Lecture 8

Linear Algebra

Goal: Develop tools to approximate complicated (potentially non-linear) functions with linear or quadratic functions.

- **Basis** (Ref.: 3.1)

Definition 1. A vector space V is a collection of objects called vectors, which may be added together and multiplied by real numbers, called scalars, satisfying:

(1) Associativity of $+$: $\forall x, y, z \in V$, $(x+y)+z = x+(y+z)$;
(2) Commutativity of $+$: $\forall x, y \in V$, $x+y = y+x$;
(3) Existence of zero: $\exists! \bar{0} \in V$ s.t. $\forall x \in V$, $x + \bar{0} = \bar{0} + x = x$;
(4) Existence of a vector additive inverse: $\forall x \in V$ $\exists! (-x) \in V$ s.t. $x + (-x) = \bar{0}$;
    (We define $x - y := x + (-y)$.)
(5) Distributivity of scalar multiplication over vector addition:
    $\forall \alpha \in \mathbb{R}$, $x, y \in V$, $\alpha \cdot (x+y) = \alpha \cdot x + \alpha \cdot y$;
(6) Distributivity of scalar multiplication over scalar addition:
    $\forall \alpha, \beta \in \mathbb{R}$, $x \in V$, $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$;
(7) Associativity of multiplication: $\forall \alpha, \beta \in \mathbb{R}$, $x \in V$, $(\alpha \cdot \beta) x = \alpha \cdot (\beta \cdot x)$;
(8) Multiplicative identity: $\forall x \in V$, $1 \cdot x = x$.

Example.

- $\mathbb{R}^n$ is a vector space;
- $M_{m \times n}$ (set of all $m \times n$ matrices) is a vector space;
- $\mathbb{R}^X$ (set of all functions $f : X \to \mathbb{R}$) is a vector space;
- $C(X)$ (set of all continuous functions $f : X \to \mathbb{R}$) is a vector space;
- $B(X)$ (set of all bounded functions $f : X \to \mathbb{R}$) is a vector space;
- $\mathbb{R}_+ = [0, \infty)$ is not a vector space (if $x \in (0, \infty)$, then $-x = -1 \cdot c \notin [0, \infty)$).

Definition 2. Let V be a vector space. A linear combination of $x_1, \ldots, x_n \in V$ is a vector of the form

$$y = \sum_{i=1}^{n} \alpha_i x_i, \text{ where } \alpha_1, \ldots, \alpha_n \in \mathbb{R},$$

α_i is called the coefficient of x_i in the linear combination.

We call a linear combination $\sum_{i=1}^{n} \alpha_i x_i$ nontrivial if $\sum_{i=1}^{n} \alpha_i^2 \neq 0$. That is, at least one coefficient is non-zero.

Definition 3. Let W be a subset of V. A span of W is the set of all linear combinations of elements of W,

$$\operatorname{span} W = \left\{ \sum_{i=1}^{n} \alpha_i x_i \mid n \in \mathbb{N},\ \alpha_1, \ldots, \alpha_n \in \mathbb{R},\ x_1, \ldots, x_n \in W \right\}.$$

The set $W \subset V$ spans V if $V = \operatorname{span} W$.

Definition 4. A set $X \subset V$ is linearly dependent if $\exists x_1, \ldots, x_n \in X,\ \alpha_1, \ldots, \alpha_n \in \mathbb{R}$ s.t.

$$\sum_{i=1}^{n} \alpha_i^2 \neq 0 \text{ and } \sum_{i=1}^{n} \alpha_i x_i = \bar{0}$$

A set $X \subset V$ is linearly independent if it is not linearly dependent. I.e., if $\forall x_1, \ldots, x_n \in X$

$$\sum_{i=1}^{n} \alpha_i x_i = \bar{0} \Leftrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

Definition 5. A basis of a vector space V is a linearly independent set of vectors in V that spans V.

