

MATHEMATICAL PRELIMINARIES

SCALAR CONCEPTS

The simplest object that we will be dealing with in this course is that of a **scalar**. A scalar α is an element of a **field** \mathcal{F} . A field is a type of algebraic structure that has certain processes defined on it. These processes manipulate elements of the field.

An example of a field is the set of **real numbers** \mathcal{R} . A field may also encompass the set of **complex numbers** \mathcal{C} . The set of **integers** \mathcal{N} is a subset of either case. For each of these examples, scalars would be the elements of these sets, which are numbers. Arithmetic is used to manipulate these numbers to perform computations.

SCALAR OPERATIONS

SCALAR ADDITION

The first fundamental process is scalar addition. There are four main properties of scalar multiplication:

Commutative Property: Given $x, y \in \mathcal{R}$, $x + y = y + x$, i.e., the result is independent of the order of scalars.

Associative Property: Given $x, y, z \in \mathcal{R}$, $(x + y) + z = x + (y + z)$, i.e., the grouping of scalars has no influence on the result.

Multiplicative Identity Property: Given $x \in \mathcal{R}$, $x + 0 = x$; the value 0 here is the identity element for the reals.

Inverse Element Property: For any $x \in \mathcal{R}$, there exists a $y \in \mathcal{R}$ such that $x + (-y) = 0$, where the value of 0 is the identity element for the reals.

SCALAR OPERATIONS

SCALAR MULTIPLICATION

Another necessary process is scalar multiplication. There are five main properties of scalar multiplication:

Commutative Property: Given $x, y \in \mathcal{R}$, $x * y = y * x$, i.e., the result is independent of the order of scalars.

Associative Property: Given $x, y, z \in \mathcal{R}$, $(x * y) * z = x * (y * z)$, i.e., the grouping of scalars has no influence on the result.

Multiplicative Identity Property: Given $x \in \mathcal{R}$, $x * 1 = x$; the value 1 here is the identity element for the reals.

Inverse Element Property: For any $x \in \mathcal{R}$, there exists a $y \in \mathcal{R}$ such that $x * y = 1$, where the value of 1 is the identity element for the reals.

Distributive Property: Multiplication distributes over addition: given $x, y, z \in \mathcal{R}$,
$$x * (y + z) = xy + xz.$$

SCALAR OPERATIONS

SCALAR FUNCTIONS AND DERIVATIVES

Both scalar addition and multiplication are examples of **functions**. A function, which is also known as a **mapping**, is a relation between a set of inputs and a set of outputs with the property that each input is related to one or more outputs. Functions can be either **bijective** (one-to-one), **surjective** (many-to-one), or **injective** (one-to-some).

SCALAR OPERATIONS

SCALAR FUNCTIONS AND DERIVATIVES

Functions typically change in response to alterations in their parameters. The measure of change is referred to as the **derivative**.

Let $f: \mathcal{R} \rightarrow \mathcal{R}$ be a real-valued function defined in an open neighborhood of a scalar $x \in \mathcal{R}$. The **tangent line** to the graph of a function f at x is the unique line through the point $(x, f(x))$ that does not meet the graph of f transversally; that is, the tangent line does not pass straight through the graph. The derivative of this function at x is the **slope** of the tangent line to the graph of f at $(x, f(x))$. To find the slope is to consider a point x_0 that is close to x , connect the points by a line segment, then find the segment's slope. As $x_0 \rightarrow x$, we recover the tangent slope:

$$\frac{d}{dx}f(x) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

SCALAR OPERATIONS

SCALAR FUNCTIONS AND DERIVATIVES: EXAMPLE

SCALAR OPERATIONS

SCALAR FUNCTIONS AND DERIVATIVES: EXAMPLE

SCALAR OPERATIONS

SCALAR FUNCTIONS AND PARTIAL DERIVATIVES

Often, a function may have more than one input that influences its output. We would still like to measure how much the function changes according to changes in one or more inputs. However, we cannot easily do this simultaneously across all inputs. Rather, we must consider the individual change across each input, which leads to the notion of the **partial derivative**.

Like ordinary derivatives, the partial derivative is defined as a limit. Let $f: \mathcal{R}^n \rightarrow \mathcal{R}$ be a real-valued function, defined on an open neighborhood, with m scalar variables $x \in \mathcal{R}^n$. The partial derivative at a point $x = (x_1, x_2, \dots, x_n)$ with respect to the i th variable x_i is defined as

$$\frac{\partial}{\partial x_i} f(x) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{h}.$$

SCALAR OPERATIONS

SCALAR FUNCTIONS AND PARTIAL DERIVATIVES: EXAMPLE

SCALAR OPERATIONS

SCALAR FUNCTIONS AND PARTIAL DERIVATIVES: EXAMPLE

MATHEMATICAL PRELIMINARIES

VECTOR CONCEPTS

Often, we will be considering a concatenation of multiple scalars. This leads to a structure called a **vector**. Each of these scalars might represent **features** about an **object**. For instance, if we are wanting to classify different types of flowers, we might measure their petal widths and lengths and estimate the average petal color. Each of these attributes would be features. Each new flower that is measured leads to a new vector.

