CIS3990-002: Mathematics of Machine Learning	Fall 2023

Lecture: Linear Algebra

Date: October 4th, 2023 — *Author: Eric Wong*

1 Linear Algebra Basics

Most likely you are familiar with basic operations on matrices and vectors. For example, $A \in \mathbb{R}^{3\times 5}$ is a matrix of real numbers with 3 rows and 5 columns, while $b \in \mathbb{R}^3$ is a vector of 3 elements. These form the basis of a system that we call linear algebra, which has several main properties. Our goal in this module will be to learn about the the fundamental properties of linear systems, and generalize these properties to abstract vector spaces that are not necessarily in the field of real numbers.

- For $m, n \in \mathbb{N}$, a matrix A is a m, n tuple of elements a_{ij} where i denotes the row and j denotes the column.
- Addition: if $C = A + B$ and $A, B \in \mathbb{R}^{m\times n}$ then $c_{ij} = a_{ij} + b_{ij}$
- Product: If $C = AB$ and $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times n}$ then $c_{ij} = \sum_{l=1}^{k} a_{il} b_{lj}$ where $C \in \mathbb{R}^{m\times n}$
- Identity: $I_m \in \mathbb{R}^{m\times m}$ is an identity matrix when it is zero everywhere except the diagonal, i.e. $I_{ij} = \mathbf{1}[i = j]$
- Associativity: $(AB)C = A(BC)$
- Distributivity: $(A + B)C = AC + BC$, $A(C + D) = AC + AD$
- Multiplication with identity: $\forall A \in \mathbb{R}^{m\times n} : I_m A = AI_n = A$
- Inverse: Let $A \in \mathbb{R}^{n\times n}$. If $AB = I$ then $B = A^{-1}$ is the inverse of A
- Transpose: Let $A \in \mathbb{R}^{m\times n}$. The matrix $B = A^\top$ such that $b_{ij} = a_{ji}$ is called the transpose.
- Symmetric: $A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^\top$

We can also add scalars to the mix (single elements).

- Scalar multiplication: Let $\lambda \in \mathbb{R}$. Then, $\lambda A = K$ where $K_{ij} = \lambda a_{ij}$.
- Associativity: $(\lambda\phi)C = \lambda(\phi C)$. Actually, scalars can be moved around: $\lambda(BC) = (\lambda B)C = B(\lambda C) = (BC)\lambda$. Also, transpose doesn't affect matrices: $(\lambda C)^\top = C^\top \lambda = \lambda C^\top$
- Distributivity: $(\lambda + \phi)C = \lambda C + \phi C$ and $\lambda(B + C) = \lambda B + \lambda C$

One of the most common uses of matrices and vectors is to represent linear systems of equations in a compact form. I.e.

$$Ax = b$$

represents a series of linear equations, where each row of A is the coefficients for each variable x and the target scalar is the corresponding row in b.

2 Groups

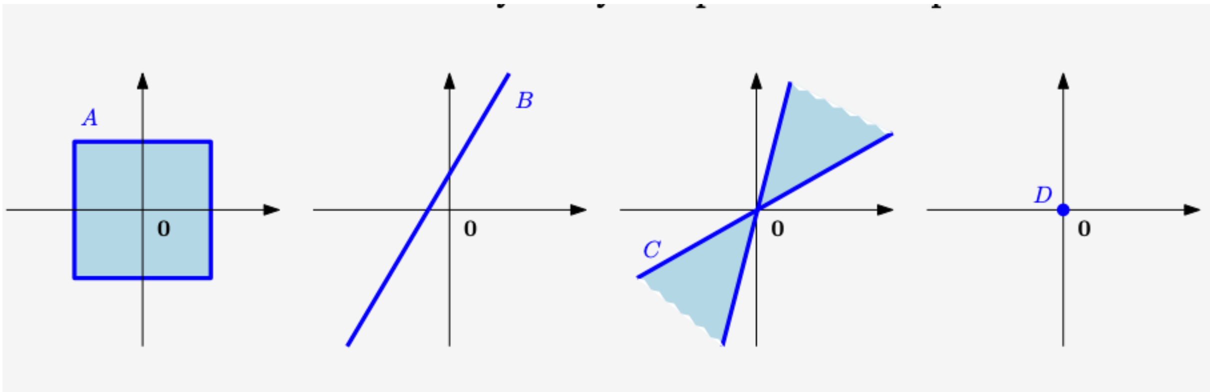
The space of matrices and vectors behaves *nicely*, in that it has these properties of associativity, distributivity, an identity and an inverse. Let's now generalize this structure.

