Linear Algebra Lecture Notes

MITZY

Last Update : 2026.02.16

Contents

1 Logic

수학에서는 객관적 '참'과 '거짓'이 명확한 문장 혹은 식을 다루며, 이를 '명제' (Statement, Proposition)라고 한다. 어떤 명제가 모든 가능한 경우에서 참이면 참인 명제라고 하며, 그렇지 않으면 (즉, 한 가지 경우라도 옳지 않을 때가 있으면) 거짓인 명제이다. 이때, 참과 거짓을 명제의 '진리값'이라고 한다. 합성 명제란 두 가지 이상의 명제가 합쳐져서 하나의 명제를 이루는 경우를 의미하는데, 두 명제의 참과 거짓, 접속어 (and, or, if then , if and only if, etc,.) 등에 따라 진리값이 결정된다. 간단하게 살펴보자. 하나의 명제 P가 주어졌을 때, 'P가 아니다.' ($\sim P$)라는 명제를 생각해볼 수 있다. 이를 '부정 명제'라고 한다. 만약, P가 참이면 $\sim P$는 거짓이고, P가 거짓이면, $\sim P$는 참이다. 'P이고 Q이다.'라는 'and'로 연결된 합성 명제를 보자. 이는 P가 참이고 Q도 참이어야 참인 명제이다. 'P또는 Q이다.'라는 'or'로 연결된 합성 명제는 둘 다 거짓일 때만 거짓이다. 즉, P가 거짓이고 Q도 거짓일 때만, 거짓이다. '만약 P라면, Q이다.'란 명제를 보자. 이는 P가 참이고, Q는 거짓일 때만, 거짓이다. 이때, P를 조건/가정, Q를 결론이라 하는데, 조건/가정이 성립할 때, 결론이 성립하지 않는 '반례' (Counter example)를 찾아서 명제를 부정할 수 있다. 이를 또한 필요조건 (Necessary condition)과 충분조건 (Sufficient condition)을 통해 이야기할 수 있다. 필요조건은 어떤 명제가 참이기 위해서는 반드시 성립해야 하는 조건을 이야기한다. 충분조건은 그 조건이 만족되면 명제가 참이 되는 조건을 이야기한다. 앞에 '만약 P라면, Q이다.' ($P \to Q$)라는 명제를 살펴보자. 이 명제에서 Q는 P이기 위한 필요조건이다. 반대로, P는 Q의 충분조건이다.

- 만약 a가 자연수라면, a는 실수이다.

여기서 살펴보자. 위 명제는 참인 명제이다. 집합 (벤 다이어그램)으로 표현해보면 심플하다. 자연수 집합은 실수 집합의 부분집합이다. 자연수가 되려면, 실수인 것은 필요하다. 즉, 실수가 아니면 자연수가 절대 될 수 없다. 하지만, 자연수가 되려면 실수인 것으로 충분하진 않다. 실수일지라도, 자연수가 아닌 수는 무수히 많기 때문이다. 즉, 충분조건은 아니다. 하지만, 자연수라면, 그 수는 반드시 실수이다. 어떤 수가 자연수라는 조건을 만족하면, 실수라는 조건을 만족하기에 충분하다.

'만약 P이면 Q이다.'라는 명제와 '만약 Q이면 P이다.'라는 명제를 간단하게 "만약 P, 그리고 오직 P일 때만 Q이다" ($P \Leftrightarrow Q$)라고 서술한다. (if and only if) 이때 두 명제는 논리적 동치 명제이다. 그리고 이때 두 명제는 서로 필요충분조건 관계이다.

수학적 증명법에는 크게 3가지가 있는데, '연역 추론 논증' (=직접 증명법), '귀류법' (Proof by contradiction), '수학적 귀납법', '대우 명제 이용' (Proof by contrapositive)이 있다. 연역 추론 논증은 참인 명제에서 시작하여, 참인 명제들을 차례로 유도하여 원하는 결론을 보이는 방법이고, 귀류법은 결론을 부정하여 가정에 모순이 생김을 보이는 방식으로 원하는 결론을 보이는 방법이다. 이는 '배중률' (Law of Excluded Middle) 원리에 기반하는데, 어떤 명제와 그 부정 중 하나는 반드시 참이므로, 부정이 모순을 일으키면 원래 명제는 참이라는 사실이다.

예를 들어, $\sqrt{2}$가 유리수가 아님을 증명할 때 귀류법을 사용할 수 있다.

Proof (귀류법). **¡1¿ Assume.** $\sqrt{2}$가 유리수라고 하자. 그러면 서로소인 정수 n, m ($m \neq 0$)가 존재하여 $\sqrt{2} = \frac{n}{m}$로 쓸 수 있다.

¡2¿ 양변을 제곱하면 $2 = \frac{n^2}{m^2}$이므로 $2m^2 = n^2$이다.

¡3¿ Claim. n은 짝수이다. **Reason.** n^2가 짝수이면 n도 짝수이기 때문이다(홀수의 제곱은 홀수).

¡4¿ 따라서 어떤 정수 k가 존재하여 $n = 2k$이고, 이를 $2m^2 = n^2$에 대입하면 $2m^2 = 4k^2$, 즉 $m^2 = 2k^2$이다.

¡5¿ 위와 같은 논리로부터 m도 짝수임을 얻는다.

¡6¿ 그러면 n과 m은 공약수 2를 가지게 되어 서로소라는 가정과 모순이다. 따라서 $\sqrt{2}$는 유리수가 아니다(즉, 무리수이다).

$\square$

다음으로, 어떤 명제를 반증하고 싶다면, 해당 명제의 '반례'를 찾으면 된다. 예를 들어, '모든 소수는 홀수이다.'라는 명제를 반증하고 싶다고 해보자. 그러면, 이 명제의 반례 (소수이면서, 홀수가 아님)를 찾으면 된다. 2는 소수이면서 짝수이므로, 위 명제의 반례이다. 따라서 위 명제는 거짓이다.

수학적 귀납법은 다음과 같은 증명 방법이다. 각 자연수 $n \in \mathbb{N}$에 대해, 어떤 명제 $P(n)$이 참이면, 모든 자연수 n에 대해, 어떤 명제 $P(n)$이 참이므로, (이는 자연수 공리계인 페아노 공리에 의해 성립한다) 다음과 같이 명제를 증명할 수 있다:

- $P(1)$이 참이다. (Base case)
- 어떤 자연수 k에 대해 $P(k)$가 참이라고 가정하고 (Induction hypothesis), $P(k+1)$이 참임을 보인다. (Induction)

수학적 귀납법은 증명에서 자주 사용되는 논리이다. 이름에서 '귀납'이란 표현이 들어가서 과학의 '귀납'과 헷갈릴 수 있는데 이는 자연수의 페아노 공리가 성립하기 때문에 성립하는 증명 방법이고, 따라서 엄밀하게는 '연역 논증'에 해당한다.

모든 자연수 n에 대해, $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$임을 수학적 귀납법을 이용해 증명하자.

Proof (수학적 귀납법). **¡1¿ Goal.** 모든 자연수 n에 대해 $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$임을 보인다.

¡2¿ (기초 단계) $n = 1$이면 $\sum_{k=1}^{1} k = 1 = \frac{1\cdot 2}{2}$이므로 성립한다.

¡3¿ (귀납 가정) 어떤 $n \in \mathbb{N}$에 대해 $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$라고 가정하자.

¡4¿ (귀납 단계) 그러면

$$\begin{aligned}\sum_{k=1}^{n+1} k &= \left(\sum_{k=1}^{n} k\right) + (n+1) = \frac{n(n+1)}{2} + (n+1) \\ &= (n+1)\left(\frac{n}{2}+1\right) = \frac{(n+1)(n+2)}{2}\end{aligned}$$

이므로 $n+1$에 대해서도 성립한다.

¡5¿ 따라서 수학적 귀납법에 의해 모든 $n \in \mathbb{N}$에 대해 성립한다.

□

1.1 강한 수학적 귀납법

수학적 귀납법에는 '강한 귀납법' (Strong induction)이라는 변형도 존재한다. 일반적인 귀납법에서는 $P(k)$가 참이라고 가정하고 $P(k+1)$이 참임을 보이지만, 강한 귀납법에서는 $P(1), P(2), \cdots, P(k)$가 모두 참이라고 가정하고 $P(k+1)$이 참임을 보인다. 강한 귀납법의 대표적인 응용은 산술의 기본 정리 (모든 2 이상의 자연수는 소수의 곱으로 유일하게 표현된다)의 존재성 증명이다.

