e_{1,1} & 0 & \cdots & 0 \\ 0 & e_{2,2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & e_{i,i} \end{bmatrix}$$

- Elements along the diagonal don't have to be the same, and they can be zero (meaning rectangular matrices still can be diagonal, technically).
- **Scaled matrix λI :** when all the elements along the main diagonal are the same and greater than 1, making it a scaled version of the **identity matrix** \uparrow .
- **Triangular matrices:** when all elements above or below the diagonal are zero, but not on both sides.
 - **Upper triangular matrix:** when all the elements **below** the diagonal are zero.
 - **Lower triangular matrix:** when all the elements **above** the diagonal are zero.

$$\text{Upper} = \begin{bmatrix} e_{1,1} & e_{1,2} & e_{1,3} \\ 0 & e_{2,2} & e_{2,3} \\ 0 & 0 & e_{3,3} \end{bmatrix} \quad \text{Lower} = \begin{bmatrix} e_{1,1} & 0 & 0 \\ e_{2,1} & e_{2,2} & 0 \\ e_{3,1} & e_{3,2} & e_{3,3} \end{bmatrix}$$

Augmented and Complex Matrices

- **Augmented (concatenated) matrix $A \mid B$:** a matrix obtained by appending the columns of two given matrices, e.g.,

$$\begin{bmatrix} 4 & 2 & 0 \\ 3 & 7 & 6 \\ 1 & 6 & 9 \end{bmatrix} \mid \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 0 & 4 \\ 3 & 7 & 6 & 2 \\ 1 & 6 & 9 & 0 \end{bmatrix}$$

- Used typically for the purpose of performing the same **elementary row operations** \downarrow on each of the given matrices.
- Matrices must have the same number of rows (or columns for vertical augmentation) for the concatenation to be applied.
- **Complex matrix:** A matrix whose elements may contain complex numbers.
- The **conjugate transpose** \uparrow discussed previously in vectors can be used here as well.
- **Transposition** \downarrow will be discussed shortly, but for now, the complex conjugate still behaves the same (just imaginary numbers change sign), e.g.,

$$\begin{bmatrix} 1 & -1+5i & 0 \\ 1 & -2 & -4 \\ 6i & -4 & 5-2i \end{bmatrix}^H = \begin{bmatrix} 1 & 1 & -6i \\ -1-5i & -2 & -4 \\ 0 & -4 & 5+2i \end{bmatrix}$$

Basic Matrix Operations

Matrix Addition and Subtraction

- **Matrix addition (subtraction)**: the operation of adding (subtracting) two matrices of **equal dimensions** $m \times n$ by adding (subtracting) the corresponding elements together, e.g.,

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 6 & 8 \\ 9 & 6 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 3 & 5 \\ 1 & -6 & 9 \\ -5 & -4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 10 \\ 1 & 0 & 17 \\ 4 & 2 & 4 \end{bmatrix}$$

- Note: there are other operations which could also be considered addition for matrices, such as the direct sum and the Kronecker sum (not discussed as of now).
- ✓ **Commutative**: $A + B = B + A$
- ✓ **Associative**: $A + (B + C) = (A + B) + C$

Matrix Scalar Multiplication

- **Matrix scalar multiplication**: the same as **vector scalar multiplication**[†] or simply, scalar multiplication, as vectors are $m \times 1$ (or $1 \times n$) matrices.
- Scalar multiplication is true when both **left scalar** and **right scalar** are equal, i.e.,

$$\lambda(A)_{ij} = (\lambda A)_{ij} = (A\lambda)_{ij} = (A)_{ij}\lambda$$

- More explicitly:

$$\begin{bmatrix} \lambda e & \lambda e & \cdots & \lambda e \\ \lambda e & \lambda e & \cdots & \lambda e \\ \vdots & \vdots & \ddots & \vdots \\ \lambda e & \lambda e & \cdots & \lambda e \end{bmatrix} = \left(\lambda \text{ or } \begin{bmatrix} e & e & \cdots & e \\ e & e & \cdots & e \\ \vdots & \vdots & \ddots & \vdots \\ e & e & \cdots & e \end{bmatrix} \text{ or } \lambda \right) = \begin{bmatrix} e\lambda & e\lambda & \cdots & e\lambda \\ e\lambda & e\lambda & \cdots & e\lambda \\ \vdots & \vdots & \ddots & \vdots \\ e\lambda & e\lambda & \cdots & e\lambda \end{bmatrix}$$

- The above is **true only** where the underlying ring (algebraic structure that generalize fields...I need to learn more about this...) **is commutative**. This fact is essential for later proofs.