More formally, a **vector space** over a **field** \mathcal{F} , such as the reals \mathcal{R} , is a set \mathcal{V} together with **two operations**: **addition** and **scalar multiplication**. These two operations satisfy **nine axioms**. Elements of this vector field $v \in \mathcal{V}$ are referred to as vectors.

VECTOR OPERATIONS

VECTOR SCALING

There are a variety of processes that can be defined for vectors. The simplest is **scaling**. Given an n -dimensional, real vector $x \in \mathcal{R}^n$ with components $[x_1, x_2, \dots, x_n]$, scaling this vector involves multiplying each component by a real scalar factor $\alpha \in \mathcal{R}$:

$$\alpha x = [\alpha x_1, \alpha x_2, \dots, \alpha x_n] \in \mathcal{R}^n.$$

In the case of a **Euclidean metric space**, notice that scaling does not change the **direction** of the vector; it only impacts its **magnitude**. The same result is true in many other metric spaces.

VECTOR OPERATIONS

VECTOR SCALING

Many of the properties of scalar multiplication carry over to the vectorial case:

Scalar Additivity: Given $\alpha, \beta \in \mathcal{R}$ and $v \in \mathcal{R}^n$, $(\alpha + \beta)v = \alpha v + \beta v$.

Distributive Property: Multiplication distributes over vector addition: given $\alpha \in \mathcal{R}$ and $v, w \in \mathcal{R}^n$, $\alpha * (v + w) = \alpha v + \alpha w$.

Multiplicative Identity Property: Multiplication by the identity element doesn't change the vector: given $v \in \mathcal{R}^n$, $1 * v = v$.

Zero Element Property: Multiplication by the zero element yields the zero vector: given $v \in \mathcal{R}^n$, $0 * v = [0, 0, \dots, 0] \in \mathcal{R}^n$.

Inverse Property: Multiplication by the negative identity element gives the additive inverse: given $v \in \mathcal{R}^n$, $-1 * v = -v$.

VECTOR OPERATIONS

VECTOR ADDITION AND ELEMENT-WISE MULTIPLICATION

Addition and **element-wise multiplication** of vectors with the same dimensionality work component-wise. For vectors $x, y \in \mathcal{R}^n$ with components $[x_1, x_2, \dots, x_n]$ and $[y_1, y_2, \dots, y_n]$, we have that

$$x + y = [x_1 + y_1, x_2 + y_2, \dots, x_n + y_n] \in \mathcal{R}^n.$$

Subtraction is a special case of addition where one vector is pre-scaled via a unit negative value. As well, for element-wise multiplication of two vectors, $x, y \in \mathcal{R}^n$, we have that

$$xy = [x_1 y_1, x_2 y_2, \dots, x_n y_n] \in \mathcal{R}^n.$$

Element-wise division of vectors is a special case of multiplication. Both of these fundamental operations can change the direction and magnitude of vectors.

VECTOR OPERATIONS

VECTOR ADDITION AND ELEMENT-WISE MULTIPLICATION

Many of the properties of scalar addition carry over to the vectorial case:

Commutative Property: Given $v, w \in \mathcal{R}^n$, $v + w = w + v$.

Associative Property: Given $v, w, x \in \mathcal{R}^n$, $(v + w) + z = v + (w + z)$.

Zero Element Property: Addition by the zero vector gives back the original vector: given $v \in \mathcal{R}^n$, $0 + v = v$.

Inverse Element Property: For any $v \in \mathcal{R}^n$, there exists a $w \in \mathcal{R}^n$ such that $v + w = [0, 0, \dots, 0] \in \mathcal{R}^n$.

VECTOR OPERATIONS

INNER PRODUCT

VECTOR OPERATIONS

INNER PRODUCT: EXAMPLE

Example: Find the inner product of the following vectors: $x = [1, -2, 5, 0]$ and $y = [5, 15, 2, 1]$.

Solution: We can simply multiply each vector component together and sum their result:

$$x^{\top}y = 1 * 5 - 2 * 15 + 5 * 2 - 1 * 0 = -15.$$

Example: Find the inner product of the following vectors: $x = [2, 0.7, 5, 1]$ and $y = [0.5, 4, 9, 18]$.

