Introduction to Linear Algebra

Linear algebra is crucial for chemists as it provides the mathematical framework for understanding and describing various chemical phenomena. For example, it is used in quantum chemistry to solve Schrödinger's equation, model molecular orbitals, and analyze atomic interactions. In spectroscopy, linear algebra helps process and interpret large data sets, such as from NMR or IR spectroscopy. It also plays a key role in molecular modeling, where vector and matrix operations are used to understand molecular geometry, predict reaction outcomes, and simulate chemical processes.

1 Scalars

Most of the computational operations rely on the manipulations of numbers one at a time. This is what we defined as **scalars**. One of the most common examples of scalars is temperature, and we can manipulate scalars using functions,

$$t_C = f(t_F) = \frac{5}{9}\left(t_F - 32\right). \tag{1}$$

In the majority of the text books, scalars are denoted by lower-case letters, for example t_C or x. It is important to remember that scalars are continuous real-valued and commonly defined as,

$$x \in \mathbb{R}, \tag{2}$$

where $\in$ means "in" and $\mathbb{R}$ are the real numbers set.

Scalars numbers have the possible operations,

- addition, $x + y$
- multiplication xy
- division $\frac{x}{y}$
- exponentiation x^y

In programming, scalars are implemented as tensors that only contain a single element, and slicing operations are prohibited.

2 Vectors

In the hierarchy of tensors, vectors are the first structure that has a single dimension, meaning they are 1^{st} oder tensors. From the computational perspective, vectors are considered a fixed-length array of scalars, commonly assumed to be vertically,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}, \tag{3}$$

where $x_i \in \mathbb{R}$, and d is the number of items in $\mathbf{x}$. Vectors are also represented as arrows with a direction and an angle (I personally not fan of this notation).

2.1 Unitary vectors

In physics, it is common to represent vectors in terms of unitary vectors,

$$\mathbf{u} = u_x\,\mathbf{i} + u_y\,\mathbf{j} + u_z\,\mathbf{k}, \tag{4}$$

where

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{5}$$

Here we present some of the most common operations for vectors.

2.2 Scalar multiplication

The operation between a scalar and a vector is called scalar multiplication. In scalar multiplication, each component of the vector is multiplied by the scalar, resulting in a new vector that points in the same (or opposite) direction but with a scaled magnitude.

$$c\,\mathbf{x} = \begin{bmatrix} cx_0 \\ \vdots \\ cx_d \end{bmatrix}, \tag{6}$$

where cx_i is the multiplication of each entry of $\mathbf{x}$ by the scalar c.

2.3 Addition

The addition of vectors involves adding corresponding components of two vectors to produce a new vector,

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_0 \\ \vdots \\ x_d \end{bmatrix} + \begin{bmatrix} y_0 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} x_0 + y_0 \\ \vdots \\ x_d + y_d \end{bmatrix} \tag{7}$$

.

We can combine scalar multiplication and addition of two vectors to compose the subtraction of two vectors, $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1)\,\mathbf{y}$.

The sum of two vectors using the physics notation, is defined as,

$$\mathbf{u} \pm \mathbf{v} = (u_x \pm v_x)\,\mathbf{i} + (u_y \pm v_y)\,\mathbf{j} + (u_z \pm v_z)\,\mathbf{k}\,. \tag{8}$$

2.4 Division

Division of vectors is not defined in the same straightforward way as addition, subtraction, or scalar multiplication. Unlike scalars, vectors don't have a direct division operation. However, there are some related concepts, i) scalar division and ii) element-wise division. For scalar division, we can use Eq. 6, where the scalar is defined as $\frac{1}{c}$, making the operation for each element of $\mathbf{x}$, $\frac{x_i}{c}$.

In programming, the division between two vectors is the element-wise division,

$$\mathbf{x} \,/\, \mathbf{y} = \begin{bmatrix} \frac{x_0}{y_0} \\ \vdots \\ \frac{x_d}{y_d} \end{bmatrix}, \tag{9}$$

where each element of the new vector is the division of elements of the two vectors.

One must be careful with element-wise operations as they are only valid for vectors with the same number of elements.

