CSE 840: Computational Foundations of Artificial Intelligence January 22, 2025

Linear Mappings

Instructor: Vishnu Boddeti *Scribe: Auden Garrard*

1 Linear Mapping

Definition 1 *Let U, V be vector spaces over the same field F. A mapping $T : U \to V$ is called a linear map if $\forall u_1, u_2 \in U$ and $\lambda \in F$:*

$$T(u_1 + u_2) = T(u_1) + T(u_2)$$
$$T(\lambda u_1) = \lambda T(u_1).$$

The set of all linear mappings from $U \to V$ is denoted is denoted $\mathcal{L}(U, V)$.

If $U = V$, then we denote $\mathcal{L}(U)$

Examples

- The zero map $0 : V \to W$, mapping every element $v \in V$ to $0 \in W$, is linear.
- The identity map $I : V \to V$, defined as $Iv = v$, is linear.
- None linear, For example, the exponential function $f(x) = e^x$ is not linear since $e^{2x} \neq 2e^x$.
- None linear, the function $f : F \to F$ given by $f(x) = x - 1$ is not linear since:

$$f(x + y) = (x + y) - 1 \neq (x - 1) + (y - 1) = f(x) + f(y).$$

Definition 2 *$T \in \mathcal{L}(U, V)$. Then Kernel of T (null-space of T) is defined as*

$$ker(T) := null(T) := \{u \in U | T_u = 0\}$$

Proposition 3

- *$ker(T)$ is a subspace of U*
- *T injective if and only if $kern(T) = 0$*

Definition 4 *The range of T (image of T) is defined as*

$$\text{range}(T) := \text{image}(T) := \{T_u | u \in U\}$$

Proposition 5

- *The* range *is always a subspace of* V
- T *is subjective if and only if* $\text{range}(T) = V$

Definition 6 *Let* v' *be any subset of* V *i.e* $v' \subset$ *The pre-image of* v' *is defined as*

$$T^{-}1(v') := \{u \in U | T_u \in v'\}$$

Proposition 7 *If* $v' \subset V$ *is a subspace of* V *then* $T^{-}1(v'$ *is a subspace of* $U)$

Theorem 8 *Let* V *be a finite-dimensional vector space,* W *be any vector space, and* $T \in \mathcal{L}(V, W)$. *Let* $(u_1, \ldots, u_n)$ *be a basis of* $\ker(T) \subset V$. *Let* $(w_1, \ldots, w_m)$ *be a basis of* $\text{range}(T) \subset W$. *Then:*

$$u_1, \ldots, u_n, T^{-1}(w_1), \ldots, T^{-1}(w_m) \subset V$$

form a basis of V. *In particular:*

$$\dim(V) = \dim(\ker(T)) + \dim(\text{range}(T)).$$

Proof: Denote $T^{-}1(w_1) = z_1, \ldots, T^{-}1(W_m) = Z_m$ Step 1: $V \subset span\{u_1, \ldots, u_2, z_1; \ldots, z_m\}$

Step One: Let $\vec{v} \in V$ consider $T_v \in \text{range}(T)$; Reminder $\vec{v_1}, \ldots, \vec{v_n}$ are basis of $\ker(T)$

$$\Rightarrow \exists \lambda_1, \ldots, \lambda_m s.t.$$

$$\begin{aligned} T_v &= \lambda_1 w_1 + \lambda_2 w_2 + \ldots \lambda_m W_m \\ &= \lambda_1 T(z_1) + \lambda_2 T(z_2) \ldots \lambda_m z_m \\ &= T(\lambda_1 z_1 + \lambda_2 z_2 + \ldots \lambda_m z_m) \end{aligned}$$

$$\Rightarrow Tv - T(\lambda_1 z_1 + \lambda_2 z_2 + \ldots \lambda_m z_m) = 0$$

$$\Rightarrow T(\underbrace{v - (\lambda_1 z_1 + \lambda_2 z_2 + \cdots + \lambda_m z_m)}_{\in \ker(T)}) = 0$$

$$\Rightarrow \exists \mu_1, \ldots \mu_m s.t.$$

$$v - (\lambda_1 z_1 + \lambda_2 z_2 + \cdots + \lambda_m z_m) = \mu_1 u_1 + \mu_2 u_2 + \ldots \mu_n u_n$$

$$\Rightarrow v = \lambda_1 z_1 + \lambda_2 z_2 + \cdots + \lambda_m z_m + \mu_1 u_1 + \mu_2 u_2 + \ldots \mu_n u_n$$