Remark 6. To emphasize the fact that only finite linear combinations are allowed, sometimes the basis is called *Hamel* basis. In spaces where basis is infinite, it is more common to work with *Schauder* basis, which allows for infinite linear combinations. This, of course, requires the notion of convergence of infinite sums (when the sum is well-defined). If you have heard about the Fourier basis, then it is a Schauder basis and not a Hamel basis.

Importance: Basis allows us to write every element as a unique linear combination. Thus, if we know a basis, then we know V.

Example.

- $\{(1,0),(0,1)\}$ is a basis of $\mathbb{R}^2$:

$$\alpha_1(1,0) + \alpha_2(0,1) = (\alpha_1, \alpha_2) = (0,0) \Leftrightarrow \alpha_1 = \alpha_2 = 0,$$

$$\forall x = (x_1, x_2) \in \mathbb{R}^2,\ x = x_1(1,0) + x_2(0,1).$$

- $\{(1,1),(-1,1)\}$ is a basis of $\mathbb{R}^2$:

$$\alpha_1(1,1) + \alpha_2(-1,1) = (\alpha_1 - \alpha_2, \alpha_1 \alpha_2) = (0,0) \Leftrightarrow \alpha_1 = \alpha_2 = 0,$$

$$\forall x = (x_1, x_2) \in \mathbb{R}^2,\ x = \frac{x_1 + x_2}{2}(1,1) + \frac{x_2 - x_1}{2}(-1,1).$$

- $\{(1,0)\}$ is not a basis of $\mathbb{R}^2$:

$$\operatorname{span}\{(0,1)\} = \{(x,0) \mid x \in \mathbb{R}\} \neq \mathbb{R}^2.$$

- $\{(1,0),(0,1),(1,1)\}$ is <u>not</u> a basis of $\mathbb{R}^2$. It is linearly dependent:
$$(1,0)+(0,1)-(1,1)=(0,0).$$

Theorem 7. *Let B be a basis for V and enumerate elements of B by a set Λ so that $B=\{v_\lambda \mid \lambda\in\Lambda\}$. Then every vector $x\in V$ has a unique representation as a linear combination of elements of B with finitely many nonzero coefficients.*

Proof. Let $x\in V$. Since B is a basis for V, span $B=V$, and
$$x=\sum_{\lambda\in\Lambda}\alpha_\lambda v_\lambda,$$
where only finite subset of coefficients $\{\alpha_\lambda\}$ is nonzero. Thus, at least one such representation exists.

Suppose that there exists another representation
$$x=\sum_{\lambda\in\Lambda}\beta_\lambda v_\lambda,$$
where, again, only finite subset of coefficients $\{\beta_\lambda\}$ is nonzero.

Thus,
$$\sum_{\lambda\in\Lambda}\alpha_\lambda v_\lambda-\sum_{\lambda\in\Lambda}\beta_\lambda v_\lambda=x-x=\bar{0}$$
or
$$\sum_{\lambda\in\Lambda}(\alpha_\lambda-\beta_\lambda)v_\lambda=\bar{0},$$
where, again, only finite subset of coefficients $\{\alpha_\lambda-\beta_\lambda\}$ is nonzero. Since B is a basis and $\sum\limits_{\lambda\in\Lambda}(\alpha_\lambda-\beta_\lambda)v_\lambda$ is its linear combination, we must have $\alpha_\lambda\equiv\beta_\lambda$. Thus, representation is unique. □

Remark 8. The representation of $\bar{0}$ has all zero coefficients.

The following properties presented without proofs hold for vector spaces.

Theorem 9. *Every vector space has a basis. Any two bases of a vector space V have the same cardinality (are numerically equivalent).*

Theorem 10. *If V is a vector space and $W\subset V$ is linearly independent, then there exists a linearly independent set B such that $W\subset B\subset \operatorname{span} B=V$.*

The second theorem says that any linearly independent set W can be extended to a basis of V.

Te first theorem says that the cardinality of V's basis B is a property of the space V, not B, and we can define the dimension of a vector space.