2.5 Dot product

The dot product, also known as the *scalar product*, is an operation that takes two vectors and returns a scalar. It measures the extent to which two vectors point in the same direction. The scalar product is defined as,

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \|u\|\|v\| \cos(\theta), \tag{10}$$

where $\|u\|$ is the norm of the vector $\mathbf{u}$ and, θ is the angle between both vectors. $\mathbf{u}^\top$ is the transpose of the vector $\mathbf{u}$, which means a *row vector*,

$$\mathbf{x}^\top = [x_0, \cdots, \mathbf{x}_d]. \tag{11}$$

The norm, magnitude or length of a vector is defined as

$$\| \mathbf{u} \| = \sqrt{x_0^2 + x_1^2 + \cdots + x_d^2} = \left(\sum_{i=0}^{d} x_i^2 \right)^{\frac{1}{2}}. \tag{12}$$

Exercise:
Compute, $\| \mathbf{u} + \mathbf{v} \|$ for $\mathbf{u} = [2, -1, 3]$ and $\mathbf{v} = [-1, 1, -1]$.

Exercise:
What is the dot product of $\mathbf{i} \cdot \mathbf{i}$ and $\mathbf{i} \cdot \mathbf{j}$?

Exercise:
What is the dot product of $\mathbf{u} \cdot \mathbf{v}$ for For, $\mathbf{u} = u_x \,\mathbf{i} + u_y \,\mathbf{j} + u_z \,\mathbf{k}$ and $\mathbf{v} = v_x \,\mathbf{i} + v_y \,\mathbf{j} + v_z \,\mathbf{k}$?

A more general definition of the dot product for d-dimensional vectors is,

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \sum_{i=1}^{d} u_i v_i. \tag{13}$$

Exercise:
What is the value of θ between $\mathbf{u} = \mathbf{i} + 3\,\mathbf{j} - \mathbf{k}$ and $\mathbf{v} = \mathbf{j} - \mathbf{k}$?

2.6 Cross product

The cross product, also known as *vector product* is an operation between two vectors that results in a new vector perpendicular to both original vectors. It is defined as,

$$\mathbf{u} \times \mathbf{v} = \|u\| \|v\| \mathbf{c} \sin(\theta), \tag{14}$$

where $\mathbf{c}$ is a unit vector perpendicular to the plane form by $\mathbf{u}$ and $\mathbf{v}$. We can find $\mathbf{c}$ using the following equations,

$$\begin{aligned} \mathbf{u}^\top \mathbf{c} &= 0 && (15) \\ \mathbf{v}^\top \mathbf{c} &= 0 && (16) \\ \mathbf{c}^\top \mathbf{c} &= 1, && (17) \end{aligned}$$

in matrix notation we get,

$$\begin{aligned} u_x c_x + u_y c_y + u_z c_z &= 0 && (18) \\ v_x c_x + v_y c_y + v_z c_z &= 0 && (19) \\ c_x c_x + c_y c_y + c_z c_z &= 1. && (20) \end{aligned}$$

We will see later in the course that these non-linear equations can be solved using optimization methods.

For three dimensional vectors and using unitary vectors, the cross product is defined as,

$$\mathbf{u} \times \mathbf{v} = (u_y v_z - u_z v_y)\,\mathbf{i} + (u_z v_x - v_x v_z)\,\mathbf{j} + (u_x v_y - u_y v_x)\,\mathbf{k}\,. \tag{21}$$

Commonly, you will find that,

$$\mathbf{u} \times \mathbf{v} = \underbrace{\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix}}_{determinant} = (u_y v_z - u_z v_y)\,\mathbf{i} + (u_z v_x - v_x v_z)\,\mathbf{j} + (u_x v_y - u_y v_x)\,\mathbf{k}\,. \tag{22}$$

Exercise:
What is the value of $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$? For each vector you can consider the physics notation.

3 Matrices

Matrices are 2^{nd}-order tensors, and they have two dimensions, n number of rows and m number of columns. Commonly, matrices in textbooks are defined as capital letters in bold font, $\mathbf{A} \in \mathbb{R}^{n,m}$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix}, \tag{23}$$

where $\mathbf{a}_i^\top$ is a m-dimensional row-vector, $\mathbf{a}_i^\top = [a_{i,1}, a_{i,2}, \cdots, a_{i,m}]$; $\mathbf{a}_i \in \mathbb{R}^m$.

Similar to vectors and scalars, there are some well defined operations for matrices.