Solution: We can simply multiply each vector component together and sum their result:

$$x^{\top}y = 2 * 0.5 + 0.7 * 4 + 5 * 9 + 18 * 1 = 66.8.$$

VECTOR OPERATIONS

OUTER PRODUCT

The **outer product** of two vectors $x \in \mathcal{R}^d$ and $y \in \mathcal{R}^n$ yields a matrix whose entries $x_i y_i \in \mathcal{R}$ are scalar values

$$xy^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}^\top = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_d y_1 & x_d y_2 & \cdots & x_d y_n \end{bmatrix} \in \mathcal{R}^{n \times m}.$$

The outer product is an operator that can act on other vectors. If we have a vector $w \in \mathcal{R}^k$, then $(xy^\top)w = x(y^\top w) = (x \cdot y)w$. Assuming that $x = y$, there is a simple geometric interpretation for this process. Recall that a vector describes a single direction in a vector space. All scalar multiples of that vector define a line lying in that direction. The exterior product of x with itself measures the projection of $x \cdot w$ onto x , then rescales x to to have that length.

VECTOR OPERATIONS

OUTER PRODUCT: EXAMPLE

Example: Find the outer product of the following vectors: $x = [1, -2, 5, 0]$ and $y = [5, 15, 2, 1]$.

Solution: We can simply multiply each vector component together to form a matrix

$$\begin{bmatrix} 1 \\ -2 \\ 5 \\ 0 \end{bmatrix} \begin{bmatrix} 5 \\ 15 \\ 2 \\ 1 \end{bmatrix}^T = \begin{bmatrix} 5 & 15 & 2 & 1 \\ -10 & -30 & -4 & -2 \\ 25 & 75 & 10 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Example: Find the outer product of the following vectors: $x = [2, 0.7, 5, 1]$ and $y = [0.5, 4, 9, 18]$.

Solution: We can simply multiply each vector component together to form a matrix

$$\begin{bmatrix} 2 \\ 0.7 \\ 5 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 4 \\ 9 \\ 18 \end{bmatrix}^T = \begin{bmatrix} 1 & 8 & 18 & 36 \\ 0.35 & 2.8 & 6.3 & 12.6 \\ 2.5 & 20 & 45 & 90 \\ 0.5 & 4 & 9 & 18 \end{bmatrix}.$$

VECTOR OPERATIONS

VECTOR DERIVATIVES

MATHEMATICAL PRELIMINARIES

MATRIX CONCEPTS

The other construct is that of a **matrix**. A matrix over a field \mathcal{F} is a rectangular array of scalars, each of which is a member of \mathcal{F} . The set of all matrices forms an **algebraic ring** \mathcal{M} with two fundamental processes: **entry-wise addition** and **matrix multiplication**. Each of these processes satisfies a series of axioms, just like in the vector case.

MATRIX OPERATIONS

MATRIX SCALING

Addition and inner-product-based multiplication of vectors are degenerate instances of corresponding matrix operations. For example, suppose that we have a factor $\alpha \in \mathcal{R}$. The **scaling** of a matrix $X \in \mathcal{R}^{n \times m}$ is the entry-wise multiplication by that factor

$$\alpha X = \begin{bmatrix} \alpha x_{1,1} & \alpha x_{1,2} & \cdots & \alpha x_{1,m} \\ \alpha x_{2,1} & \alpha x_{2,2} & \cdots & \alpha x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha x_{n,1} & \alpha x_{n,2} & \cdots & \alpha x_{n,m} \end{bmatrix} \in \mathcal{R}^{n \times m}.$$

Notice that scaling is **commutative**; hence, the pre- or post-multiplication of the matrix by the scalar does not change the result.

MATRIX OPERATIONS

MATRIX SCALING: EXAMPLE

Example: Find the result of multiplying the following matrix by the scalar $\alpha = 0.2$

$$\begin{bmatrix} 9 & 4 & 10 & -2 \\ 6 & 8 & -4 & 0 \\ 18 & 5 & 10 & 19 \\ 0 & 1 & -8 & 4 \end{bmatrix}.$$

Solution: We simply scale each entry of the matrix by $\alpha = 0.2$, which leads to

$$\begin{bmatrix} 1.8 & 0.8 & 2 & -0.4 \\ 1.2 & 1.6 & -0.8 & 0 \\ 3.6 & 1 & 2 & 3.8 \\ 0 & 0.2 & -1.6 & 0.8 \end{bmatrix}.$$