Transposition

- **Transpose**^T: an operation where a matrix is flipped over its **diagonal**[†], i.e., it switches the row and column indices of the matrix, e.g.,

$$\begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 0 \\ 3 & 7 & 1 \\ 4 & 8 & 2 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 0 & 1 & 2 \end{bmatrix}$$

- Formally, the element of the i -th row, j -th column of matrix \mathbf{A} when transposed becomes the element of the j -th row, i -th column of matrix \mathbf{A}^T , i.e.,

$$\mathbf{A}_{i,j} = \mathbf{A}_{j,i}^T$$

- Alternatively, with regard to dimensionality, if \mathbf{A} is an $m \times n$ matrix, then \mathbf{A}^T is an $n \times m$ matrix.
 - Thus, a transposed matrix that is transposed again will produce the original matrix, i.e.,

$$(\mathbf{A}_{j,i}^T)^T = \mathbf{A}_{i,j}$$

- Revisiting complex matrices[†]:
 - Hermitian matrix**: a square complex matrix whose transpose is equal to every entry being replaced with its complex conjugate[†].
 - Denoted: $\mathbf{A}^T = \overline{\mathbf{A}}$
 - Skew-Hermitian matrix**: a Hermitian matrix whose transpose is equal to the negation of its complex conjugate.
 - Denoted: $\mathbf{A}^T = -\overline{\mathbf{A}}$

Diagonal and Trace

- The main diagonal[†] of a matrix can be extracted and turned into a vector.
 - Not to be confused with diagonalization[‡] of a matrix, which is a result of matrix decomposition[‡] resulting from eigendecomposition[‡].
- Trace** $\text{tr}(\mathbf{A})$: the sum of all diagonal elements, defined only for square matrices.

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n e_{i,i} = e_{1,1} + e_{2,2} + \cdots + e_{n,n}$$

- The trace is a linear mapping (two vector spaces that preserves the operations of vector addition and scalar multiplication.), i.e.,

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) \quad \text{tr}(\lambda \mathbf{A}) = \lambda \text{tr}(\mathbf{A})$$

- Additionally, a matrix and its transpose have the same trace, as elements along the main diagonal are not affected, i.e., $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$

Broadcasting

- Broadcasting**: duplication of a vector so that the dimensionality matches a larger matrix, allowing for simplification of element wise addition or multiplication.
 - Technically not a valid operation in linear algebra at face value, but is used commonly in applied linear algebra and machine learning.

Matrix Multiplication

Standard Matrix Multiplication

- **Standard matrix multiplication AB :** a binary operation that produces a matrix from two matrices whose **inner dimensions match**.
 - Multiplication of matrices can be thought of as going from right to left ($A \leftarrow B$), or rather, **A pre-multiplies B** , or **B post-multiplies A** .
 - The number of **columns (n)** in the first matrix must be **equal to** the number of **rows (m)** in the second matrix.
 - **Matrix product:** the product, whose size is equal to **rows of the first matrix** and the number of **columns of the second matrix**.

$$(m_1 \times n_1)(m_2 \times n_2) = m_1 \times n_2$$

- **Transposing** \uparrow a matrix switches the dimensions, so it can enable computation in some cases, e.g,

$$\underbrace{A}_{5 \times 7} = \underbrace{A^T}_{7 \times 5} \quad \underbrace{A^T}_{7 \times 5} \underbrace{B}_{5 \times 2} = \underbrace{C}_{7 \times 2}$$

- Revisiting the **dot (inner) product** \uparrow and the **outer product** \uparrow of vectors:
 - The dot product must have equal-length vectors since the transpose of left vector makes inner dimensions much match, e.g.,

$$\underbrace{v}_{5 \times 1} \underbrace{w}_{5 \times 1} \rightarrow \underbrace{v^T}_{1 \times 5} \underbrace{w}_{5 \times 1} = 1 \times 1 \text{ (scalar)}$$

- However, the transpose of the right vector makes the 1×1 dimensions match, making the outer dimensions irrelevant to the validity of the computation, e.g.,

$$\underbrace{v}_{6 \times 1} \underbrace{w}_{9 \times 1} \rightarrow \underbrace{v}_{6 \times 1} \underbrace{w^T}_{1 \times 9} = 6 \times 9 \text{ (matrix)}$$

Standard Matrix Multiplication Perspectives

- There are four ways to compute and conceptualize the process of multiplication of two matrices, all of which give the same result.
- I.e, if **A** is an **$m \times n$** matrix and **B** is an **$n \times p$** matrix, then their product **C** =

$$\begin{bmatrix} a_{11}b_{11} + \dots + a_{1n}b_{n1} & a_{11}b_{12} + \dots + a_{1n}b_{n2} & \dots & a_{11}b_{1p} + \dots + a_{1n}b_{np} \\ a_{21}b_{11} + \dots + a_{2n}b_{n1} & a_{21}b_{12} + \dots + a_{2n}b_{n2} & \dots & a_{21}b_{1p} + \dots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \dots + a_{mn}b_{n1} & a_{m1}b_{12} + \dots + a_{mn}b_{n2} & \dots & a_{m1}b_{1p} + \dots + a_{mn}b_{np} \end{bmatrix}$$

- **The element perspective:** building the product matrix directly, one element at a time, via the computation of the **dot product**[↑] between the **rows of the left matrix** and the **columns of the right matrix**.