- Groups: Let $\mathcal{G}$ be a set and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ be defined on $\mathcal{G}$. Then $G = (\mathcal{G}, \otimes)$ is called a group if
  1. Closure: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
  2. Associativity: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
  3. Neutral element: $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \otimes e = e \otimes x = x$
  4. Inverse element: $\forall x \in \mathcal{G} \exists y \in \mathcal{G} : x \otimes y = y \otimes x = e$. We write x^{-1} to denote the inverse element of x. This does not always mean $\frac{1}{x}$ and is with respect to the operator $\otimes$.
  5. (Commutivity) If $f \forall x, y \in \mathcal{G} : x \otimes y = y \otimes x$ then $\mathcal{G}$ is an Abelian group
- Examples of Abelian groups: $(\mathcal{Z}, +), (\mathcal{R} \setminus \{0\})$
- Examples of not-groups: $(\mathcal{N} + 0, +), (\mathcal{Z}, \cdot), (\mathcal{R}, \cdot)$
- $(\mathcal{R}^n, +), (\mathcal{Z}^n, +)$ are Abelian if using component wise addition
- Matrices and addition: $(\mathcal{R}^{m \times n}, +)$ is Abelian with component-wise addition
- Matrices and multiplication: $(\mathcal{R}^{m \times n}, \cdot)$ is only a group if the inverse always exists
- General Linear Group: set of invertible matrices $A \in \mathcal{R}^{n \times n}$ is a group with respect to matrix multiplication, but is not Abelian (not commutative)

3 Vector Spaces

Groups have an operation with structure that stays within the group. This can be referred to as an *inner* operation (i.e. elementwise addition) as the operator stays within the group. We can also consider an *outer* operation which takes in an element outside of the group.

- Real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with operations $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $\cdot : \mathcal{R} \times \mathcal{V} \to \mathcal{V}$
- Distributivity:
  1. $\forall \lambda \in \mathbb{R}, \forall x, y \in \mathcal{V} : \lambda \odot (x + y) = \lambda \odot x + \lambda \odot y$
  2. $\forall \lambda, \psi \in \mathbb{R}, x \in \mathcal{V} : (\lambda + \psi) \odot x = \lambda \odot x + \psi \odot y$
- $x \in V$ are called vectors, the neutral element is 0, and the inner operator is vector addition while the outer operation is multiplcation by scalars.
- A subspace of a vector space is a vector space: if $\mathcal{U} \subset \mathcal{V}$ and $V = (\mathcal{V}, +, \odot)$ is a vector space, then if $U = (\mathcal{U}, +, \cdot)$ is a vector space we call it a subspace of V restricted to $\mathcal{U}$.



- Subspaces inherit properties from the higher space, including Abelian, distributivity, associativity, and neutral element. To show that U is a subspace, we need to show that $0 \in \mathcal{U}$ and U is closed with respect to both inner and outer operations (i.e. $\forall \lambda \forall x \in \mathcal{U} : \lambda x \in \mathcal{U}$ and $\forall x, y \in \mathcal{U} : x + y \in \mathcal{U}$).

- Example 2.12 from the textbook.

This structure gives us the nice properties we expect in linear algebra (i.e. we can do operations on vectors that result in more vectors).

4 Linear Independence

- Linear combination is a combination of scaled vectors:

$$v = \sum_i \lambda_i x_i$$

- If there exists λ such that $0 = \sum_i \lambda_i x_i$ with at least one $\lambda_i \neq 0$ then they are linearly dependent. If no such non-zero solution exists, they are linearly independent.