대우 명제 증명이란 '만약 A이면, B이다.'의 진리값과 대우 명제 '만약 B가 아니면, A가 아니다.'의 진리값이 같다는 것을 이용하는 것이다. (논리적 동치) 즉, 대우 명제의 참/거짓을 판단하는 방법으로 원래 명제의 참/거짓을 판단하는 증명 방법이다.

1.2 구성적 증명과 비구성적 증명

수학에서는 '존재성'을 증명할 때, 두 가지 접근법이 있다. 구성적 증명 (Constructive proof)은 해당 대상을 직접 만들어서 존재함을 보이는 방법이고, 비구성적 증명 (Non-constructive proof)은 존재하지 않는다고 가정하면 모순이 발생함을 보이는 방법이다. 귀류법을 통한 존재성 증명은 대표적인 비구성적 증명이다.

마지막으로 $\forall, \exists, s.t.,$ 등등에 대해 살펴보고 집합론으로 넘어간다.

- $\forall$: '모든' (For all, For every) 을 의미하는 논리 기호이며, '임의의'로도 읽는다.
- $\exists$: '존재한다' (there exists) 를 의미하는 논리 기호이며, '적어도 하나'로도 읽는다.
- $\exists!$: '유일하게 존재한다' (there exists a unique)를 의미한다.
- $s.t.,$: Such that의 약자이며, 조건을 서술할 때 사용한다.
- $i.e.,$: Id est의 약자이며, '즉', '다시 말해서'라는 의미이다.
- $e.g.,$: For example의 약자이며, '예를 들어'라는 의미이다.
- $etc.$: et cetra의 약자이며, '등등' 이라는 의미이다.
- □ 또는 ■ : 증명 완료를 나타내는 기호이다.

사실, 이런 논리 기호를 만든 것은 수학에서 관심을 갖는 요소가 '존재성'과 '유일성', '보편성'이기 때문이다. '존재성'이란 어떤 성질/식을 만족하는 대상이 존재하는가?라는 의미이고, '유일성'이란 그 대상이 '유일한가?'라는 의미이다. '보편성'이란 가급적 넓은 대상에게 일관되게 성립하는 성질인가?라는 의미이다. 예를 들어, 함수의 최댓값과 최솟값을 생각해보자. 제일 먼저 생각할 것은 주어진 함수의 최댓값과 최솟값이 '존재하는가?'이다. 만약, 그것이 존재한다면, 그 다음 생각할 것은 '유일한가?'이다. 유일하지 않다는 것은 '최댓값' (또는 최솟값)이 2개 이상 존재한다는 것이다. 보편성이란 그 최댓값과 최솟값이 보편적으로 갖는 성질을 의미한다. 예를 들어, '미분 가능한 함수'의 최댓값 (또는 최솟값)에서는 미분계수가 0이라는 '보편성'이 성립한다. (= 페르마 정리) 주로, 어떤 성질을 만족하는 대상들이 모인 그 집합의 크기가 크면 클수록 더 보편적인 명제라고 볼 수 있다. '정의'는 보편성을 만족해야 하고, '정리' (theorem : 수학적 참인 명제)도 가능한 보편적인 것이 더 좋은 정리이다.

2 Basic Set theory

집합은 보통 대문자 $A, B, \cdots$로 표기하고, 그 집합의 원소는 소문자 $a, b, \cdots$로 표기한다. a가 집합 A의 원소이면 다음과 같이 표기한다.

$$a \in A$$

반대로, a가 집합 A의 원소가 아니면 $a \notin A$로 표기한다.
집합 어떤 전체집합(Universal set) U가 고정되어 있고 $A \subseteq U$라 하자. A의 여집합 (complement)은 $A^c := U \setminus A = \{x \in U : x \notin A\}$로 정의한다.
원소가 없는 집합을 공집합이라 하고, $\emptyset$로 표기한다.
원소의 개수가 유한한 집합을 '유한 집합' (Finite set)이라 하고, 무한한 집합을 '무한 집합' (Infinite set)이라 한다.
두 집합 A, B가 있을 때, A의 모든 원소가 B의 원소이면, A를 B의 '부분집합' (subset)이라 하고 다음과 같이 표기한다:

$$A \subseteq B.$$

또한, $A \subseteq B$ 이고 (and), $B \subseteq A$이면, 두 집합은 같다 ($A = B$).
참고로, 공집합은 모든 집합의 부분집합이다. 즉, 임의의 집합 A에 대해 $\emptyset \subseteq A$이다. 하지만 이는 $\emptyset \in A$를 의미하지 않는다. 예를 들어 $A = \{1\}$이면 $\emptyset \notin A$이다.
집합을 표기하는 것에는 두 가지 방법이 있는데, 모든 원소를 적어주거나, 원소의 조건을 적어주는 방법이 있다.

$$\begin{aligned} A &= \{1, 2\} \\ B &= \{x : \text{x is an even natural number}\} \end{aligned}$$

또한, 임의의 집합 A의 부분집합들의 전체 모임을 '멱집합' (Power set)이라고 하고 $P(A)$ 또는 2^A로 표기한다. 예를 들어, 집합 $A = \{1, 2\}$의 멱집합은 다음과 같다.

$$\{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$$

유한 집합 A의 원소 개수가 n이면, $|P(A)| = 2^n$이다.

$$A \cup B : \{x : x \in A \text{ or } x \in B\} \tag{1}$$
$$A \cap B : \{x : x \in A \text{ and } x \in B\} \tag{2}$$
$$A - B = A \cap B^c : \{x : x \in A \text{ and } x \notin B\} \tag{3}$$
$$A \triangle B = (A - B) \cup (B - A) : \text{대칭차집합 (Symmetric difference)} \tag{4}$$

(1)은 합집합 (Union)을 의미하고, A의 원소 또는 B의 원소인 x들의 집합이다.
수학에서 '또는'은 둘 중 하나만 만족하면 충분하기 때문에, A의 원소이면서 B의 원소인 것도 포함된다.
(2)는 교집합 (Intersection) 을 의미하고, A의 원소이고, B의 원소인 x들의 집합이다.
(3)은 차집합 (Set difference) 을 의미하고, A의 원소이면서, B의 원소는 아닌 x들의 집합이다.
(4)는 대칭차집합으로, A 또는 B에는 속하지만 둘 모두에 속하지는 않는 원소들의 집합이다.
교집합이 공집합인 두 집합, $A \cap B = \emptyset$, 을 우리는 '서로소' (disjoint) 라고 한다. 또한, 수학자들은 편의상 다음이 성립한다고 약속하였다.

$$\begin{aligned} A \cup \emptyset &= A \\ A \cap \emptyset &= \emptyset \end{aligned}$$

다음은 집합 연산의 기본적인 공식들이다. 드모르간의 법칙과 분배 법칙을 서술한다.

Theorem 2.1.