Step Two: $u_1, u_2, \ldots, u_n; z_1, z_2, \ldots, z_m$ are linearly independent.

$$\text{Assume that } \mu_1 u_1 + \ldots \mu_n u_n + \lambda_1 z_1 + \cdots + \lambda_m z_m = 0^{\star}$$

$$\text{Now consider: } \lambda_1 w_1 + \ldots \lambda_m w_m = \lambda_1 T(z_1) + \ldots \lambda_m T(z_m)$$

$$= \lambda_1 T(z_1) + \ldots \lambda_m T(z_m) + \underbrace{\mu_1 T(\underline{u_1}) + \cdots + \mu_n T(\underline{u_n})}_{0}$$

$$T(\underbrace{\lambda_1 z_1 + \cdots + \lambda_m z_m + \mu_1 u_1 + \cdots + \mu_n u_n}_{0 \text{ by } \star}) = 0$$

$$\Rightarrow \lambda_1 w_1 + \cdots + \lambda_m w_m = 0$$
$$\Rightarrow w_1 \ldots w_m \text{ basis } \lambda_1 = 0 \ldots \lambda_m = 0$$
$$\Rightarrow \mu_1 u_1 + \cdots + \mu_u u_u = 0$$
$$\mu_1 = \mu_2 = \cdots = \mu_n = 0 \text{ since } u_1, \ldots, u_n \text{ are basis}$$

□

Example

Consider a matrix $A \in \mathbb{R}^{3\times 3}$ representing a linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$. Suppose the rank of A is 2 (i.e., the dimension of the image of T is 2), and the nullity of A is 1 (i.e., the dimension of the kernel is 1). The Rank-Nullity Theorem tells us that:

$$\dim(\mathbb{R}^3) = \dim(\ker(A)) + \dim(\text{range}(A)),$$

so:

$$3 = 1 + 2.$$

Proposition 9 $T \in \mathcal{L}(V, V)$, V *is* **finite_dim**. *Then the following statements are equivalent:*

- T *is injective*
- T *is surjective*
- T *is bijective*

Proof: Direct consequence of theorem. Also, only holds in finite dimensional spaces □

Real world examples

- Proposition 5 importance
  - The range of the subspace is crucial in Principal Component Analysis (PCA), especially when reducing dimensions. PCA projects the data onto a lower-dimensional subspace (the range) that captures as much of the data's variance as possible.
- Pre-images
  - Support Vector Machines (SVMs) work by mapping data to a higher-dimensional space to make the problem linearly separable. The pre-images help make it possible to understand how the original data corresponds to points or subsets in the feature space

2 Matrices and Linear Maps

Notation:

$$A = \text{The } j\text{-row} \left\{ \underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{n-columns} \right. = (a_{ij})_{i=1,\dots,m,j=1,\dots,n}$$

Proposition 10 *Consider $T \in \mathcal{L}(v, w)$, v,w* **finite_dim***, let $\vec{v_1}, \dots, \vec{v_n}$ be a basis of V, $w_1, \dots, w_m$ be a basis of W*

- $$V = \lambda_1 \vec{v_1} + \cdots + \lambda_n \vec{v_n}$$
  $$T(v) = T(\lambda_1 \vec{v_1} + \cdots + \lambda_n \vec{v_n})$$
  $$= \lambda_1 T(\vec{v_1}) + \cdots + \lambda_n T(\vec{v_n})$$

- *Each image vector $T\vec{v_j}$ can be expressed in basis $w_1, \dots, w_m$: there exist co-coefficients $a_1 j, \dots, a_m j$ s.t,*
  $$T(v_j) = a_{1j} w_1 + \cdots + a_{mj} w_m$$

- *we now stack a these co-efficient in a matrix:*
  $$\begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

Notation: Let $T : V \to W$ be linear. let B a basis of V, C basis of W. We denote by $M(T, B, C)$ the matrix corresponding to T with respect to bases B and C