Definition 11. Let V be a vector space. The dimension of V, denoted $\dim V$, is the cardinality of any basis of V. If $\dim V = n$ for some $n \in \mathbb{N}$, then V is finite-dimensional. Otherwise V is infinite-dimensional.

Example.

- $\dim \mathbb{R}^n = n$, basis is a set $\{e_i\}_{i=1}^n$, where e_i has ith coordinate equal 1 and all other coordinates equal zero;
- $\dim M_{m\times n} = mn$, basis is a set $\{E_{ij}\}_{i=1\ldots m, j=1\ldots n}$, where E_{ij} is a matrix with 1 on the ith row, jth column and zeros otherwise.

In the case of finite-dimensional vector spaces we can not have too many linearly independent vectors: if $\dim V = n$, then any set of more than n vectors is linearly dependent.

Theorem 12. *Suppose* $\dim V = n \in \mathbb{N}$*. If $W \subset V$ and $|W| > n$, where $|W|$ denotes the cardinality of W, then W is linearly dependent.*

Proof. If W is linearly independent, then by Theorem 10 we can extend it to a basis B for V that contains W. Yet, $|B| \geq |W| > n = \dim V$, which is a contradiction with Theorem 9. □

Theorem 13. *Suppose* $\dim V = n \in \mathbb{N}$ *and $W \subset V$, $|W|$=n. Then*

(1) *If W is linearly independent, then* $\operatorname{span} W = V$*, so W is a basis of V;*
(2) *If* $\operatorname{span} W = V$*, then W is linearly independent, so W is a basis of V.*

Proof. (1) If $\operatorname{span} W \neq V$, then by Theorem 10 W can be extended to a basis B, so that $W \subset B \subset \operatorname{span} B = V$. Thus, $|B| > |V| = n$. However, by Theorem 12, B is linearly dependent. So $\operatorname{span} W \neq V$ and W is a basis of V.

(2) If W is not linearly independent, then it has nontrivial linear combination $\sum_{i=1}^{n} \alpha_i v_i = \bar{0}$. Thus, for $\alpha_j \neq 0$ $v_j = -\frac{1}{\alpha_j} \sum_{i\neq j} \alpha_i v_i$ and $\operatorname{span}(W \setminus v_j) = \operatorname{span} W = V$. If $W' := W \setminus v_j$ is still linearly dependent, we can repeat the procedure until we get $\tilde{W} = W \setminus v_{j_1}, \ldots, v_{j_k}$ such that $\tilde{W}$ is linearly independent (the process terminates because W has finite number of elements). Therefore, $\tilde{W}$ is a basis of V. Yet, $|\tilde{W}| < |W| = n = \dim V$ while all bases must have cardinality $= n$. Thus, we get a contradiction and If W is linearly independent, so W is a basis of V. □

- **Linear Transformations** (Ref.: 3.2)

Suppose we have a function from one vector space to another. We want to characterize functions that preserve the algebraic structure (so that, for example, $x + y$ is mapped to $f(x) + f(y)$).

Definition 14. Let X and Y be two vector spaces. We say that $T : X \to Y$ is a linear transformation if for all $x_1, x_2 \in X$, $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2)$$

Let $L(X, Y)$ denote the set of all linear transformations from X to Y.

That is, given any two vectors, the image of their sum under a linear function is equal to the sum of their images, and the image of the product of a scalar and a vector is equal to the scalar times the image of the vector. It is in this sense that we can say that a linear function preserves the algebraic structure of the vector space on which it is defined.