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \end{aligned}$$

벤 다이어그램을 그려보면 위 등식이 항상 성립함을 알 수 있다. 하지만 꼭 두 집합의 합집합, 교집합만 고려할 필요는 없다. 유한한 개수 또는 무한한 개수의 집합들의 합집합, 교집합도 수학에서 자주 등장한다. 이를 위해서 우리는 index를 도입한다.
그 전에 집합의 크기를 측정하는 방법 중 대표적인 방법 하나는 원소의 개수를 세는 것이다. 예를 들어, 다음의 유한

집합 A의 원소의 개수는 4이므로 크기, $|A| = 4$이다.
다른 말로 이를 '기수' (Cardinality)라고 한다.

$$A = \{-1, 0, 1, 2\}$$

가장 대표적인 무한집합은 '수', 자연수 $\mathbb{N}$, 정수 $\mathbb{Z}$, 유리수 $\mathbb{Q}$, 실수 $\mathbb{R}$이다. 이들은 원소의 개수가 모두 무한하기 때문에 무한집합의 대표적인 예시이다.
하지만, 수학이 발전하면서 무한을 두 종류로 구분하는데 '셀 수 있는 무한' (Countably infinite)과 '셀 수 없는 무한' (Uncountably infinite)으로 나뉜다.
자연수, 정수, 유리수는 그 크기가 같은 무한집합이고 이들은 셀 수 있는 무한집합 (Countably infinite)이다. 실수는 셀 수 없는 무한집합 (Uncountably infinite)이고, 이 두 무한집합을 구분한 것이 칸토어의 집합론이다. 이에 대한 정의는 뒤의 '함수' 부분에서 살펴볼 예정이다.
가산 집합, Countable set, 이라고 하면 유한 집합 (Finite set)과 자연수 집합과 크기가 같은 무한 집합 (Countably infinite set)을 포함하는 집합이고, 비가산 집합 (Uncountable set)은 실수집합 (혹은 그 부분집합)과 크기가 같은 무한 집합을 나타내는 표현이다.
예를 들어 다음과 같은 표기가 있다고 하자.

$$\cup_{n\in\mathbb{N}} A_n := \{x : x \in A_n \text{ for at least one } n \in \mathbb{N}\}$$
$$\cap_{n\in\mathbb{N}} A_n := \{x : x \in A_n \text{ for every } n \in \mathbb{N}\}$$

위의 집합은 인덱스 집합이 자연수 집합이므로 countably infinite 개수의 집합 A_n들이 있고, 그것의 합집합 (또는 교집합)을 정의하고 있는 것이다. 인덱스 집합이 유한 집합이면, 다루기 쉽다 ($e.g., \cup_{n=1}^{N} A_n$). 하지만, 인덱스 집합이 비가산 무한집합이면 이 경우에는 다루기가 상당히 어려워진다. 이때는 Axiom of choice, Zorn's lemma 등 몇 가지의 보조정리와 공리들이 필요하고 학부 수학과 기준으로 '집합론'이라는 과목에서 이러한 내용들을 다룬다. 하지만, 본 강의에서는 이에 대해 자세하게 다루지 않겠다. 우리가 주로 볼 집합족 (family/collection of sets)은 인덱스 집합이 자연수 집합이거나, 유한 집합인 경우가 거의 대다수이기 때문이다.

2.1 관계와 동치관계

집합론에서 빠뜨릴 수 없는 중요한 개념이 '관계' (Relation)이다. 두 집합 A, B의 곱집합 $A \times B$의 부분집합 R을 A에서 B로의 관계라고 한다. $(a, b) \in R$이면 $a \sim b$ 또는 aRb로 표기한다. 관계 중에서 특별한 성질을 만족하는 것이 '동치 관계' (Equivalence relation)이다.

Def. 동치 관계 : 집합 A에서의 관계 $\sim$이 다음 세 조건을 만족하면, 이를 동치 관계라고 한다:

1. 반사성 (Reflexive) : 모든 $a \in A$에 대해, $a \sim a$.
2. 대칭성 (Symmetric) : $a \sim b$이면, $b \sim a$.
3. 추이성 (Transitive) : $a \sim b$이고 $b \sim c$이면, $a \sim c$.

Example 2.1. *정수 집합 $\mathbb{Z}$에서 합동 관계 (Congruence relation)를 생각하자. 양의 정수 n에 대해, $a \equiv b \pmod{n}$은 n이 $a - b$를 나누는 것으로 정의한다. 이 관계는 동치 관계이다.*

2.2 분할

동치 관계와 밀접하게 연결된 개념이 '분할' (Partition)이다.

Def. 분할 : 집합 A의 분할이란 다음 조건을 만족하는 A의 부분집합들의 모임 $\{A_i\}_{i\in I}$이다:

1. 각 $A_i \neq \emptyset$.
2. $i \neq j$이면 $A_i \cap A_j = \emptyset$ (서로소).
3. $\bigcup_{i\in I} A_i = A$ (합집합이 전체).

동치 관계와 분할 사이에는 다음과 같은 근본적인 대응 관계가 있다.

Theorem 2.2. *집합 A의 동치 관계 $\sim$이 주어지면, 동치류 $[a] = \{b \in A : a \sim b\}$들의 모임은 A의 분할을 이룬다. 역으로, A의 분할이 주어지면, '같은 부분집합에 속하는 관계'는 동치 관계를 정의한다.*

이 정리는 선형대수학에서 매우 중요한데, 예를 들어 '닮음' (Similarity) 관계는 정사각행렬들의 집합에서 동치 관계이며, 닮음 관계에 의한 동치류는 동일한 선형연산자를 다른 기저에서 표현한 행렬들의 모임이 된다. 또한, 벡터 공간의 동형 관계도 동치 관계이며, 같은 차원의 벡터 공간들이 하나의 동치류를 이룬다.

다음은 집합의 '곱'이다. 이 개념을 활용해서 우리는 $\mathbb{R}$의 곱집합으로서 $\mathbb{R}^n$을 정의해볼 수 있다. 여기서의 곱은 정확히는 Cartesian product라고 하는데, 두 집합 A, B가 주어져 있을 때, 이것의 곱집합 $A \times B$는 다음을 만족하는 순서쌍 (a, b)의 집합으로 정의한다:

$$A \times B := \{(a, b) : a \in A \text{ and } b \in B\}.$$

예를 들어, $A = \{1, 2\}$, $B = \{-1, -2\}$이면, 이들의 곱집합은 다음과 같다:

$$A \times B = \{(1, -1), (1, -2), (2, -1), (2, -2)\}.$$

이러한 방식으로 우리는 $\mathbb{R}^2$를 정의해볼 수 있다:

$$\mathbb{R}^2 := \mathbb{R} \times \mathbb{R} = \{(a, b) : a \in \mathbb{R} \text{ and } b \in \mathbb{R}\}.$$

다음은 함수이다.

3 Functions

수학에서 제일 중요한 두 가지를 뽑으라고 하면 집합과 함수이다. 사실 모든 수학의 과목들이 집합과 함수에 대해서 배우는 것이다. 예를 들어, 선형대수학은 벡터 공간이라는 집합과 두 벡터 공간을 사상하는 함수인 선형 변환에 대해서 배우는 것이고, 확률론은 확률 공간이라는 집합과 두 확률 공간을 사상하는 함수인 확률 변수에 대해서 배우는 것이고, 미적분학은 거리 공간이라는 집합과 두 거리 공간을 사상하는 함수의 연속, 미분, 적분에 대해서 배우는 것이다. 이처럼 집합과 함수는 결국 수학의 모든 것이다. 함수의 정확한 정의는 다음과 같다:
function: 두 집합 A, B가 주어져 있고, 이들의 곱집합 $A \times B$을 생각하자. 이 곱집합의 부분집합이 다음의 조건을 만족하면, 이를 함수라고 한다:

$$[(a, b) \text{ and } (a, b')] \to [b = b']$$

이때, a와 b 사이에는 어떤 관계 (relation)가 성립하는데, 이 관계를 곧 함수(즉, $f(a) = b$) 라고 한다. 또한, A를 정의역 (domain), B를 공역 (codomain)이라 하고, 함수 $f : A \to B$로 표기한다. 함수값들의 집합 $f(A) = \{f(a) : a \in A\}$을 f의 치역 (image 또는 range)이라 한다. 예를 들어, $\mathbb{R} \times \mathbb{R}$의 부분집합을 생각해보자. 이 부분집합이 다음의 조건을 만족한다면:

$$\{(x, x^3 + 1) : x \in \mathbb{R}\},$$

이는 함수 $f : \mathbb{R} \to \mathbb{R}$이고, $f(x) = x^3 + 1$로 적는다. 그리고 정의상, 위 함수는 $\mathbb{R}^2$의 부분집합이기 때문에, 우리는 위 함수의 '그래프'를 2차원 좌표평면에 표현할 수 있는 것이다.

Injective, Surjective, Bijective : 함수 $f : A \to B$가 다음의 조건을 만족하면 이를 단사 함수 (Injective) 라고 한다:

$$[f(a) = f(a')] \to [a = a'].$$

함수 $f : A \to B$가 다음의 조건을 만족하면 이를 전사 함수 (Surjective) 라고 한다:

$$[b \in B] \to [b = f(a) \text{ for at least one } a \in A].$$

함수 $f : A \to B$가 Injective이면서 동시에 Surjective이면 이를 Bijective (일대일 대응 함수, 전단사 함수 One-to-One correspondence) 라고 한다. 일대일 대응 함수의 가장 중요한 성질은 '역함수'가 존재한다는 것이다.