3.1 Scalar multiplication

Matrices can be scaled by being multiply by a scalar,

$$c\,\mathbf{A} = \begin{bmatrix} c\,a_{11} & c\,a_{12} & \cdots & c\,a_{1m} \\ c\,a_{21} & c\,a_{22} & \cdots & c\,a_{2m} \\ \vdots & & & \vdots \\ c\,a_{n1} & c\,a_{n2} & \cdots & c\,a_{nm} \end{bmatrix} = \begin{bmatrix} c\,\mathbf{a}_1^\top \\ c\,\mathbf{a}_2^\top \\ \vdots \\ c\,\mathbf{a}_n^\top \end{bmatrix}. \tag{24}$$

3.2 Matrix-Matrix element-wise operations

There are three main element-wise operations between matrices, i.g., addition, subtraction and the Hadamard product. For the latter, the mathematical symbol used is $\odot$.

$$\mathbf{A} \pm \mathbf{B} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1m} \pm b_{1m} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2m} \pm b_{2m} \\ \vdots & & & \vdots \\ a_{n1} \pm b_{n1} & a_{n2} \pm b_{n2} & \cdots & a_{nm} \pm b_{nm} \end{bmatrix} = \begin{bmatrix} (\mathbf{a}_1 \pm \mathbf{b}_1)^\top \\ (\mathbf{a}_2 \pm \mathbf{b}_2)^\top \\ \vdots \\ (\mathbf{a}_n \pm \mathbf{b}_n)^\top \end{bmatrix} \tag{25}$$

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}\, b_{11} & a_{12}\, b_{12} & \cdots & a_{1m}\, b_{1m} \\ a_{21}\, b_{21} & a_{22}\, b_{22} & \cdots & a_{2m}\, b_{2m} \\ \vdots & & & \vdots \\ a_{n1}\, b_{n1} & a_{n2}\, b_{n2} & \cdots & a_{nm}\, b_{nm} \end{bmatrix} = \begin{bmatrix} (\mathbf{a}_1 * \mathbf{b}_1)^\top \\ (\mathbf{a}_2 * \mathbf{b}_2)^\top \\ \vdots \\ (\mathbf{a}_n * \mathbf{b}_n)^\top \end{bmatrix}, \tag{26}$$

where

$$\mathbf{a}_i * \mathbf{b}_j = \begin{bmatrix} a_{i1}\, b_{j1} \\ a_{i2}\, b_{j2} \\ \cdots \\ a_{im}\, b_{jm} \end{bmatrix} \tag{27}$$

3.3 Matrix-vector multiplication

Matrix-vector multiplication is defined as,

$$\underbrace{\mathbf{A}}_{(n,m)} \underbrace{\mathbf{x}}_{(m,1)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m \end{bmatrix} = \underbrace{\begin{bmatrix} \sum_{i=1}^m a_{1i}x_i \\ \sum_{i=1}^m a_{2i}x_i \\ \vdots \\ \sum_{i=1}^m a_{ni}x_i \end{bmatrix}}_{(n,1)} \tag{28}$$

In matrix-vector multiplications, one has to be careful with the inner-index.

3.4 Matrix-Matrix multiplication

Matrix-matrix multiplication is defined as,

$$\begin{aligned} \underbrace{\mathbf{A}}_{(n,m)} \underbrace{\mathbf{B}}_{(m,k)} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{bmatrix} && (29) \\ &= \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_k \end{bmatrix} && (30) \\ &= \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_k \end{bmatrix} && (31) \\ &= \underbrace{\begin{bmatrix} \mathbf{a}_1^\top \mathbf{b}_1 & \mathbf{a}_1^\top \mathbf{b}_2 & \cdots & \mathbf{a}_1^\top \mathbf{b}_k \\ \mathbf{a}_2^\top \mathbf{b}_1 & \mathbf{a}_2^\top \mathbf{b}_2 & \cdots & \mathbf{a}_2^\top \mathbf{b}_k \\ \vdots & & & \vdots \\ \mathbf{a}_n^\top \mathbf{b}_1 & \mathbf{a}_n^\top \mathbf{b}_2 & \cdots & \mathbf{a}_n^\top \mathbf{b}_k \end{bmatrix}}_{(n,k)} && (32) \end{aligned}$$

3.5 Miscellaneous matrices

Here, I will present some additional information and concepts related to matrices.