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 1a + 2c & \dots \\ \dots & \dots \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 1a + 2c & 1b + 2d \\ \dots & \dots \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & \dots \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix} \end{aligned}$$

- This is generally the most common perspective as the others methods have all have two “pseudo steps” to reach the final product, rather than through one direct method.
- **The layer perspective:** the building of the product matrix one layer at a time, followed by a “flattening” to make the final product.
 - Each layer is the same size ($m_1 \times n_2$) as the product, but is only a **rank 1 matrix**[↓], or essentially the representation of only one column’s worth of information.
 - Can be thought of as a **left matrix of columns** and a **right matrix of rows**, resulting in the computation of the **outer product**[↑].

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 1a & 1b \\ 3a & 3b \end{bmatrix} \\ \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} 2c & 2d \\ 4c & 4d \end{bmatrix} \\ \begin{bmatrix} 1a & 1b \\ 3a & 3b \end{bmatrix} \Downarrow \begin{bmatrix} 2c & 2d \\ 4c & 4d \end{bmatrix} &= \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix} \end{aligned}$$

- **The column perspective:** the building of the product matrix by columns, where the first column is the sum of the two columns of the left matrix weighted (scaled) by the elements of the first column of the right matrix.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \left(a \begin{bmatrix} 1 \\ 3 \end{bmatrix} + c \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad b \begin{bmatrix} 1 \\ 3 \end{bmatrix} + d \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right) = \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix}$$

- **The row perspective:** similar to the column perspective, but building up by row.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \left(1 \begin{bmatrix} a & b \end{bmatrix} + 2 \begin{bmatrix} c & d \end{bmatrix} \right) = \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix}$$

Various Notes on Matrix Multiplication

Diagonal Matrix Multiplication

- Square **diagonal matrices**[†] are often used to scale another matrix by the elements along such diagonal.
- You can scale the columns or rows depending on the placement of the diagonal matrix **D** relative to original matrix **A**.

• **Pre-multiplication** of the diagonal results in the scaling by **rows**, i.e., **DA**:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} a1 & a2 & a3 \\ b4 & b5 & b6 \\ c7 & c8 & c9 \end{bmatrix}$$

• **Post-multiplication** of the diagonal results in the scaling by **columns**, i.e., **AD**:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} 1a & 2b & 3c \\ 4a & 5b & 6c \\ 7a & 8b & 9c \end{bmatrix}$$

- Again, when all the elements along the diagonal are the same, then it is simply a scaled version of the **identity matrix**[†], or sometimes referred to as a **scaled matrix**[†].

Order of Operations

- “And love is evol, spell it backwards, I’ll show ya”
- $(\mathbf{LOVE})^T = \mathbf{E}^T \mathbf{V}^T \mathbf{O}^T \mathbf{L}^T$: reversing the order of multiplication on a set of matrices is valid if the same operation (e.g., **transpose**[†] or **inverse**[‡]) can be applied to each matrix, e.g., (**diagonal** highlighted for easier time seeing transpose)

$$\begin{aligned} \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right)^T &= \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix}^T = \begin{bmatrix} 1a + 2c & 3a + 4c \\ 1b + 2d & 3b + 4d \end{bmatrix} \\ \begin{bmatrix} a & b \\ c & d \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T &= \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1a + 2c & 1b + 2d \\ 3a + 4c & 3b + 4d \end{bmatrix} \end{aligned}$$

Matrix Vector Multiplication

- Multiplying a matrix by a vector always results in a vector, regardless order.
- The order of multiplication does impact the orientation and size (dimensionality) of the product vector, i.e.,
pre-multiplication → **row vector** (weighted combinations of the rows of **A**)
post-multiplication → **column vector** (weighted combinations of the columns of **A**)

$$\underbrace{\mathbf{w}^T}_{1 \times n} \underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{v}}_{1 \times n} \quad \underbrace{\mathbf{A}}_{m \times n} \underbrace{\mathbf{w}}_{m \times 1} = \underbrace{\mathbf{v}}_{m \times 1}$$