Composite function: 함수 $f : A \to B$, $g : B \to C$가 있다고 하자. 두 함수의 합성 함수 $(g \circ f)(x) : x \in A \to C$ (composite function)는 다음과 같이 정의한다:

$$\{(a, c) : \text{ For some } b \in B, f(a) = b \text{ and } g(b) = c\}.$$

합성 함수가 잘 정의되기 위해서는 반드시 f의 공역이 g의 정의역과 같거나, 그 부분집합이어야 한다.

Inverse function: 항등함수 (Identity function)를 먼저 정의하자. 모든 $x \in A$에 대해서, 함수 $I : A \to A$가 다음을 만족하면, I를 항등함수라고 한다:

$$I(x) = x$$

어떤 함수 $f : A \to B$와 $g : B \to A$가 다음을 만족하면 g를 f의 역함수 (또는 f를 g의 역함수) 라고 한다:

$$g \circ f = I_A, \qquad f \circ g = I_B.$$

즉, 모든 $a \in A$에 대해 $g(f(a)) = a$이고, 모든 $b \in B$에 대해 $f(g(b)) = b$이다. 역함수가 존재하는 함수 f를 가역함수 (Invertible function)라고 한다. 다음은 역함수의 존재성과 관련된 매우 중요한 정리이다.

Theorem 3.1. *함수 f가 역함수가 존재할 필요충분조건은 f가 일대일 대응 함수인 것이다.*

앞으로 종종 쓰일 표기법 하나를 정리하고 간다.

$$f^{-1}(B) := \{a : f(a) \in B\}$$

이를 역상 (pre-image)이라고 하는데, 편의상 자주 쓰이는 표기이다.

앞에서 무한 집합 중 가산 집합과 비가산 무한 집합을 정의함에 있어서 함수가 사용된다고 언급했는데, 이에 대해 살펴보자.

Countable set : 가산 집합에는 유한 집합과 가산 무한 집합이 있고, 이들은 다음과 같이 정의한다:
집합 A가 유한 집합이라는 것은 $|A| < |\mathbb{N}|$이다. 집합 A가 가산 무한 집합이라는 것은, $|A| = |\mathbb{N}|$이고, 일대일 대응 함수 $f : A \to \mathbb{N}$가 존재한다. 대표적으로 정수와 유리수 집합이 있다. 사실, 수학에서 두 집합 사이의 함수들을 생각했을 때, 일대일 대응 함수가 존재하는 것 (= 가역함수가 존재하는 것)은 매우 특수한 상황이다. 앞으로는 '가산 무한 집합'이라 하면 자연수 집합을 떠올리면 된다. 다음은 셀 수 있는 집합인지 판정할 때 사용되는 유용한 정리들이다.

Theorem 3.2. *셀 수 있는 집합의 임의의 부분집합은 항상 셀 수 있는 집합이다.*

Theorem 3.3. *셀 수 있는 두 집합의 합집합은 셀 수 있는 집합이다.*

Theorem 3.4. *셀 수 있는 두 집합의 곱집합은 셀 수 있는 집합이다.*

Theorem 3.5. *셀 수 있는 가산 개의 집합의 합집합은 셀 수 있는 집합이다.*

이러한 내용들로부터 칸토어는 양의 유리수 $\mathbb{Q}_+$가 셀 수 있는 집합임을 증명하였고, 양의 유리수, 0, 음의 유리수 모두 셀 수 있는 집합이므로 이들의 합집합인 유리수 전체집합 $\mathbb{Q}$가 가산 집합임을 증명하였다. **Uncountable set** : 집합 A가 비가산 집합이라는 것은, $|A| > |\mathbb{N}|$이다. 대표적으로 실수 집합이 있다. 더 나아가서 임의의 두 실수 $a < b$에 대해, 다음의 집합들도 모두 비가산 무한 집합이다.

$$\begin{aligned}
(a, b) &:= \{x : a < x < b\} \subset \mathbb{R} \\
[a, b] &:= \{x : a \leq x \leq b\} \subset \mathbb{R} \\
(a, b] &:= \{x : a < x \leq b\} \subset \mathbb{R}
\end{aligned}$$

칸토어는 $(0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$ 집합이 셀 수 없는 집합임을 대각선 논법 (Cantor's diagonal argument)으로 증명하였다. 즉, 실수는 무한집합 유리수와 그 크기가 다른 무한집합인 것이다. 이는 $|\mathbb{R}| = 2^{|\mathbb{N}|}$로 표현되며, 연속체 가설 (Continuum hypothesis)은 $|\mathbb{N}|$과 $|\mathbb{R}|$ 사이의 크기를 가진 집합이 존재하지 않는다는 주장인데, 이는 ZFC 공리계에서 독립적인 명제임이 알려져 있다.

4 Vector space

수학에서 가장 핵심이 되는 두 대상은 집합과 함수이다. 선형대수학에서 가장 핵심이 되는 집합은 '벡터 공간' (Vector space)이고, 가장 핵심이 되는 함수는 '선형 사상' (linear map, linear transformation) 이다. 우리는 행렬 (Matrix)이라는 선형 사상을 선형대수학에서 매우 중요하게 다룬다.

체**, Field** :
덧셈 (+), 곱셈 (×)연산과 그와 관련된 법칙들이 정의된 집합을 '체'라고 한다.

1. 집합의 임의의 두 원소를 더하거나 곱해도, 그 집합의 원소이다. (닫힘 조건)
2. 덧셈과 곱셈에 대하여 교환 법칙이 성립한다.
3. 덧셈과 곱셈에 대하여 결합 법칙이 성립한다.
4. 덧셈과 곱셈에 대하여 항등원, 역원이 집합의 원소로 존재한다. (단, 곱셈의 역원은 0이 아닌 원소에 대해)
5. 분배 법칙이 성립한다.

위 정의에 입각하여, 자연수 ($\mathbb{N}$), 정수 ($\mathbb{Z}$)는 '체'가 아니고, 유리수 ($\mathbb{Q}$), 실수 ($\mathbb{R}$), 복소수 ($\mathbb{C}$) 모두 '체'임을 쉽게 확인해볼 수 있다. 벡터 공간에서는 '체'의 원소를 '스칼라' 라고 한다. 참고로, 유한체 (Finite field)도 존재하는데, 예를 들어 $\mathbb{F}_p = \{0, 1, \cdots, p-1\}$ (p는 소수)는 modular 연산 하에서 체가 된다. 유한체 위의 선형대수학은 코딩 이론, 암호학 등에서 중요하게 응용된다.

벡터 공간**, Vector space over a field** :
벡터 공간 (V) 은 '덧셈', '스칼라 곱' 두 연산에 대하여 다음의 법칙을 만족하는 집합이다. 임의의 두 원소 $x, y \in V$와 체 F 의 원소 $a \in F$에 대하여,

1. $x + y \in V$ (덧셈에 대한 닫힘)

2. $ax \in V$ (스칼라 곱에 대한 닫힘)

3. $x + 0 = x$를 만족하는 $0 \in V$이 존재한다. (영벡터의 존재)

4. $x + z = 0$인 $z \in V$가 존재한다. (역원의 존재)

5. $1 \cdot x = x$ (스칼라 곱의 항등원)

6. $a(bx) = (ab)x$ (스칼라 곱의 결합법칙)

7. $(a + b)x = ax + bx,\ a(x + y) = ax + ay$ (분배법칙)

또한, 교환법칙, 결합법칙 등이 모두 성립한다. 선형대수학에서 벡터공간의 '체'는 실수 집합 ($\mathbb{R}$)과 복소수 집합 ($\mathbb{C}$)인 경우를 다루는데, 우리는 편의상 실수 집합에 한정해서 생각하도록 하겠다. (복소 벡터 공간은 후반부에서 다룬다.)
위 성질에서 1번과 2번 성질을 '선형성' (linearity) 이라고 표현한다. 즉, 어떤 집합 X의 임의의 두 원소 x, y와 실수 a, b에 대하여,

$$ax + by \in X$$

를 만족하면 집합 X는 선형성을 갖는 집합이다. 즉, 벡터 공간은 선형성을 갖는 집합이다.
본 강의록에서 벡터 공간은 $V/\mathbb{R}$로 표기한다.