3.5.1 Transpose

The transpose of a matrix, $\mathbf{A}^\top$, is obtained by interchanging its rows and columns. So if, if $\mathbf{A} \in \mathbb{R}^{n,m}$, its transpose $\mathbf{A}^\top \in \mathbb{R}^{m,n}$.

$$\underbrace{\mathbf{A}}_{(n,m)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad \underbrace{\mathbf{A}^\top}_{(m,n)} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix} \tag{33}$$

3.5.2 Symmetric and antiSymmetricmatrices

If a square matrix, $\mathbf{A} \in \mathbb{R}^{n,n}$, is equal to it's transpose, $\mathbf{A} = \mathbf{A}^\top$, meaning $a_{ij} = a_{ji}$, this matrix is said to be **symmetric**.

- If square matrix is equal to its transpose, it is said to be **Symmetric**.

$$\mathbf{A} = \mathbf{A}^\top \tag{34}$$

- If a square matrix is equal to minus one times its transpose, $a_{ij} = -a_{ji}$, it is said to be **antiSymmetric**.

$$\mathbf{A} = -\,\mathbf{A}^\top \tag{35}$$

3.5.3 Hermitian and antiHermitian matrices

So far we have assumed that the matrix elements are real numbers, $a_{ij} \in \mathbb{R}$. However, for some systems it is convenient to defined matrices that have complex numbers, $a_{ij} \in \mathbb{C}$ [1]. the Hermitian transpose of a matrix, defined as $\mathbf{A}^\dagger$, also called the conjugate transpose of the matrix, is obtained by interchanging its rows and columns and taking its complex conjugate of its elements, [2]

$$\underbrace{\mathbf{A}}_{(n,m)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad \underbrace{\mathbf{A}^\dagger}_{(m,n)} = \begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{n1}^* \\ a_{12}^* & a_{22}^* & \cdots & a_{n2}^* \\ \vdots & & & \vdots \\ a_{1m}^* & a_{2m}^* & \cdots & a_{nm}^* \end{bmatrix} \tag{36}$$

- If a square matrix is equal to its Hermitian transpose, it is said to be **Hermitian**.

$$\mathbf{A} = \mathbf{A}^\dagger \tag{37}$$

- If a square matrix is equal to minus one times its Hermitian transpose, $a_{ij} = -1a_{ji}^*$ it is said to be **antiHermitian**.

$$\mathbf{A} = -\,\mathbf{A}^\dagger \tag{38}$$

[1] $z = x_{\text{real}} + iy_{\text{imaginary}}$

[2] the complex conjugate of $z = x + iy$ is $z^* = x - iy$

3.6 Properties of matrix operations

Let's assume we have three matrices, $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, and two scalars c and d. We can defined some additional properties for matrix operations.

- $(c+d)\,\mathbf{A} = c\,\mathbf{A} + d\,\mathbf{A}$
- $c(d\,\mathbf{A}) = (cd)\,\mathbf{A}$
- $\mathbf{A}(\mathbf{B}\,\mathbf{C}) = (\mathbf{A}\,\mathbf{B})\,\mathbf{C}$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\,\mathbf{B} + \mathbf{A}\,\mathbf{C}$
- $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$, **multiplication of matrices is not commutative**
- $\mathbf{A}\,\mathbf{I} = \mathbf{I}\,\mathbf{A} = \mathbf{A}$

For the last property, $\mathbf{I}$ is an identity matrix, defined as,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{39}$$

where only the diagonal elements are equal to 1, $I_{ii} = 1$.

Exercise: Show that for rectangular matrices, $\mathbf{A}\,\mathbf{I} = \mathbf{A}$.

4 System of linear equations

Studying linear equations is essential because they form the foundation for solving real-world problems in science, engineering. For example, let's consider the equation for the logarithm of the equilibrium constant for an isothermal reaction,

$$\ln K = -\frac{\Delta G_{rxn}}{RT} = -\frac{\Delta H_{rxn} - T\Delta S_{rxn}}{RT} \tag{40}$$

$$\underbrace{\ln K}_{y} = \underbrace{-\frac{\Delta H_{rxn}}{R}}_{a} \underbrace{\left(\frac{1}{T}\right)}_{x} + \underbrace{\frac{\Delta S_{rxn}}{R}}_{b} \tag{41}$$