Convenient propertied of matrix Let V, W be vector spaces, consider the bases fixed. Let $S, T \in \mathcal{L}(V, W)$

- Linear properties of mapping extend to matrices as well:
  $$M(S + T) = M(S) + M(T)$$
  $$M(\lambda S) = \lambda M(S)$$

- For $v = \lambda_1 \vec{v_1} + \cdots + \lambda_n \vec{v_n}$, we have that:
  $$\underbrace{T(v)}_{\text{image of } v \text{ under } T} = \underbrace{M(T) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}}_{\text{matrix-vector product}}$$
  where $(\vec{v_1}, \dots, \vec{v_n})$ is a basis of V.

- $T : C \to V; S : V \to W$ linear, then $M(S \circ T) = M(S) \times M(T)$
- $\circ$ is the composition of the maps S and T

Additional Properties

- **Addition and Scalar Multiplication**
  - Commutativity of addition:
  $$A + B = B + A$$
  - Associativity of addition:
  $$A + (B + C) = (A + B) + C$$
  - Distributive property for scalar multiplication:
  $$c(A + B) = cA + cB$$
  - Distributive property for scalars:
  $$(c + d)A = cA + dA$$
  - Multiplication by scalar 1:
  $$1A = A$$
- **Matrix Multiplication**
  - Associativity:
  $$A(BC) = (AB)C$$
  - Distributive Property:
  $$A(B + C) = AB + AC \quad \text{and} \quad (A + B)C = AC + BC$$
  - Not Commutative:
  $$AB \neq BA \quad \text{in general (Matrix multiplication is not commutative).}$$
  - Identity Matrix (I):
  $$AI = IA = A$$
- **Transpose of a Matrix**
  - Transpose of transpose:
  $$(A^T)^T = A$$
  - Transpose of addition:
  $$(A + B)^T = A^T + B^T$$
  - Transpose of scalar multiplication:
  $$(cA)^T = cA^T$$
  - Transpose of product:
  $$(AB)^T = B^T A^T \quad \text{(Order reverses under transpose).}$$

There are more Properties that will be covered in later lectures

Real world examples

- Matrices are widely used in AI because they provide an efficient way to store and manipulate data in areas such as optimization, neural networks, and computer vision and many more.

3 Indivertible maps and Matrices

Definition 11 *$T \in \mathcal{L}(V,W)$ is called invertible if there exist a linear map $S \in \mathcal{L}(w,v)$ such that*

$$S \circ T = Id_v \text{ and } T \circ S = Id_w$$

The map S is called the inverse of T, denoted by T^{-1}

Remark 12 *Inverse maps **exist** and are **unique***

Proposition 13 *A linear map is invertible if and only if it is bijective*

Proof: "$\Rightarrow$" invertible $\Rightarrow$ injective:

$$\text{suppose } T(u) = T(v).$$

$$\text{Then } u = T^{-1}(T(u))$$
$$= T^{-1}(T(v)) = v$$
$$\Rightarrow u = v \Rightarrow \text{injective} \Rightarrow \text{surjective:}$$
$$w \in W \text{Then } w = T(T^{-1}(w)$$
$$\Rightarrow w \in \text{range of } T \Rightarrow \text{surjective}$$

"$\Leftarrow$" injective and surjective $\Rightarrow$ invertible Let $w \in W$ There exists *unique* $\vec{v} \in V$ s.t $T(u) = w$

Define the mapping: $S(w) = \vec{v}$. Clearly have $T \circ S = Id$ let $\vec{v} \in V$. Then

$$T((s \circ T)\vec{v}) = (T \circ S)(T\vec{v}) = Id \circ T\vec{v} = T\vec{v}$$

$$\Rightarrow (S \circ T)\vec{v} = \vec{v}$$
$$\Rightarrow S \circ T = Id$$
$$\Rightarrow \text{S is inverse of T}$$