- The order of multiplication does not matter when multiplying a vector with a **symmetric matrix**[↑]:

$$\begin{array}{ll}
 \mathbf{S}^T = \mathbf{S} & \text{symmetric matrix} \\
 \mathbf{S}\mathbf{w} = \mathbf{v} & \text{column vector} \\
 (\mathbf{S}\mathbf{w})^T = \mathbf{v}^T & \text{apply transpose} \rightarrow \text{row vector} \\
 \mathbf{w}^T \mathbf{S}^T = \mathbf{v}^T & \text{love} \leftrightarrow \text{evol}^{\uparrow} \\
 \mathbf{w}^T \mathbf{S} = \mathbf{v}^T & \text{symmetric matrix} \leftrightarrow \text{transpose} \\
 \mathbf{v}^T = \mathbf{v} & \text{row vector} \leftrightarrow \text{column vector}
 \end{array}$$

- If $\mathbf{S}^T \neq \mathbf{S}$, then order does matter, resulting in different product vectors.
- Multiplication of vectors and matrices form the basis of **linear transformations**[↓].
- When multiplication between a matrix and vector is equal to the multiplication between a scalar and the same vector, then the scalar is the **eigenvalue**[↓] and the vector is the **eigenvector**[↓].

Additive and Multiplicative Matrices

- **Multiplicative identity matrix**: commonly referred to as the **identity matrix**[↑], where both **pre** and **post** multiplication are equal and both produce the original matrix.

$$\mathbf{I}\mathbf{A} = \mathbf{A}\mathbf{I} = \mathbf{A}$$

- However, addition of the multiplicative identity matrix does not yield the same product, i.e.,

$$\mathbf{A} + \mathbf{I} \neq \mathbf{A}$$

- **Additive identity matrix**: the complement to the multiplicative matrix that uses the zero matrix (matrix of all zeros), hence why it is commonly referred to as simply the zero matrix.
- Multiplication by the zero matrix of course does not yield the original matrix (unless the original matrix was a zero matrix), but the addition of zero will yield the original product, i.e.,

$$\mathbf{A}\mathbf{0} = \mathbf{0}\mathbf{A} \neq \mathbf{A}, \quad \mathbf{A} + \mathbf{0} = \mathbf{A}$$

Creating Symmetric Matrices

- **Additive method**: using a square matrix \mathbf{A} added to its own transpose will create a **symmetric matrix**[↑], i.e.,

$$\mathbf{S} = \mathbf{A} + \mathbf{A}^T$$

- Non-square matrices result in invalid **matrix addition**[↑] due to unequal dimensions.

- Optionally, a scaling factor of $\frac{1}{2}$ can be used to make the symmetric matrix more closely resemble the original matrix since all values are doubled during addition.
- **Multiplicative method:** multiplying a matrix \mathbf{A} of any dimension with the transpose of itself to create a symmetric matrix \mathbf{S} .
- The order of matrix multiplication matters if the matrix is a non-square matrix, i.e.,

$$\underbrace{\mathbf{A}^T}_{n \times m} \underbrace{\mathbf{A}}_{m \times n} = \underbrace{\mathbf{S}}_{n \times n} \quad \underbrace{\mathbf{A}}_{m \times n} \underbrace{\mathbf{A}^T}_{n \times m} = \underbrace{\mathbf{S}}_{m \times m}$$

- Note, again the order of multiplication does impact the dimensionality of the symmetric matrix produced, i.e.,
 pre-multiplication \rightarrow size $n \times n$ = number of columns in the original matrix
 post-multiplication \rightarrow size $m \times m$ = number of rows in the original matrix
- Proving multiplicative matrices using love \leftrightarrow evol rule[†]:

$$\mathbf{A}^T \mathbf{A} = (\mathbf{A}^T \mathbf{A})^T = \mathbf{A}^T \mathbf{A}^{TT} = \mathbf{A}^T \mathbf{A} \quad \mathbf{A} \mathbf{A}^T = (\mathbf{A} \mathbf{A}^T)^T = \mathbf{A}^{TT} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T$$

- Example of producing a symmetric matrix:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix} = \begin{bmatrix} a^2 + b^2 + c^2 & ad + be + cf \\ ad + be + cf & d^2 + e^2 + f^2 \end{bmatrix}$$

- Notice that the diagonal of the symmetric matrix is the columns of the original (or rows of transposed matrix) squared while the off diagonal elements are the cross terms (or rows of the original and columns of transpose).
- The multiplicative method is widely used signal processing and statistics; the product is often called a covariance matrix, i.e., a matrix where the diagonal elements are the variances and the off diagonal elements are the covariances.