Example 4.1. *벡터 공간의 예시들을 살펴보자.*

1. $\mathbb{R}^n$*: 가장 대표적인 유한 차원 벡터 공간이다.*

2. $M_{m\times n}(\mathbb{R})$*: 모든* $m \times n$ *실수 행렬의 집합. 차원은* mn*이다.*

3. $P_n(\mathbb{R})$*: 차수가* n *이하인 모든 실수 계수 다항식의 집합. 차원은* $n + 1$*이다.*

4. $C[a, b]$*: 폐구간* $[a, b]$*에서 연속인 실수값 함수의 집합. 이는 무한 차원 벡터 공간이다.*

5. $\{0\}$*: 영벡터만으로 이루어진 집합. 차원이* 0*인 벡터 공간이다.*

벡터**, Vector** :
벡터공간 $V/\mathbb{R}$의 원소를 '벡터'라고 한다.
선형대수학은 '유한 차원 벡터공간'과 벡터 공간을 정의역, 공역으로 갖는 함수, 구체적으로 선형 변환 (= 행렬)을 배우는 과목이다. 유한 차원 벡터 공간 중 가장 대표적인 예시가 $\mathbb{R}^n$, 즉 유클리드 공간이다.
$a_1, a_2, \cdots, a_n \in \mathbb{R}$일 때, 이들을 n-tuple 형태로 서술한 것, 즉

$$(a_1, a_2, \cdots, a_n)$$

이 $\mathbb{R}^n$의 원소를 표기하는 방식이다. 이 때 각각의 실수 $a_1, \cdots, a_n$ 들을 벡터 $(a_1, \cdots, a_n)$의 '성분' (entry, component)이라고 한다.
하나의 집합이 주어졌을 때, 우리는 그 집합의 '부분집합' (subset)을 생각할 수 있다. 선형대수학에서는 벡터 공간이라는 집합이 주어졌을 때, 이 집합과 구조가 동일한 부분 집합인 부분공간 (부분 벡터공간)을 정의한다.
$V/\mathbb{R}$의 부분 공간**, Vector subspace**:
벡터 공간 $V/\mathbb{R}$의 부분집합 W가 있을 때, W의 임의의 두 원소 $x, y \in W$가 벡터 공간의 정의를 만족시킬 때, V의 부분공간이라고 한다. 즉, $a, b \in \mathbb{R}$에 대하여, $ax + by \in W$이고, $0 \in W$이면, W는 V의 부분공간이다.

Remark 4.1. *부분공간 판정 시 가장 자주 하는 실수는 영벡터 포함 여부를 확인하지 않는 것이다.* $0 \notin W$*이면* W*는 부분공간이 될 수 없다. 예를 들어,* $\{(x, y) \in \mathbb{R}^2 : x + y = 1\}$*은* $\mathbb{R}^2$*의 부분공간이 아니다 (원점을 포함하지 않으므로).*

앞서 우리는 선형대수학에서 다루는 집합이 '유한 차원' 벡터 공간이라고 하였다. 하지만, 아직 차원이 무엇인지 정의하지 않았다. 이를 정의하기 위해 우선 '선형 결합', '선형 독립'이라는 중요한 두 개념을 먼저 짚고 넘어가겠다.
선형 결합**, linear combination** : 벡터 공간 $V/\mathbb{R}$이 있고, 임의의 원소 $x_1, x_2, \cdots, x_n \in V$이 있다. 또한, 임의의 실수들 $a_1, a_2, \cdots, a_n \in \mathbb{R}$ 이 있다. 이때, 다음과 같이 정의되는 V의 부분집합 W를 $x_1, \cdots, x_n$들의 '선형 결합'이라고 한다.

$$W := \{a_1x_1 + a_2x_2 + \cdots + a_nx_n\} \subseteq V$$

$W \subseteq V$가 부분 공간이 됨은 쉽게 알 수 있다. 또한, W는 집합 $\{x_1, \cdots, x_n\}$으로 생성 (span)되는 집합이라고 한다.
이를 벡터 공간의 정의와 연결지어 생각해보면, 벡터 공간은 선형성을 갖는 집합이고, 몇 개의 벡터를 임의로 뽑았을 때, 그 벡터들의 선형 결합은 반드시 벡터 공간의 부분집합 (부분 공간)이어야 한다. 만약, 부분 공간이 아닌 벡터 공간 전체가 몇 개의 벡터들의 선형결합과 같은 집합이면, 즉 위에서 $W = V$라면, 우리는 벡터 공간의 모든 원소들을 그 몇 개의 벡터들의 선형결합으로 표현할 수 있다는 것이다. 즉, 벡터 공간의 모든 원소들은 (유한한 크기의) 어떤 벡터

집합의 선형 결합으로 표현된다는 것을 추측해볼 수 있다. 그리고, 이 유한 개의 벡터들이 어떤 조건을 만족한다면, 우리는 이 벡터 집합을 '기저 벡터' (basis)라고 부른다. 기저 집합은 반드시 선형 독립 조건을 만족해야만 한다.

선형 독립**, linearly independent** : 벡터 공간 $V/\mathbb{R}$의 n개의 원소 $x_1, \cdots, x_n \in V$의 선형 결합으로 영벡터를 표현할 때,

$$a_1x_1 + \cdots a_nx_n = 0$$

위 식을 만족하는 계수 $a_1 = a_2 = \cdots = 0$인 경우뿐이라면, 즉, 위 방정식의 유일한 해일 때, 우리는 벡터 $x_1, \cdots, x_n$이 선형 독립이라고 한다. 선형 독립이 아니면, 선형 종속 (linearly dependent) 이라고 한다. 즉, 영벡터의 표현에서 '적어도 하나는' 0이 아닌 계수 $a_i, i \in [n]$가 존재한다.

Theorem 4.1. *벡터공간 V와 그 부분집합 $S_1 \subseteq S_2 \subseteq V$가 있다. 이때, S_1이 선형종속이면, S_2도 선형종속이다. 만약, S_2가 선형독립이면, S_1도 선형독립이다.*

Theorem 4.2. *선형 종속인 벡터 집합에서는 적어도 하나의 벡터는 다른 벡터들의 선형 결합으로 표현된다. 즉, 바꿔 말하면 다른 벡터들의 선형 결합으로 표현되는 벡터가 '존재한다'.*

Theorem 4.3 (교체 정리, Replacement theorem)**.** *벡터공간 V의 부분집합 G가 V를 생성하고, $|G| = n$이라고 하자. V의 선형독립인 부분집합 L에 대해 $|L| \leq n$이고, G의 원소 중 $n - |L|$개를 적절히 선택하여 L과 합치면 V를 생성하는 집합을 만들 수 있다.*

Corollary 4.4. *유한 차원 벡터공간에서 선형독립인 벡터의 개수는 생성 집합의 크기를 넘을 수 없다.*

Theorem 4.5. *벡터공간 V와 선형 독립인 부분집합 S가 주어져 있다. S에 포함되지 않는 벡터 $v \in V$에 대하여, $S \cup \{v\}$가 선형 종속일 필요충분조건은 v가 S의 선형 결합으로 표현되는 것이다. 즉, $v \in \mathrm{span}(S)$이다.*

기저와 차원**, Basis and Dimension** : 벡터 공간 V의 부분 집합 $\beta = \{x_1, \cdots, x_n\}$가 다음의 조건을 만족할 때, 이 벡터 집합을 V의 기저 벡터 (basis vector) 라고 한다.

1. $\mathrm{span}(\beta) = V$
2. $\{x_1, \cdots, x_n\}$이 선형 독립이다.

앞서 서술한 **Theorem**들에 의해, 위 두 조건을 동시에 만족하는 집합의 크기, cardinality는 유일함을 알 수 있다.
주의할 것은, 하나의 벡터 공간은 무수히 많은 기저 집합을 가질 수 있다. 하지만, 그 모든 기저집합의 '불변량'이 기저 집합의 크기, 즉 원소의 개수라는 의미이다. 이를 수학에서는 '차원'이라 정의한다.
즉, 집합 β의 원소의 개수, $|\beta|$를 '차원' (Dimension) 이라고 한다. 예를 들어, $\mathbb{R}^n$은 차원이 n이다. 즉, 선형 독립인 n개의 벡터로 $\mathbb{R}^n$의 모든 벡터를 표현할 수 있다. 유클리드 공간의 가장 대표적인 기저는 '표준 기저' (standard basis)로서 다음과 같다.

$$\begin{aligned} e_1 &= (1, 0, \cdots, 0) \\ e_2 &= (0, 1, 0, \cdots, 0) \\ &\cdots \\ e_n &= (0, 0, \cdots, 1) \end{aligned}$$

위 n개의 벡터들의 선형결합으로 우리는 $\mathbb{R}^n$의 어떤 벡터이든 표현할 수 있다.