Theorem 15. *$L(X, Y)$ is a vector space.*

Proof. The scalar multiplication and vector addition of $T_1, T_2 : X \to Y$ are defined by

$$(\eta T_1 + \gamma T_2)(x) = \eta T_1(x) + \gamma T_2(x).$$

Let us chech that $\eta T_1 + \gamma T_2$ is a linear transfornation.

$$\begin{aligned}(\eta T_1 + \gamma T_2)(\alpha_1 x_1 + \beta x_2) &= \eta T_1(\alpha_1 x_1 + \beta x_2) + \gamma T_2(\alpha_1 x_1 + \beta x_2) \\ &= \eta(\alpha_1 T_1(x_1) + \beta T_1(x_2)) + \gamma(\alpha_1 T_2(x_1) + \beta T_2(x_2)) \\ &= \alpha_1(\eta T_1(x_1) + \gamma T_2(x_1)) + \beta(\eta T_1(x_2) + \gamma T_2(x_2)) \\ &= \alpha_1(\eta T_1 + \gamma T_2)(x_1) + \beta(\eta T_1 + \gamma T_2)(x_2),\end{aligned}$$

so $\eta T_1 + \gamma T_2 \in L(X, Y)$.

We also need to check all of the vector space axioms, which is straightforward. □

Hence, every linear combination of linear functions is a linear function. Moreover, a composition of two linear functions is linear.

Theorem 16. *If $R : X \to Y$ and $S : Y \to Z$ are linear transformations, then $S \circ R : X \to Z$ is a linear transformation.*

Proof.

$$\begin{aligned}S \circ R(\alpha_1 x_1 + \beta x_2) &= S(R(\alpha_1 x_1 + \beta x_2)) = S(\alpha_1 R(x_1) + \alpha_2 R(x_2)) \\ &= \alpha_1 S(R(x_1)) + \alpha_2 S(R(x_2)) = \alpha_1 (S \circ R)(x_1) + \alpha_2 (S \circ R)(x_2).\end{aligned}$$

□

Definition 17. Let $T \in L(X, Y)$. The image of T is $\operatorname{Im} T := T(X) = \{T(x) \mid x \in X\}$, the kernel of T is $\ker T := \{x \in X \mid T(x) = \bar{0}\}$, and the rank of T is $\operatorname{rank} T := \dim(\operatorname{Im} T)$.

Theorem 18. *Let X be a finite-dimensional vector space and $T \in L(X, Y)$. Then $\operatorname{Im} T$ and $\ker T$ are vector subspaces of Y and X respectively, and*

$$\dim X = \dim \ker T + \operatorname{rank} T = \dim \ker T + \dim \operatorname{Im} T.$$

Proof. Let us show that $\operatorname{Im} T$ is a vector subspaces of Y. We need to show that if $\alpha, \beta \in \mathbb{R}$ and $y_1, y_2 \in \operatorname{Im} T$, then $\alpha y_1 + \beta y_2 \in \operatorname{Im} T$. If $y_1, y_2 \in \operatorname{Im} T$, then $\exists x_1, x_2 \in X$ such that $y_1 = T(x_1), y_2 = T(x_2)$. Thus,

$$\alpha y_1 + \beta y_2 = \alpha T(x_1) + \beta T(x_2) = T(\alpha x_1 + \beta x_2).$$

Because $\alpha x_1 + \beta x_2 \in X$, $\alpha y_1 + \beta y_2 \in T(X)$, and $\operatorname{Im} T$ is a vector subspaces of Y.

Let us show that $\ker T$ is a vector subspaces of X. If $x_1, x_2 \in \ker T$, then $T(x_1) = T(x_2) = \bar{0}$. Thus,

$$T(\alpha x_1 + \beta x_2) = \alpha T(x_1) + \beta T(x_2) = \bar{0} + \bar{0} = \bar{0},$$

and $\alpha x_1 + \beta x_2) \in \ker T$. So $\ker T$ is a vector subspaces of X.

We are left with showing $\dim X = \dim \ker T + \operatorname{rank} T$. Let $V = \{v_1, \dots, v_k\}$ be a basis for $\ker T$, so that $\dim \ker T = k$ (note that $\ker T \subset X$ so $\dim \ker T \le \dim X$). If $\ker T = \{\bar{0}\}$, take $k = 0$ so $V = \emptyset$. Extend V to a basis B of X with $W = \{v_1, \dots, v_k, w_1, \dots, w_r\}$. We claim that $\{T(w_1), \dots, T(w_r)\}$ is a basis for $\operatorname{Im} T$, so that $\operatorname{rank} T = r$.