Hadamard Multiplication

- Covered briefly in while introducing other **properties of vectors**[†]; it is often denoted with either \circ , or \odot .
- To reiterate, now with more context, the Hadamard product takes two matrices with equal dimensions and multiplies each element with the corresponding element in the other matrix, i.e.,

$$\begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix} \odot \begin{bmatrix} b_{1,1} & \cdots & b_{1,n} \\ \vdots & \ddots & \vdots \\ b_{m,1} & \cdots & b_{m,n} \end{bmatrix} = \begin{bmatrix} a_{1,1}b_{1,1} & \cdots & a_{1,n}b_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1}b_{m,1} & \cdots & a_{m,n}b_{m,n} \end{bmatrix}$$

- ✓ **associative**, ✓ **distributive**, and ✓ **commutative**, unlike standard matrix multiplication which is not commutative.

Matrix Norms

- **Matrix norm:** a **vector norm**[†] in a vector space whose elements are matrices.
 - More specifically, given a field K (of either real or complex numbers) and the vector space $K^{m \times n}$ of matrices of the size $m \times n$ with entries in the field K , then a matrix norm is a norm in the vector space $K^{m \times n}$.
 - Thus, a matrix norm is a function $\|\mathbf{A}\| : K^{m \times n} \rightarrow \mathbb{R}$ that satisfies the following properties for all scalars $\lambda \in K$ and all matrices $\mathbf{A}, \mathbf{B} \in K^{m \times n}$:
 - $\|\lambda \mathbf{A}\| = |\lambda| \|\mathbf{A}\|$ (absolutely homogenous)
 - $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ (sub-additive or satisfying the triangle inequality)
 - $\|\mathbf{A}\| \geq 0$ (positive)
 - $\|\mathbf{A}\| = 0 \leftrightarrow \mathbf{A} = \mathbf{0}_{m,n}$ (definite)
 - **Submultiplicative norm:** special cases of square matrices that also follow the following condition:
 - $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$ for all matrices \mathbf{A} and \mathbf{B} in $K^{n \times n}$
- There are many types of norms, and special cases for each category of norms, but for now just the Frobenius norm, the induced 2-norm, and the Schatten p-norm will be discussed.

Frobenius Norm

- **Frobenius inner (dot) product** $\langle \mathbf{A}, \mathbf{B} \rangle_F$: binary operation that takes two matrices of equal dimensions and computes a component-wise **inner product**[†] of two matrices as if they were vectors. There are several ways of thinking about how the Frobenius inner product is computed, either:
 - Element-wise multiplication \rightarrow summation of all elements;
 - Vectorization of both matrices \rightarrow vector dot product;
 - **Vectorization** $\text{vec}(\mathbf{A})$: a **linear transformation**[‡] that converts a matrix into a **column vector**, i.e., a $mn \times 1$ column vector is obtained by stacking the columns of the matrix \mathbf{A} on top of one another:

$$\text{vec}(\mathbf{A}) = [a_{1,1}, \dots, a_{m,1}, a_{1,2}, \dots, a_{m,2}, \dots, a_{1,n}, \dots, a_{m,n}]^T$$

- Or the most efficient way, simply taking the **trace**[†] of $\mathbf{A}^T \mathbf{B}$
- Thus, the Frobenius inner product can be defined as:

$$\langle \mathbf{A}, \mathbf{B} \rangle_F = \sum_{i,j} \mathbf{A}_{i,j} \mathbf{B}_{i,j} = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B}) = \text{tr}(\mathbf{A}^T \mathbf{B})$$

- **Frobenius (Euclidean) norm**: the square root of the Frobenius inner product of a matrix with itself, i.e.,

$$\text{norm}(\mathbf{A}) = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle_F} = \sqrt{\text{tr}(\mathbf{A}^T \mathbf{A})}$$

- This works since the use of the **multiplicative method**[†] create a symmetric matrix whose diagonal can be traced, resulting in a vector consisting of each element being squared then summed, thus the **norm of that vector**[†] can easily be taken, yielding the norm of the matrix.
- This is often the most commonly used matrix norm and is often simply called denoted as the $\text{norm}(\mathbf{A})$ of a matrix.

Induced p-Norm

- **Induced p-norm** $\|\mathbf{A}\|_p$: Essentially measures the effect of a matrix on a particular vector norm, particularly if its longer or shorter, hence $\sup()$ in following definition.
- If p (often 2, making it commonly referred to as the 2-norm) is used for both K^n and K^m , then the induced p-norm can be defined as:

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \quad 1 \leq p \leq \infty$$

- The notation does not differ from the Schatten p-norm or other entry-wise p-norms.