Theorem 4.6. *모든 벡터공간은 기저를 갖는다.*

앞서 선형대수학은 '유한 차원 벡터 공간'을 배운다고 하였는데 이는 $\dim(V) < \infty$인 벡터 공간을 의미한다. 무한 차원 벡터 공간에 대해서는 '함수해석학'이라는 과목에서 다룬다.

앞서 우리는 벡터 공간의 부분집합이 벡터 공간의 정의를 만족할 때, 이를 부분 공간 (subspace)이라고 한다 하였다. 부분 공간도 벡터 공간이므로 기저를 가지고 있다. 그렇다면, 부분 공간의 기저와 원래 벡터 공간의 기저는 어떤 관계가 있을까?

Theorem 4.7. *n차원의 벡터 공간 V와 부분 공간 $W \subset V$가 주어져 있다. W의 기저 집합을 확장하여, V의 기저 집합을 만들 수 있다. 즉, $k(< n)$개의 선형 독립인 벡터 집합이 W를 생성한다면, $n - k$개의 선형 독립인 벡터를 추가하면, n개의 선형 독립인 V의 기저 집합이 된다.*

Theorem 4.8. *벡터 공간 $V/\mathbb{R}$의 기저 집합을 $U = \{x_1, \cdots, x_n\}$라고 하자. 그러면, 임의의 벡터 $x \in V$를 기저 벡터들의 선형 결합으로 표현할 수 있고, 더 나아가서 '유일하게' 표현된다. 즉, 다음을 만족하는 $a_1, \cdots, a_n \in \mathbb{R}$이 '유일하게' 존재한다.*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = x$$

이때, $(a_1, \cdots, a_n)$을 x의 '좌표' (coordinate) 라고 한다.

4.1 부분 공간의 합과 교

Theorem 4.9. *벡터 공간 V의 두 부분공간 W_1, W_2에 대해, $W_1 \cap W_2$도 V의 부분공간이다.*

Theorem 4.10. *두 부분공간 W_1, W_2에 대해, $W_1 + W_2 := \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}$도 부분공간이다. 이를 W_1과 W_2의 합이라 한다.*

Theorem 4.11 (차원 공식)**.** *유한 차원 벡터 공간 V의 부분공간 W_1, W_2에 대해 다음이 성립한다:*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

사실상 선형대수학 전체를 관통하는 가장 중요한 흐름 중 하나는 '기저 찾기'이다. 왜냐하면 선형대수학에서는 기저집합을 찾으면 그 벡터공간을 '알고 있다.'고 보기 때문이다. 그리고 앞서 이야기한대로 선형대수학에서 두 벡터 공간의 차원이 같으면, 두 벡터 공간은 '벡터공간으로서 같다' 라고 표현한다. 이를 '동형' (isomorphic)이라고 하는데 집합론의 '동치 관계' (equivalence relation)에서 유래했다. 이 말은 수학에 존재하는 모든 형태의 (유한 차원)벡터 공간들을 모두 모아놨을 때, 이들을 우리가 '차원'을 기준으로 분할 (partition)할 수 있다는 의미이다. 그리고 같은 차원의 벡터공간들은 벡터공간으로서 같기 때문에, '차원'을 기준으로 분할하는 것은 충분하다.

5 Matrix

이 장에서는 행렬에 대해 아주 기본적인 내용들만 살펴볼 것이다. 행렬에 대한 깊은 이야기는 앞으로 지속적으로 선형 변환과 함께 다뤄질 예정이다.

Def. 행렬 : 체 F (특히 $F = \mathbb{R}$)의 원소들을 직사각형 모양으로 배열한 것을 행렬이라고 한다. 일반적으로 $m \times n$ 행렬은

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

와 같이 쓴다. 행렬의 가로줄을 행 (row), 세로줄을 열 (column)이라 하고, a_{ij}는 i행 j열 성분이다.

행렬의 덧셈과 뺄셈, 실수배는 모두 '성분별 연산' (element-wise operation) 이다. 또한, 덧셈과 뺄셈은 두 행렬이 서로 행의 개수가 같고, 열의 개수가 같을 때 정의될 수 있다. 사실, 이는 행렬도 벡터 공간의 원소로 볼 수 있기 때문이다. 행의 개수가 서로 같고 열의 개수가 서로 같은 행렬들의 집합은 벡터 공간이다. 이를 $M_{m\times n}(\mathbb{R})$이라고 표기하자. 즉, $A \in M_{m\times n}(\mathbb{R})$은 성분이 모두 실수인 임의의 m행 n열 크기의 행렬 집합 (= 벡터 공간)이 있고 그 원소인 행렬 A를 의미한다. 즉, 넓게 보면 행렬도 벡터이다. 행렬 중 몇 가지 중요한 행렬들이 있다.

- 전치 행렬(Transpose) : A의 행과 열을 서로 뒤바꾼 행렬 (A^t, A^T).
- 대칭 행렬(Symmetric matrix) : A와 전치 행렬 A^t가 같을 때, 이를 대칭행렬이라 한다.
- 정사각 행렬 : 행과 열의 개수가 같은 행렬.
- 상삼각, 하삼각 행렬 : 대각성분을 기준으로 아래의 성분이 모두 0인 행렬을 상삼각 행렬, 위의 성분이 모두 0인 행렬을 하삼각 행렬이라고 한다.
- 대각 행렬(Diagonal matrix) : 상삼각 행렬이면서 하삼각 행렬인 행렬.
- 블록 대각 행렬 (Block Diagonal matrix) : 주대각선을 따라 0이 아닌 항들을 가질 수 있는 정방형 블록들을 포함하고, 나머지 항의 값은 모두 0인 행렬.
- 단위 행렬 (Identity matrix, $\mathbf{I}$) : 대각 성분이 모두 1인 대각 행렬.
- 직교 행렬(Orthogonal matrix) : A의 열벡터들 (또는 행벡터들)이 서로 *정규직교*(orthonormal)하는 행렬. 즉, $A^TA = AA^T = \mathbf{I}$.

특히, 대칭 행렬과 직교 행렬은 매우 중요한 행렬이고 후반부에서 이 행렬들의 중요한 성질을 자세하게 다룰 것이다. 참고로, 대각 행렬은 정사각 행렬에만 적용되지는 않는다. 더 일반화시켜 비정방 행렬 (=정사각 행렬이 아닌 행렬)에 대해서도 주대각선이 아닌 항들이 전부 0인 행렬을 (비정방) 대각행렬이라고 부른다.
행렬의 덧셈과 뺄셈은 쉽게 정의되지만 행렬의 곱셈은 조금 특수하게 정의된다. 이는 행렬과 벡터의 곱셈도 마찬가지이다. 행렬의 곱셈은 뒤에서 정의하겠다.

6 Linear transformation

앞에서 우리는 벡터공간에 대해 꼭 알아야 할 내용들을 살펴보았다. 이제 두 벡터 공간을 사상하는 함수에 대해 생각할 것이다. 그런데, 이 함수는 단순히 사상시킬 뿐만 아니라, 벡터공간의 구조를 보존해주는 함수이다. 이를 우리는 선형 사상 (또는 선형 변환)이라고 부른다.