- If $y \in \operatorname{Im} T$, then $y = T(x)$ for some $x \in X$, $x = \sum_{i=1}^{k} \alpha_i v_i + \sum_{i=1}^{r} \beta_i w_i$. Thus,

$$y = T\left(\sum_{i=1}^{k} \alpha_i v_i + \sum_{i=1}^{r} \beta_i w_i\right) = \sum_{i=1}^{r} \beta_i T(w_i),$$

  as for all $i = 1, \dots, n$, $v_i \in \ker T$ so that $T(v_i) = \bar{0}$. Thus, $\operatorname{span}\{T(w_1), \dots, T(w_r)\} = \operatorname{Im} T$.
- If $\{T(w_1), \dots, T(w_r)\}$ is linearly dependent, then $\exists$ nontrivial linear combination $\sum_{i=1}^{r} \beta_i T(w_i) = \bar{0}$.

$$\bar{0} = \sum_{i=1}^{r} \beta_i T(w_i) = T\left(\sum_{i=1}^{r} \beta_i w_i\right),$$

  so that nontrivial linear combination $\sum_{i=1}^{r} \beta_i w_i \in \ker T$. However, basis of $\ker T$, V is independent of $\{w\}_i$, and we must have $\beta_i = 0$ for all $i = 1, \dots, r$.

Since W is a basis of X, $\dim X = k + r = \dim \ker T + \operatorname{rank} T$. ☐

Definition 19. $T \in L(X, Y)$ is <u>invertible</u> if there exists a function $S : Y \to X$ such that

$$S(T(x)) = x \,\forall x \in X,$$

$$T(S(y)) = y \,\forall y \in Y.$$

The transformation S is called the <u>inverse</u> of T and is denoted T^{-1}.

In other words, $S \circ T = I_X$ and $T \circ S = I_Y$, where I_X and I_Y are the identity mappings in X and Y, respectively.

When is the transformation invertible? First, T must not glue points, i.e., $\forall x_1 \neq x_2$ we must have $T(x_1) \neq T(x_2)$ (such T is called *one-to-one* or *injection*). Second, the image of T must equal Y, i.e., $T(X) = Y$ (such T is called *onto* or *surjection*). For each point $y \in Y$ we must have $x \in X$ such that $T(x) = y$, otherwise we will not get $T(S(y)) = y$.

Theorem 20. *If $T \in L(X,Y)$ is invertible, then $T^{-1} \in L(Y,X)$, i.e., T^{-1} is linear.*

Proof. Suppose that $\alpha, \beta \in \mathbb{R}$, $y_1, y_2 \in Y$. Since T is invertible, $\exists! x_1, x_2 \in X$ such that $y_1 = T(x_1), y_2 = T(x_2)$ and $T^{-1}(y_1) = x_1, T^{-1}(y_2) = x_2$. Therefore,

$$\begin{aligned} T^{-1}(\alpha y_1 + \beta y_2) &= T^{-1}(\alpha T(x_1) + \beta T(x_2)) = T^{-1}(T(\alpha x_1 + \beta x_2)) \\ &= \alpha x_1 + \beta x_2 = \alpha T^{-1}(y_1) + \beta T^{-1}(y_2), \end{aligned}$$

and $T^{-1} \in L(Y,X)$. □

Theorem 21. *$T \in L(X,Y)$ is one-to-one if and only if $\ker T = \{\bar{0}\}$.*

Proof. Suppose $T \in L(X,Y)$ is one-to-one. If $x \in \ker T$, then $T(x) = \bar{0}$. Since T is linear,

$$T(\bar{0}) = T(0 \cdot \bar{0}) = 0 \cdot T(\bar{0}) = \bar{0}.$$

Since T is one-to-one, $x = \bar{0}$, so $\ker T = \bar{0}$.