Schatten p-Norm

- **Schatten p-norm** $\|\mathbf{A}\|_p$: arises when applying the p-norm to the vector of **singular values of a matrix**[‡] (for now, a special set of scalars with a certain matrix).
- If the singular values of a matrix are denoted by σ , then the Schatten p-norm is defined the sum of all singular values of matrix, i.e.,

$$\|\mathbf{A}\|_p = \left(\sum_{i=1}^r \sigma_i^p \right)^{1/p}$$

- All Schatten norms are submultiplicative.
- Often used cases are when $p = 1, 2$, and ∞ ;
 - $p = 2$: yields the Frobenius norm.
 - $p = \infty$: yields the induced vector 2-norm.
 - $p = 1$: yields the nuclear norm, which is the sum of all the singular values of the matrix.

Matrix Rank

Rank Terminology

- **Matrix rank** r , $\text{rank}(\mathbf{A})$: a non-negative integer ($\mathbb{N}_0 = 0, 1, 2, 3, \dots$) that describes the dimension of the vector space **spanned**[↑] by its columns.
 - Rank can also be thought of as communicating the **dimensionality of the information** in a matrix, but not the dimensionality of the matrix itself.
 - E.g., the columns of a 2×3 matrix are in \mathbb{R}^3 , but one column could be embedded in a two-dimensional **subspace**[↑], which would make it a linearly dependent set.
 - Thus, rank can also be thought of as the maximal number of **linearly independent**[↑] columns of a matrix, which also means rank is dimension of the vector space spanned by its rows, i.e.,

$$\text{rank}(C(\mathbf{A})) = \text{rank}(R(\mathbf{A})) = \text{rank}(\mathbf{A})$$

Maximum Rank

- The maximum possible rank is equal to the smaller of the two dimensions, either the **rows** or **columns**, i.e.,

$$\max(r) = r \in \mathbb{N}_0 \mid 0 \leq r \leq \min(m, n)$$

- **Full rank**: a square matrix with maximum possible rank, making it **invertible**[↓].
- **Full column rank**: when $m > n$ and $\max(r) = n$.
- **Full row rank**: when $m < n$ and $\max(r) = m$.
- **Rank deficient (reduced rank, degenerate, low-rank, singular, non-invertible)**: when the rank is less than the maximum possible rank.

Computing Rank

- In smaller matrices, it may be easily computed by counting the number of columns in a linearly independent set.
- **Row reduction**[↓] to the row echelon form also allows for calculation of rank by counting number of pivots.
- Compute the **singular value decomposition**[↓] and count the number of non-zero singular values.
- Compute the **eigendecomposition**[↓] and count the number of non-zero eigenvalues.

Rank After Addition and Multiplication

- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$
- $\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$
- The key take way here is that the rank is not necessarily the same as either of the original matrices. However, these rules aren't used often in applied linear algebra.

Rank of $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A} \mathbf{A}^T$

- The rank of a matrix when multiplied with itself is the same as the rank of the original matrix (and the transpose of said matrix), regardless of pre/post-multiplication, i.e.,

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A} \mathbf{A}^T)$$

- Explanation 1:
 - Columns \mathbf{c}_j are combinations of columns in \mathbf{A}^T , i.e., combining rows of \mathbf{A} results in column \mathbf{c}_j of product matrix $\rightarrow \mathbf{A}^T \mathbf{a}_j = \mathbf{c}_j$
 - This means the dimensionality of the subspace spanned by the columns of \mathbf{C} is the same as the dimensionality subspace spanned the columns of \mathbf{A}^T , which means the rank is the same, i.e.,

$$\mathbf{C}(\mathbf{C}) = \mathbf{C}(\mathbf{A}^T) \rightarrow \text{rank}(\mathbf{C}) = \text{rank}(\mathbf{A}^T)$$

- Explanation 2: one can use null spaces[↓] to show that the dimensionality of the $\mathbf{A}^T \mathbf{A}$ is the same as \mathbf{A} . I don't understand this yet, so this will hopefully be explained better later.
- Explanation 3: one can use singular value decomposition[↓] to show that $\mathbf{A}^T \mathbf{A}$ has the same singular values (squared) as \mathbf{A} , indicating the rank is also equal. Again, I don't understand this yet, so this will hopefully be explained better later.
- The implications of this fact show that creating symmetric full rank matrices, via the multiplicative method[↑], of desired size (pre=column sized, post=row sized) can easily be done.

Full Rank via "Shifting"

- First, "shifting" a matrix is different from a shift matrix[↓], which may or may not be explicitly defined, for now.
- "Shifting" $\tilde{\mathbf{A}}$: a binary operation that involves the addition of a square matrix with a scaled identity matrix[↑] in order to create a full rank matrix, i.e.,

$$\tilde{\mathbf{A}} = \mathbf{A} + \lambda \mathbf{I}$$

- E.g., below is a rank deficient matrix (rank 2) that has been "shifted" by a slightly scaled identity matrix in order to create the full rank (3):

$$\begin{bmatrix} 1 & 3 & 3 \\ 5 & -7 & -7 \\ -5 & 2 & 2 \end{bmatrix} + 0.1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.01 & 3 & 3 \\ 5 & -6.99 & -7 \\ 5 & 2 & 2.01 \end{bmatrix}$$

- Determining the correct scaling factor without changing the information in the original matrix can be very difficult, but generally small shifts are better.