Def. Linear transformation : 정의역과 공역이 모두 벡터공간 V, W인 함수 $T : V \to W$를 생각하자. 모든 $x, y \in V$, $a \in \mathbb{R}$에 대해, 다음이 성립하면 T를 선형 변환이라고 한다.

$$T(x+y) = T(x) + T(y),\ T(ax) = aT(x)$$

위 성질을 하나의 식으로 표현하면, 다음과 같다:

$$T(ax+y) = aT(x) + T(y).$$

참고로, 정의역과 공역이 같은 벡터 공간이면, 즉 $V = W$이면 선형변환 $T : V \to W$를 선형 연산자 (linear operator)라고 한다. 이 함수는 추후 고윳값, 고유벡터, 내적을 다룰 때 다시 등장할 것이다. 꼭 2개의 벡터에 대해서만 적용될 필요 없이 임의의 유한한 개수의 벡터 $x_1, x_2, \cdots, x_n$과 임의의 유한한 개수의 실수 $a_1, a_2, \cdots, a_n$에 대해서 다음이 성립한다:

$$T\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i T(x_i)$$

선형 변환 $T : V \to W$과 관련하여 매우 중요한 두 벡터 공간인 상공간 (Range, Image)과 영공간 (Null space, Kernel)에 대해 살펴보자.

Def. Range and Null space: T의 함숫값을 원소로 가지는 W의 부분 공간을 Range라 하고, $R(T)$로 표기하겠다. 즉, 다음의 집합이다:

$$R(T) := \{T(x) : x \in V\}.$$

$T(x) = 0$이 되게 하는 $x \in V$를 원소로 가지는 V의 부분공간을 Null space라 하고, $N(T)$로 표기하겠다. 즉, 다음의 집합이다:

$$N(T) := \{x \in V : T(x) = 0\}.$$

다음은 Range, $R(T)$의 기저 집합을 찾는 방법을 알려주는 정리이다.

Theorem 6.1. *벡터 공간 V, W와 선형 변환 $T : V \to W$, V의 기저집합 $\beta = \{v_1, v_2, \cdots, v_n\}$에 대하여 다음이 성립한다.*

$$R(T) = span(T(\beta)) = span(T(v_1), T(v_2), \cdots, T(v_n))$$

Proof. **¡1¿ Goal.** $R(T) = \text{span}(T(\beta))$임을 보인다.

¡2¿ Show. $\text{span}(T(\beta)) \subseteq R(T)$. **Reason.** 각 $T(v_i)$는 정의에 의해 $R(T)$의 원소이고, $R(T)$는 부분공간이므로 (선형결합에 닫힘) $T(v_i)$들의 모든 선형결합을 포함한다.

¡3¿ Show. $R(T) \subseteq \text{span}(T(\beta))$. **Reason.** 임의의 $y \in R(T)$를 택하면 어떤 $v \in V$가 존재하여 $y = T(v)$이다. β가 기저이므로 $v = \sum_{i=1}^{n} a_i v_i$로 쓸 수 있고, 선형성으로부터

$$y = T(v) = T\left(\sum_{i=1}^{n} a_i v_i\right) = \sum_{i=1}^{n} a_i T(v_i) \in \text{span}(T(\beta))$$

이다.

¡4¿ Conclusion. 두 포함관계가 모두 성립하므로 $R(T) = \text{span}(T(\beta))$이다.

□

다음은 영공간의 차원, 상공간의 차원이다.

Def. Rank and Nullity $N(T)$의 차원을 nullity (영공간의 차원), $R(T)$의 차원을 rank (상공간의 차원)라 한다.

Theorem 6.2. *벡터공간 V, W와 선형 변환 $T : V \to W$에 대하여, V가 유한 차원이면 다음이 성립한다:*

$$nullity + rank = dim(V)$$

7 합성 변환과 역변환

앞에서 함수에 대한 이야기를 했을 때, 합성 함수와 역함수에 대한 정의를 이야기했다. 선형 변환도 함수이기 때문에 합성 함수와 역함수를 이야기해야만 한다. 그리고 이 두 함수 모두 '선형 변환'임을 살펴볼 것이다.

Def. 합성 변환 : 벡터 공간 V, W, Z와 두 선형 변환 $T : V \to W$, $U : W \to Z$를 생각하자. 두 선형 변환의 합성함수 $UT : V \to Z$는 선형변환이고, 합성 변환이라고 한다.

어떤 함수가 있을 때, 이 함수의 역함수가 존재하는가는 수학에서 매우 중요한 주제이다. 또한, 이때 원래의 함수와 그 역함수는 많은 성질들을 공유한다. 예를 들어, 원래 함수가 연속이면 그 역함수도 연속이고, 원래 함수가 미분 가능하면 적절한 조건 하에서 그 역함수도 미분 가능하다. (= 역함수 정리) 이제 선형 변환이란 함수가 언제 역함수가 존재하는지, 그 역함수도 선형변환인지에 대해 살펴보자.

벡터 공간 V, W와 선형 변환 $T : V \to W$를 생각하자. 이때, 선형 변환 $UT = \mathbf{I}_V$, $TU = \mathbf{I}_W$인 함수 U를 T의 역함수라고 한다. 이때, $\mathbf{I}_U$, $\mathbf{I}_W$는 각각 항등 변환이다. 역함수가 존재하는 T를 가역이라 하고, 역함수를 T^{-1}로 표기한다. T가 가역이면, 그 역함수는 유일함을 기억하자.

우선, 함수 시간에 다루었듯이 역함수가 존재하는 것의 필요충분조건은 함수가 '전단사 함수' (일대일 대응 함수)인 것이다.

Theorem 7.1. *가역인 선형변환 $T : V \to W$에 대해, 역함수 T^{-1}도 선형변환이다. 앞으로는 이를 '역변환'이라 부르겠다.*

Proof. **¡1¿ Goal.** $T^{-1} : W \to V$가 선형임을 보인다. 즉, 임의의 $y_1, y_2 \in W$와 $c \in \mathbb{R}$에 대해

$$T^{-1}(cy_1 + y_2) = cT^{-1}(y_1) + T^{-1}(y_2)$$

임을 보이면 된다.

¡2¿ T가 가역이므로, $x_1 := T^{-1}(y_1)$, $x_2 := T^{-1}(y_2)$가 유일하게 존재한다.

¡3¿ T의 선형성으로부터

$$T(cx_1 + x_2) = cT(x_1) + T(x_2) = cy_1 + y_2$$

이다.

¡4¿ 양변에 T^{-1}를 적용하면

$$T^{-1}(cy_1 + y_2) = cx_1 + x_2 = cT^{-1}(y_1) + T^{-1}(y_2)$$

를 얻는다.

□

더 나아가, 다음의 중요한 정리를 제공해준다.

Theorem 7.2. *선형변환 $T : V \to W$가 가역이면, $dim(V) = dim(W)$이다.*

이를 증명하기 위해선 몇 가지의 정리들이 필요하다. 이에 대해서 살펴보자.

Theorem 7.3. *벡터공간 V, W와 선형 변환 $T : V \to W$에 대하여, 다음은 서로 필요충분관계이다:*
T는 단사함수 (=일대일 함수)이다. $\leftrightarrow$ $N(T) = \{0\}$이다.

즉, 영공간의 유일한 원소가 영벡터뿐이므로, (영벡터는 벡터공간의 정의상 항상 원소로 포함되어져야 한다) 영공간의 차원도 0이다.

Theorem 7.4. *벡터 공간 V, W의 차원이 같을 때, 선형 변환 $T : V \to W$에 대하여, 다음 세 명제는 서로 필요충분관계이다.*

1. *T는 단사이다.*
2. *T는 전사이다.*
3. *$rank(T) = dim(V)$*

Theorem 7.3 $\sim$ 7.4의 증명은 생략한다. 이 두 정리를 이용해서, Theorem 7.2.를 증명하자.

Proof. **¡1¿ Assume.** $T : V \to W$가 가역이라고 하자. 그러면 T는 전단사(특히 단사, 전사)이다.

¡2¿ T가 단사이므로 $\mathrm{Ker}(T) = \{0\}$이고, 따라서 $\mathrm{nullity}(T) = 0$이다.

¡3¿ (랭크-널리티 정리) $\dim(V) = \mathrm{rank}(T) + \mathrm{nullity}(T)$이므로 $\dim(V) = \mathrm{rank}(T)$이다.

¡4¿ T가 전사이므로 $\mathrm{Im}(T) = W$이고, 따라서 $\mathrm{rank}(T) = \dim(W)$이다.

¡5¿ **Conclusion.** $\dim(V) = \dim(W)$이다.