We can determined experimentally the value of $\ln K$ at two different temperatures and ask, what are the values of ΔS_{rxn} and ΔH_{rxn} for this chemical process? For example, for the $2NO_2 \longleftrightarrow N_2O_4$ reaction we have the following experimental values of $\ln K$ and T,

$$4.3 = -\frac{\Delta H_{rxn}}{R}\left(\frac{1}{273K}\right) + \frac{\Delta S_{rxn}}{R} \tag{42}$$

$$0.26 = -\frac{\Delta H_{rxn}}{R}\left(\frac{1}{325K}\right) + \frac{\Delta S_{rxn}}{R}, \tag{43}$$

These equations can be rewritten in matrix for, leading to *two linear equations*,

$$\begin{bmatrix} 4.3 \\ 0.26 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{273} \\ 1 & -\frac{1}{325} \end{bmatrix} \begin{bmatrix} \frac{\Delta S_{rxn}}{R} \\ \frac{\Delta H_{rxn}}{R} \end{bmatrix} \tag{44}$$

We can solve these equations, by first solving one in terms of the variable $\frac{\Delta H_{rxn}}{R}$, and the substitute the results in the other equation. For example, from Eq. 43 we get,

$$\frac{\Delta S_{rxn}}{R} = 0.26 + \frac{\Delta H_{rxn}}{R}\left(\frac{1}{325K}\right), \tag{45}$$

substituting this into Eq. 43 we now get,

$$\begin{aligned} 4.3 &= -\frac{\Delta H_{rxn}}{R}\left(\frac{1}{273K}\right) + 0.26 + \frac{\Delta H_{rxn}}{R}\left(\frac{1}{325K}\right) && (46)\\ 4.3 - 0.26 &= \frac{\Delta H_{rxn}}{R}\left(\frac{1}{325K} - \frac{1}{273K}\right) && (47)\\ \frac{\Delta H_{rxn}}{R} &= \frac{4.34}{-5.86\times 10^{-4}K^{-1}} && (48)\\ \Delta H_{rxn} &= \frac{4.34}{-5.86\times 10^{-4}K^{-1}}\left(0.0083145\frac{kJ}{Kmol}\right) = -57.9\frac{kJ}{mol} && (49) \end{aligned}$$

Given the value of ΔH_{rxn}, we can solve for ΔS_{rxn} in Eq. 43,

$$\Delta S_{rxn} = 0.26R - 57.9\frac{kJ}{mol}\left(\frac{1}{325K}\right) = -0.176\frac{kJ}{Kmol}. \tag{50}$$

Linear set of equations have the general form of,

$$y = ax + by + zc \qquad \text{or more generally} \qquad y = \sum_i^n w_i x_i \tag{51}$$

and each can be represented as a vector-vector multiplication,

$$y = \mathbf{w}^\top \mathbf{x} = \underbrace{\begin{bmatrix} w_0 & w_1 & \cdots & w_n \end{bmatrix}}_{\text{linear parameters}} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_i^n w_i x_i \tag{52}$$

When multiple linear equations are combined we get

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix} = \underbrace{\begin{bmatrix} w_{10} & w_{11} & \cdots & w_{1n} \\ w_{20} & w_{21} & \cdots & w_{2n} \\ \vdots & & & \vdots \\ w_{k0} & w_{k1} & \cdots & w_{kn} \end{bmatrix}}_{\mathbf{W}} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}, \tag{53}$$

where each row of the matrix $\mathbf{W}$ represents the linear weights of an equation. From this equation we can observe the following,

- Swapping rows does not affect the over all representation, it simple "orders" the equations differently.
- Swapping columns does not affect the over all representation, as long as we also change the order of the corresponding elements in $\mathbf{x}$.
- Each individual equation can be multiplied by a scalar function and this will not alter the over all solution.

4.1 Gauss method

For equations that have more than 2 unknown variables, there exist methods/recipes to solve a set of linear equations. Here, I will present the Gauss method. Let's consider the following set of linear equations.

$$25 = x_0 + 3x_1 + 6x_2 \tag{54}$$
$$58 = 2x_0 + 7x_1 + 14x_2 \tag{55}$$
$$19 = 2x_1 + 5x_2 \tag{56}$$

which is equal to

$$\left.\begin{array}{ccc|c} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{array}\right.$$

• multiply by (-2) row 1 and sum it to row 2.

$$\begin{array}{ccc|c} -2 & -6 & -12 & -50 \\ 0 & 1 & 2 & 8 \\ 0 & 2 & 5 & 19 \end{array}$$

• multiply by (-2) row 2 and sum it to row 3.

$$\begin{array}{ccc|c} -2 & -6 & -12 & -50 \\ 0 & -2 & -4 & -16 \\ 0 & 0 & 1 & 3 \end{array}$$

the last row indicates that $x_2 = 3$.