Matrix Spaces

Systems of Linear Equations

Linear Equations

- **Linear equation:** an equation that may be put in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + b = 0$$

- **Coefficients (parameters)** a_1, a_2, \dots, a_n : typically real numbers; sometimes they can be arbitrary expressions as long as they don't contain any new variables, though often $\{x_{n+1}, x_{n+2}, x_{n+3}, \dots\}$ are replaced with $\{s, r, t, \dots\}$ for convenience.
 - **Variables** x_1, x_2, \dots, x_n : the unknowns of the equation.
 - **Constant term** b : typically a real number; technically also a coefficient, but not attached to any variables.
- **System of linear equations:** a finite collection of linear equations involving the same set of variables, e.g.,

$$\begin{cases} 3x + 2y - z &= 1 \\ 2x + 2y + 4z &= -2 \\ -x + \frac{1}{2}y - z &= 0 \end{cases}$$

- Note: some variables can have a coefficient of 0, so some equations may appear to not contain the same set at first glance.
- Together, all individual equations are used to indicate one object, hence "system."

Solutions

- **Solution:** an assignment of values to variables such that all assignments yield a valid equation, i.e.,

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

- **Linear system solution:** an assignment of values to the variables such that all equations are simultaneously satisfied, e.g., using the above system of equations:

$$\begin{array}{ll} x = 1 & 3(1) + 2(-2) - (-2) = 1 \\ y = -2 & 2(1) + 2(-2) + 4(-2) = -2 \\ z = -2 & -(1) + \frac{1}{2}(-2) - (-2) = 0 \end{array}$$

- **Solution set:** the set of all possible solutions that satisfy all variables in a system of linear equations; there are three possibilities for such system:
 1. The system has infinitely many solutions (the lines are the same).
 2. The system has a single unique solution (the lines intersect at single point).
 3. The system has no solution (the lines are parallel).
- **Inconsistent:** a system that has no solution.
- **Consistent:** a system that has at least one solution.

Elementary Operations

- Solve system of equations is easily done if such systems are first converted to a form of matrix multiplication, i.e., taking this general form of a system of equations:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= b_m \end{cases}$$

and then turning it into the **matrix vector multiplication**[†] form, specifically the **post-multiplication form** in order to create a **column vector of constants**:

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

- **Augmented matrices**[†] are a useful intermediate that is used to represent the final product and perform elementary operations on.
 - Creating an augmented matrix can easily be done by taking the above form, **dropping the variables** temporarily, then **concatenating the vector of constants** onto the **matrix of coefficients**, i.e.,

$$\left[\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \cdots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} & b_m \end{array} \right]$$

- **Elementary operations**: a set of operations that can be performed on a matrix while **maintaining equivalence** (when two systems have the same set of solutions); the operations are as follows:
 1. Interchange two rows: $R_i \updownarrow R_j$
 2. Multiply one row by a nonzero number: $kR_i \rightarrow R_i, k \neq 0$
 3. Add a multiple of one row to a different row: $R_i + kR_j \rightarrow R_i, i \neq j$

Gaussian Elimination

- **Gaussian elimination (row reduction)**: an algorithm for solving systems of linear equations by using a sequence of elementary operations on a matrix until the **row echelon form** is obtained.
 - **Gauss-Jordan elimination**: when Gaussian elimination is used until a matrix reaches the **reduced row echelon form**; sometimes it is computationally beneficial to stop and before such form is reached.
 - The **reduced row echelon form is unique**, unlike the row echelon form, i.e., it is independent of the sequences of row operations used.

Row Echelon form

- **Row echelon form**: a matrix that has been converted into a pseudo **upper triangular matrix**[↑] using Gaussian elimination, e.g. (* = any number; can be zero),

$$\begin{bmatrix} 1 & * & * & * & * & * \\ 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- More specifically, a matrix is in row echelon form if:
 - All rows consisting of only zeroes are on the bottom.
 - **Pivots**: the leading coefficient of a nonzero row is always strictly to the right of the leading coefficient (sometimes required to be 1) of the row above it.
- The row echelon form allows for determination of the **matrix rank**[↑] by counting the number of pivots, e.g., the above example is rank 4.
 - If $\text{rank}(r) = \text{number of columns}(n)$, then it has a unique solution.
 - If $r < n$, then it has infinite solutions.
- **Reduced row echelon form** $\text{rref}(M)$: a matrix that satisfies all the requirements of the row echelon form while additionally satisfying the following:
 - The leading coefficient of each nonzero row must be 1 (often called a leading 1).
 - Columns containing a leading 1 have zeros in all other entries (below **and above**).
 - E.g., using the row echelon form above:

$$\begin{bmatrix} 1 & * & 0 & 0 & * & 0 \\ 0 & 0 & 1 & 0 & * & 0 \\ 0 & 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- **Parameters:** the nonleading variables ($*x_n$; the leftover variables associated with nonzero elements) are replaced with new variables $\{s, r, t, \dots\}$ when converting back systems of equations, hence why **coefficients**[†] are sometimes referred to as parameters.
- Any matrix can be brought to the (reduced) row echelon form by a sequence of elementary row operations; this theorem is the basis of the Gaussian algorithm.

The Gaussian Algorithm

1. Create an augmented matrix from a system of linear equations in order to perform elementary operations more easily.
 - Steps for how to do this is under **elementary operations**[†], as well valid operations for step 2.
2. Use Gaussian/Gaussian-Jordan elimination until the row echelon form/rref is obtained; the majority of the work involves this step (assuming not a zero matrix).
 - 2.1 Find the first column from the left containing a nonzero entry (call it a).
 - 2.2 Create a leading 1 in the top row by multiplying by a^{-1} .
 - 2.3 Subtract multiples of row that row from rows below; make each entry below the leading 1 zero.
 - 2.4 Repeat steps 2.1–2.3, on the next row.
 - Sometimes interchanging rows can make this process more efficient, look for invalid pivots to save computational effort.
 - 2.5 Stop when either no rows remain in step 4, or the remaining rows consist of entirely of zeros to obtain the row echelon form, continue to obtain rref.
 - 2.6 Repeat step 2.3, but now subtracting multiples of that row from rows above the row; make each entry above the leading 1 zero.
3. There a few different possibilities once solved, i.e.,
 - If a row with the form $[0 \ 0 \ 0 \ \dots \mid a \neq 0]$ occurs, then the system is **inconsistent**, i.e., at least one line is parallel with another—no solution possible.
 - If a row with all zeros occurs, then that row added no new information and was simply a multiple of another row.
 - If reduced form is found, then continue.
4. Map the matrix back to the equations.
5. Back-substitution of the nonleading coefficients (parameters) to solve for leading coefficients.
6. Double check for validity by using leading variables in the original system.

Homogeneous Equations

- **Homogeneous systems:** when a system of linear equations have zeros in all the constant terms, i.e.,

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n &= 0 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n &= 0 \\ &\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n &= 0 \end{aligned}$$

- Equivalent to a matrix equation of the form $\mathbf{A}\mathbf{v} = \mathbf{0}$.

Homogeneous Solution Sets

- **Trivial solution:** obtained by assigning the value of zero to each of the variables.
- If the system has a **non-singular matrix**[↓], then it is also the only solution.
- If the system has a **singular matrix**[↓] then there are infinite solutions and consists of a solution set with the following properties:
 - If \mathbf{u} and \mathbf{v} are two vectors representing solutions, then the vector sum $\mathbf{u} + \mathbf{v}$ is also a solution.
 - If \mathbf{u} is a vector representing a solution, and λ is any scalar, then $\lambda\mathbf{u}$ is also a solution.
 - These are exactly the same properties required for the solution set to be a **subspace**[↑] of \mathbb{R}^n .
 - Numerical solutions to a homogeneous can be found with **singular value decomposition**[↓].
- There is a close relationship between the solutions to a linear system and the solutions to the corresponding homogeneous system:

$$\mathbf{Ax} = \mathbf{b} \quad \text{and} \quad \mathbf{Ax} = \mathbf{0}$$

- Specifically, if \mathbf{p} is any specific solution to the linear system $\mathbf{Ax} = \mathbf{b}$, then the entire solution set can be described as

$$\{\mathbf{p} + \mathbf{v} : \mathbf{v} \text{ is any solution to } \mathbf{Ax} = \mathbf{0}\}$$

- Geometrically, $\mathbf{Ax} = \mathbf{b}$ is simply translation of the solution set for $\mathbf{Ax} = \mathbf{0}$.
- The above only applies if $\mathbf{Ax} = \mathbf{b}$ has at least one solution.

The Determinant

-

Matrix Inverse

-

—To Be Defined—

- Diagonalization
- Matrix decomposition
- Linear transformations
- Eigendecomposition
- Change of Basis
- Singular values of a matrix
- Non-singular matrix
- Singular matrix
- Null spaces
- Singular value decomposition