Still need to show S us a linear mapping. Let $Y_1, Y_2 \in W, \alpha \in F$:

$$S(Y_1 + Y_2) = S(Y_1) + S(Y_2) \text{ and } S(\alpha Y_1) = \alpha S(Y_1)$$

Let $x_1, x_2 \in V$ s.t $T(x_i) = y_i$ Then $S(y_i) = x_i$

$$\begin{aligned} S(Y_1 + Y_2) &= S(T(x_1) + T(x_2)) \\ &= S(T(x_1 + x_2)) \\ &= x_1 + x_2 \\ &= S(Y_1) + S(Y_2) \end{aligned}$$

$$\begin{aligned} S(\alpha Y_1) &= S(\alpha T(x_1)) \\ &= S(T(\alpha x_1)) \\ &= \alpha x_1 \\ &= \alpha S(x_1) \end{aligned}$$

$$\Rightarrow S \text{ is a linear transformation}$$

□

Example using Gaussen elemation

$$A = \begin{pmatrix} 4 & 7 \\ 2 & 6 \end{pmatrix} \quad \text{Augment with the identity matrix:} \quad \left(\begin{array}{cc|cc} 4 & 7 & 1 & 0 \\ 2 & 6 & 0 & 1 \end{array}\right)$$

$$\text{Step 2: Divide the first row by 4:} \quad \left(\begin{array}{cc|cc} 1 & \frac{7}{4} & \frac{1}{4} & 0 \\ 2 & 6 & 0 & 1 \end{array}\right)$$

$$\text{Step 3: Subtract 2 times the first row from the second row:} \quad \left(\begin{array}{cc|cc} 1 & \frac{7}{4} & \frac{1}{4} & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 1 \end{array}\right)$$

$$\text{Step 4: Multiply the second row by } \frac{2}{5}: \quad \left(\begin{array}{cc|cc} 1 & \frac{7}{4} & \frac{1}{4} & 0 \\ 0 & 1 & -\frac{1}{5} & \frac{2}{5} \end{array}\right)$$

$$\text{Step 5: Subtract } \frac{7}{4} \text{ times the second row from the first row:} \quad \left(\begin{array}{cc|cc} 1 & 0 & \frac{3}{10} & -\frac{7}{10} \\ 0 & 1 & -\frac{1}{5} & \frac{2}{5} \end{array}\right)$$

$$\text{The inverse is:} \quad A^{-1} = \begin{pmatrix} \frac{3}{10} & -\frac{7}{10} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix}$$

4 Inverse Matrices

Definition 14 *A square matrix $A \in F^{nxn}$ is invertible if there exist a square matrix $B \in F^{nxn}$ such that: $A \times B = B \times A =$ Identity matrix $= Id$*

The matrix B is called the inverse matrix , and is denoted by A^{-1}

Proposition 15 *The inverse matrix represents the inverse of the corresponding linear map, that is:* $T : V \to V$

$$\underbrace{M(T^{-1})}_{\text{matrix of invers map}} = \underbrace{(M(I))^{-1}}_{\text{inverse matrix of the original map}}$$

In particular, a matrix in Invertible if and only if the corresponding map is Invertible.

Remark 16

- *The inverse matrix does not always exist*
- $(A^{-1})^{-1} = A; (A \times B)^{-1} = B^{-1} \times A^{-1}$
- A^t *invertible* $\Leftrightarrow$ A *invertible*
$$(A^t)^{-1} = (A^{-1})^t$$
- $A \in F^{nxn}$ *invertible* $\Leftrightarrow \operatorname{rank}(A) = n$
- *The set of all invertible matrices is called general linear group:*
$$GL(n, F) = A \in F^{nxn} | A \textit{ is invertible}$$

Additional Properties

Inverse of a Matrix

- If A is invertible:
$$A^{-1}A = AA^{-1} = I$$
- Double Inverse Property:
$$(A^{-1})^{-1} = A$$

Take inverse by hand $2x2$ matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \underbrace{\frac{1}{ad - bc}}_{\text{determinant of a } 2x2 \text{ matrix}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example

$$\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}^{-1} = \frac{1}{4 \times 6 - 7 \times 2} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

Remark 17 *Anything bigger than a 2x2 matrix look into using Gauss-Jordan elimination method.*

References

https://www.math.ucdavis.edu/ anne/WQ2007/mat67-Lh-Linear$_M$*aps.pdf*

https://math.mit.edu/ dyatlov/54summer10/matalg.pdf

https://en.wikipedia.org/wiki/Linear$_m$*ap*