MATRIX OPERATIONS

MATRIX ADDITION

Addition also works in an element-wise fashion. Given two equally sized matrices $X, Y \in \mathcal{R}^{n \times m}$, their addition is defined as

$$X + Y = Y + X = \begin{bmatrix} x_{1,1} + y_{1,1} & x_{1,2} + y_{1,2} & \cdots & x_{1,m} + y_{1,m} \\ x_{2,1} + y_{2,1} & x_{2,2} + y_{2,2} & \cdots & x_{2,m} + y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} + y_{n,1} & x_{n,2} + y_{n,2} & \cdots & x_{n,m} + y_{n,m} \end{bmatrix} \in \mathcal{R}^{n \times m}.$$

Subtraction is again a special case of addition. Addition is commutative; hence, the order of the matrices is interchangeable without changing the results.

MATRIX OPERATIONS

MATRIX ADDITION: EXAMPLE

Example: Find the result for the following

$$\begin{bmatrix} 5 & 2 & 0 \\ 4 & 9 & 2 \\ 10 & -3 & -1 \\ 5 & 12 & 16 \end{bmatrix} + \begin{bmatrix} -11 & 0 & 6 \\ 7 & 1 & 0 \\ -6 & -8 & 12 \\ 2 & 10 & 4 \end{bmatrix}$$

Solution: We simply add the corresponding entries together

$$\begin{bmatrix} 5 - 11 & 2 + 0 & 0 + 6 \\ 4 + 7 & 9 + 1 & 2 + 0 \\ 10 - 6 & -3 - 8 & -1 + 12 \\ 5 + 2 & 12 + 10 & 16 + 4 \end{bmatrix} = \begin{bmatrix} -6 & 2 & 6 \\ 11 & 10 & 2 \\ 4 & -11 & 11 \\ 7 & 22 & 20 \end{bmatrix}.$$

MATRIX OPERATIONS

MATRIX TRANSPOSE

Matrices also have a **transpose** operation. A given matrix $X \in \mathcal{R}^{n \times m}$ is said to be transposed when its columns are converted into rows and vice versa, which leads to a matrix $X^T \in \mathcal{R}^{m \times n}$ with

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}^T = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,m} & x_{2,m} & \cdots & x_{n,m} \end{bmatrix} \in \mathcal{R}^{m \times n}.$$

MATRIX OPERATIONS

MATRIX TRANSPOSE: EXAMPLE

Example: Find the transpose of the following matrix

$$\begin{bmatrix} 2 & 0 & 3 & -8 \\ 1 & -5 & -2 & 6 \\ 7 & 4 & 0 & 9 \end{bmatrix}.$$

Solution: Each row of the original matrix becomes a column and each column of the original matrix becomes a row:

$$\begin{bmatrix} 2 & 1 & 7 \\ 0 & -5 & 4 \\ 3 & -2 & 0 \\ -8 & 6 & -9 \end{bmatrix}.$$

MATRIX OPERATIONS

MATRIX INVERSION

MATRIX OPERATIONS

MATRIX INVERSION: EXAMPLE

MATRIX OPERATIONS

MATRIX MULTIPLICATION

MATRIX OPERATIONS

MATRIX MULTIPLICATION: EXAMPLE

Example: Perform the following multiplication

$$\begin{bmatrix} 3 & 4 & 2 \end{bmatrix} * \begin{bmatrix} 13 & 9 & 7 & 15 \\ 8 & 7 & 4 & 6 \\ 6 & 4 & 0 & 3 \end{bmatrix}$$

Solution: We simply compute a series of dot products

$$\begin{bmatrix} 3 * 13 + 4 * 8 + 2 * 6 \\ 3 * 9 + 4 * 7 + 2 * 4 \\ 3 * 7 + 4 * 4 + 3 * 0 \\ 3 * 15 + 4 * 6 + 2 * 3 \end{bmatrix}^T = \begin{bmatrix} 83 \\ 63 \\ 37 \\ 75 \end{bmatrix}^T.$$

MATRIX OPERATIONS

EIGENDECOMPOSITIONS

MATRIX OPERATIONS

MATRIX DERIVATIVES

FUNCTION OPERATIONS

GRADIENTS OF VECTOR-BASED FUNCTIONS ON SCALAR DOMAINS

FUNCTION OPERATIONS

GRADIENTS OF VECTOR-BASED FUNCTIONS ON VECTOR DOMAINS

FUNCTION OPERATIONS

GRADIENTS OF MATRIX-BASED FUNCTIONS ON SCALAR DOMAINS

FUNCTION OPERATIONS

GRADIENTS OF MATRIX-BASED FUNCTIONS ON VECTOR AND MATRIX DOMAINS