• substitute $x_2 = 3$ in row 2 and solve for x_1.

$$-2x_1 - 4(3) = -16 \quad \rightarrow \quad x_1 = 2 \tag{57}$$

• substitute $x_2 = 3$ and $x_1 = 2$ in row 1 and solve for x_0.

$$-2x_0 - 6(2) - 12(3) = -50 \quad \rightarrow \quad x_0 = 1 \tag{58}$$

The Gauss-Jordan method is an extension of the Gauss method where we further do operations but now from the lower to the upper rows, aiming to find a diagonal matrix. Let's start from the last matrix equation that we have and,

• multiply by ($\frac{1}{4}$) row 2 and sum it to row 3 to row 2.

$$\begin{array}{ccc|c} -2 & -6 & -12 & -50 \\ 0 & -\frac{1}{2} & 0 & -1 \\ 0 & 0 & 1 & 3 \end{array}$$

• multiply row 2 by (-2)

$$\begin{array}{ccc|c} -2 & -6 & -12 & -50 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}$$

• multiply row 3 by (12) and sum it to row 1.

$$\begin{array}{ccc|c} -2 & -6 & 0 & -14 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}$$

• multiply row 2 by (6) and sum it to row 1.

$$\left.\begin{array}{ccc|c} -2 & 0 & 0 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right.$$

• multiply row 1 by (-$\frac{1}{2}$).

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

These new steps in the Gauss-Jordan method are known as, *back-substitution*, and the goal is to generate the diagonal matrix where each variable is only equal to a value.

4.2 Determinant of a matrix

In Section 2.6, we saw that the determinant can be used to compute the cross product of two vectors. The determinant helps determine whether a system of linear equations has a unique solution, no solution, or infinitely many solutions.

- A non-zero determinant means the system has a unique solution.
- A zero determinant indicates either no solution or infinitely many solutions.

The determinant of a 2×2 matrices,

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc, \tag{59}$$

and for a 3×3 matrix is,

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & g & i \end{bmatrix} = a(ei - fh) - b(di - fg) + c(dh - eg) \tag{60}$$

Going back to our previous example, we can see that the value of the determinant is equal to 1,

$$\det \begin{bmatrix} 1 & 3 & 6 \\ 2 & 7 & 14 \\ 0 & 2 & 5 \end{bmatrix} = 1(7 * 5 - 2 * 14) - 3(2 * 5 - 0 * 14) + 6(2 * 2 - 0 * 7) = 1 \tag{61}$$

5 Useful identities

There are several useful identities related to the inverse, transpose, and Hermitian transpose of matrix multiplication.

$$(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1} \quad \text{inverse of a matrix product} \tag{62}$$

$$(\mathbf{A}\,\mathbf{B})^{\top} = \mathbf{B}^{\top}\,\mathbf{A}^{\top} \quad \text{transpose of a matrix product} \tag{63}$$

$$(\mathbf{A}\,\mathbf{B})^{\dagger} = \mathbf{B}^{\dagger}\,\mathbf{A}^{\dagger} \quad Hermitiantransposeofamatrixproduct \tag{64}$$

The above identities can be found in the Matrix Cookbok.

Similarly to "common numbers" where the inverse of a number[3] $x^{-1}\, x = x\, x^{-1} = 1$, a similar concept exist only for **nonsingular square matrices**,

$$\mathbf{A}\ \mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}, \tag{65}$$

where $\mathbf{I}$ is the identity matrix, defined as,

$$\mathbf{I} \quad = \quad \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{66}$$

where only the diagonal elements are equal to 1, $I_{ii} = 1$.

The definition of nonsingular matrix (for a square matrix, at least) is "has an inverse". The magnitude of a number is the square root of the number times its complex conjugate, $zz^* = z^*z = |z|^2$; this number is never negative. Similarly, the "square" of a matrix can be considered the value of this matrix times its Hermitian transpose; either $\mathbf{A}^\dagger\,\mathbf{A}$ or $\mathbf{A}\ \mathbf{A}^\dagger$. These matrices are not always the same. **A** matrix for which they are the same is called a *normal* matrix.