Suppose that $\ker T = \bar{0}$. If $T(x_1) = T(x_2)$, then $T(x_1 - x_2) = T(x_1) - T(x_2) = \bar{0}$. Thus, $x_1 - x_2 = \bar{0}$ and $x_1 = x_2$, so that T is one-to-one. □

- **Isomorphisms** (Ref.: 3.3)

We will now show that two vector spaces of the same dimension are "equivalent" from an algebraic point of view. Thus, all n dimensional vector spaces are "equivalent" to $\mathbb{R}^n$ and $L(X,Y)$, where $\dim X = m, \dim Y = n$, is "equivalen" to $M_{m \times n}$. What does "equivalence" mean formally?

Definition 22. Two vector spaces X and Y are isomorphic if there exists an invertible linear function (one-to-one and onto) from X to Y. A function with these properties is called an isomorphism.

Theorem 23. *Let X and Y be two vector spaces, and let $V = \{v_\lambda \mid \lambda \in \Lambda\}$ be a basis for X. Then a linear transformation $T : X \to Y$ is completely defined by its value on V, that is:*

(1) *Given any set $\{y_\lambda \mid \lambda \in \Lambda\} \subset Y$, $\exists T \in L(X,Y)$ s.t. $T(v_\lambda) = y_\lambda$ for all $\lambda \in \Lambda$.*
(2) *If $S, T \in L(X,Y)$ and $S(v_\lambda) = T(v_\lambda)$ for all $\lambda \in \Lambda$, then $S = T$.*

Proof. (1) Any $x \in X$ has a unique representation $x = \sum_{\lambda\in\Lambda} \alpha_{\lambda\in\Lambda} v_\lambda$, and only finite set of indices has non-zero coefficients α. Define $T(x) = \sum_{\lambda\in\Lambda} \alpha_\lambda y_\lambda$, which is well-defined as there are only finitely many non-zero terms. Thus, $T(x) \in Y$. Moreover, T is linear:

$$T(ax_1 + bx_2) = T\left(\sum_{\lambda\in\Lambda} a\alpha_\lambda v_\lambda + \sum_{\lambda\in\Lambda} b\beta_\lambda v_\lambda\right) = a\alpha_\lambda y_\lambda + b\beta_\lambda y_\lambda = aT(x_1) + bT(x_2).$$

(2) Suppose that $S(v_\lambda) = T(v_\lambda)$ for all $\lambda \in \Lambda$. For any $x \in X$, $x = \sum_{i=1}^{n} \alpha_i v_{\lambda_i}$. Thus,

$$S(x) = S\left(\sum_{i=1}^{n} \alpha_i v_{\lambda_i}\right) = \sum_{i=1}^{n} \alpha_i S(v_{\lambda_i}) = \sum_{i=1}^{n} \alpha_i T(v_{\lambda_i}) = T\left(\sum_{i=1}^{n} \alpha_i v_{\lambda_i}\right) = T(x).$$

□

Theorem 24. *Two vector spaces X and Y are isomorphic if and only if* $\dim X = \dim Y$.

Proof. Let $T : X \to Y$ be an isomorphism and let $\{v_\lambda \mid \lambda \in \Lambda\}$ be a basis of X. Then one can check that $\{T(v_\lambda) \mid \lambda \in \Lambda\}$ is a basis of Y. Hence, $\dim X = |\Lambda| = \dim Y$.

In the opposite direction, suppose that $\dim X = \dim Y$. Then we can enumerate bases of spaces X and Y by the same set Λ. Thus, we choose a basis $\{v_\lambda \mid \lambda \in \Lambda\}$ for X and a basis $\{u_\lambda \mid \lambda \in \Lambda\}$ for Y.

We define the map T by formula:

$$T\left(\sum_{\lambda} \alpha_\lambda v_\lambda\right) = \sum_{\lambda} \alpha_\lambda u_\lambda.$$

Following the proof of Theorem 23, T is linear. Also T is invertible, since its inverse is explicit:

$$T^{-1}\left(\sum_{\lambda} \alpha_\lambda u_\lambda\right) = \sum_{\lambda} \alpha_\lambda v_\lambda.$$

We conclude that T gives an isomorphism between X and Y. □