Since the matrices $\mathbf{A}^\dagger\,\mathbf{A}$ and $\mathbf{A}\ \ \mathbf{A}^\dagger$ are the "square" of something, that they should be a negative in some sense. These matrices are in fact symmetric and nonnegative, in the sense that for *any* vector $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}^\dagger\,\mathbf{A}\,\mathbf{x} \geq 0$. Such matrices are said to be positive semidefinite. In particular if,

positive semidefinite matrix (analogous to nonnegative numbers)	for every $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} \geq 0$
positive definite matrix (analogous to positive numbers)	for every $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} > 0$
negative semidefinite matrix (analogous to nonpositive numbers)	for every $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} \leq 0$
negative definite matrix (analogous to nonnegative numbers)	for every $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} < 0$
indefinite matrix	for some $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} > 0$ for other $\mathbf{x}$, $\mathbf{x}^\dagger\,\mathbf{A}\,\mathbf{x} > 0$

We are now in position to define several different types of matrices; this can be thought of as the "taxonomy" of matrices, see Table 1 for a summary.

[3] except for $x = 0$

Table 1: **Taxonomy of matrices**, Types of m-by-n matrices $\mathbf{A}$, with elements $\mathbf{a_{ij}}$

Square	Number of rows and columns is the same	$m = n$
Diagonal	Square matrix that is zero except on the diagonal	$a_{ij} = 0$ unless $i = j$
Upper-triangular	Square matrix that is zero below the diagonal	$a_{ij} = 0$ if $i > j$
Lower-triangular	Square matrix that is zero above the diagonal	$a_{ij} = 0$ if $i < j$
Symmetric	Square matrix that is "symmetric" with respect to reflection across the diagonal	$\mathbf{A} = \mathbf{A}^{\top}$ $a_{ij} = a_{ji}$
Antisymmetric	Square matrix that is "antisymmetric" with respect to reflection across the diagonal	$\mathbf{A} = -\mathbf{A}^{\top}$ $a_{ij} = -a_{ji}$
Hermitian	Square matrix that is equal to its Hermitian conjugate/transpose	$A = A^{\dagger}$ $a_{ij} = a_{ji}^{*}$
AntiHermitian	Square matrix that is equal to minus one times its Hermitian conjugate/transpose	$\mathbf{A} = -\mathbf{A}^{\dagger}$ $a_{ij} = -a_{ji}^{*}$
Orthogonal	Square matrix whose transpose is equal to its inverse	$\mathbf{A}^{-1} = \mathbf{A}^{T}$ $\mathbf{A}^{T}\,\mathbf{A} = \mathbf{A}\ \mathbf{A}^{T} = \mathbf{I}$
Unitary	Square matrix whose Hermitian transpose is equal to its inverse	$\mathbf{A}^{-1} = \mathbf{A}^{\dagger}$ $\mathbf{A}^{\dagger}\,\mathbf{A} = \mathbf{A}\ \mathbf{A}^{\dagger} = \mathbf{I}$
Idempotent	Square matrix whose square is equal to itself	$\mathbf{A}\,\mathbf{A} = \mathbf{A}$
Nilpotent	Square matrix that, raised to some power d, equals the zero matrix. d is the degree of the matrix.	$\mathbf{A}^{n} \neq 0$; $n <$ degree $\mathbf{A}^{n} = 0$; $n \leq$ degree
Normal	Square matrix that commutes with its Hermitian conjugate	$\left[\mathbf{A}, \mathbf{A}^{\dagger}\right] = 0$ $\mathbf{A}\ \mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}\,\mathbf{A}$
Row-stochastic	Sum of the elements in any row is equal to one; all elements are positive	$a_{ij} \geq 0$ $\sum_{j}^{n} a_{ij} = 1$
Column-stochastic	Sum of the elements in any column is equal to one; all elements are positive	$a_{ij} \geq 0$ $\sum_{i}^{m} a_{ij} = 1$
Doubly-stochastic	Matrix is both row- and column-stochastic	$a_{ij} \geq 0$ $\sum_{j}^{n} a_{ij} = \sum_{i}^{m} a_{ij} = 1$