$\square$

합성 명제에서 '만약 P이면, Q이다.'가 참인 경우, Q는 P의 '필요조건'이었음을 기억하자. 즉, 두 벡터공간의 차원이 같은 것은 선형변환의 가역성을 위한 필요조건이다. 또한, Theorem 7.4.은 우리에게 두 벡터 공간의 차원이 같을 때 선형변환이 가역이기 위한 조건을 알려주는 정리이다. T가 단사이면, 자동으로 전사이고, T가 전사이면, 자동으로 단사이므로 전단사 함수가 되어 가역이다. rank(T) = dim(V)이면, T는 전단사이므로, 자동으로 가역이다. 그러한 점에서 Theorem 7.4.은 의미가 있다. 특히 3번째 조건은 '역행렬' 이야기에서도 다시 등장할 것이므로, 잘 기억해두자.

이제 매우 중요한 이야기인 선형변환으로서 행렬에 대한 이야기를 할 것이다. 사실, 모든 선형변환은 행렬로 표현된다. 그리고 이로부터 우리는 행렬의 곱, 역행렬, 벡터 공간의 동형성 등등을 정의할것이다.

8 선형 변환의 행렬 표현

먼저, 순서 기저부터 정의하자. 순서 기저 (ordered basis)란, 순서가 주어져 있는 기저로서 $\mathbb{R}^3$를 예로 들면, 두 기저 집합 $\beta = \{e_1, e_2, e_3\}$, $\gamma = \{e_1, e_3, e_2\}$가 있다. 기저로서 둘은 같은 집합이지만, 순서 기저로는 순서가 다르기 때문에 다른 기저이다. 이제부터 우리는 주로 '순서 기저'로 이야기할 것이다. 또한, 앞에서 배운 정리에 의해 (Theorem 4.6.), 우리는 좌표벡터를 다음과 같이 표기할 수 있다.

Coordinate vector. 유한 차원 벡터 공간 V의 순서기저를 $\beta = \{v_1, v_2, \cdots, v_n\}$라 하고, $x \in V$에 대하여, 다음을 만족하는 $a_1, \cdots, a_n$이 유일하게 존재한다:

$$x = \sum_{i=1}^{n} a_i v_i.$$

이때 β에 대한 x의 '좌표 벡터' $(a_1, a_2, \cdots, a_n)$는 $[x]_\beta$로 표기한다.

Matrix representation. 유한 차원 벡터 공간 V, W와 각각의 순서기저 β, γ가 주어져 있다. 그리고 선형 변환 $T : V \to W$가 주어져 있다.

$$\begin{aligned} \beta &= \{v_1, \cdots, v_n\} \\ \gamma &= \{w_1, \cdots, w_m\} \end{aligned}$$

그러면, $T(v_1), \cdots, T(v_n)$은 모두 W의 원소이다. 즉, $w_1, \cdots, w_m$의 선형결합으로 표현될 것이고, 이때의 계수는 '유일하게' 존재한다. (좌표) 따라서,

$$\begin{aligned} T(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ &\vdots \\ T(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \end{aligned}$$

위 계수들을 모두 모으면 우리는 $m \times n$ 크기의 행렬을 얻을 수 있고, 이를 T의 '행렬 표현'이라 하고, $[T]_\beta^\gamma$로 표기한다.

즉, 다음과 같은 행렬을 얻을 수 있다. $\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$

두 벡터 공간의 '순서기저'가 주어져 있다면, 하나의 선형 변환의 행렬 표현은 '유일하다'는 것을 생각할 수 있다. 이를 바꿔 표현하면, 선형 변환과 행렬 사이에 '일대일 대응'관계가 성립한다고 할 수 있다. 또한, 영변환의 행렬표현은 영행렬이고, 항등변환의 행렬 표현은 항등 행렬이다. (스스로 생각해볼 것.) 앞에서 우리는 행렬을 '벡터공간의 원소'로 바라보았다. 그런데, 벡터 공간의 순서기저가 주어져 있다면 우리는 모든 선형 변환은 행렬로 표현하는 것이 가능하다. 따라서, 선형변환의 집합도 결국 벡터 공간이 아닐까 생각해볼 수 있다. 실제로, 두 선형 변환의 합과 실수배도 결국 선형변환이다. 즉, 두 벡터 공간 V, W가 주어져 있고 선형변환 $T, U : V \to W$일 때,

$$(aT + U)(v) = aT(v) + U(v)$$

와 같은 방식으로 선형변환의 합과 스칼라 배를 정의하면, U에서 V로의 모든 선형 사상은 벡터 공간이 됨을 알 수 있다. 이는 곧 쌍대 공간의 기본이 된다. (Dual space) 쌍대 공간은 선형대수학에서 심화적인 내용이기 때문에 본 강의에서는 자세하게 다루지는 않는다. 다만, 후반부에서 간단하게 살펴볼 예정이다.

9 행렬의 곱셈과 역행렬

유한 차원 벡터 공간 V, W, Z와 두 선형 변환 $T : V \to W$, $U : W \to Z$가 주어져 있고, 각각의 벡터 공간의 순서 기저를 α, β, γ라 하자.

$$\alpha = \{v_1, \cdots, v_n\}$$
$$\beta = \{w_1, \cdots, w_m\}$$
$$\gamma = \{z_1, z_2, \cdots, z_p\}$$

두 행렬 $A = [U]_\beta^\gamma \in M_{p\times m}(\mathbb{R})$, $B = [T]_\alpha^\beta \in M_{m\times n}(\mathbb{R})$가 유일하게 존재한다. 우리는 두 선형 변환의 합성 함수인 $(U \circ T)(v) : V \to Z$가 선형 변환임을 알고 있다. 즉, $[UT]_\alpha^\gamma \in M_{p\times n}(\mathbb{R})$ 행렬이 유일하게 존재함을 알 수 있다. 행렬의 곱셈은 다음이 성립하도록 정의된다:

$$AB = [UT]_\alpha^\gamma$$

행렬의 곱셈을 어떻게 계산하는지는 이미 알고 있을 것이라 생각된다. 여기서는 '왜' 행렬의 곱셈을 그렇게 정의하는지 이야기했다. 행렬의 곱셈은 합성 변환이고, 합성 변환의 행렬 표현을 생각했을 때, 행렬의 곱셈은 사실상 그렇게 정의할 수밖에 없음을 이야기하는 것이다. 그리고, 행렬의 표현은 교환 법칙이 성립하지 않음도 자연스럽다. 두 함수를 합성할 때, 함수의 합성 순서를 바꾸면 당연히 서로 다른 함수가 나오거나, 합성 함수가 정의되지 않을 수도 있다. 행렬의 곱셈도 마찬가지로, 곱셈 순서를 바꾸면 서로 다른 행렬이 나오거나, 곱셈이 정의되지 않는 것이다.

Theorem 9.1. *항등 행렬은 행렬 곱셈에 대한 항등원이다. 즉,* $A\mathbf{I} = \mathbf{I}A = A$*이다. 또한, 행렬의 곱셈은 분배 법칙은 성립한다. 즉,* $A(B + C) = AB + AC$*이다.*

Theorem 9.2. *행렬의 곱셈* AB*와* B^TA^T*가 정의된다고 하자. 그러면,* $(AB)^T = B^TA^T$*가 성립한다.*

행렬과 벡터의 곱셈도 이와 유사하게 정의된다. 순서 기저가 각각 α, β인 두 벡터 공간 V, W가 있고 선형 변환 $T : V \to W$가 있다. 또한, $|\alpha| = n, |\beta| = m$이라 하자. 그러면, 임의의 $v \in V$에 대해, 다음이 성립한다:

$$[T(v)]_\beta = [T]_\alpha^\beta [v]_\alpha$$

이때, $[T(v)]_\beta \in M_{m\times 1}(\mathbb{R})$이고, $[T]_\alpha^\beta \in M_{m\times n}(\mathbb{R}), [v]_\alpha \in M_{n\times 1}(\mathbb{R})$이다. 이는 결국 다음의 사실을 알려준다. 행렬과 벡터의 곱셈은 곧 두 벡터 공간 사이의 한 선형변환과 일대일 대응된다. 선형 변환과 그것의 행렬 표현 사이에는 (순서 기저가 주어져 있으면) 일대일 대응 관계가 성립하므로, 한 벡터 공간에서 다른 벡터 공간으로의 선형 변환은 모두 '행렬과 벡터의 곱셈'으로 표현된다. 그리고 사실 행렬과 벡터의 곱셈의 본질은 하나의 벡터 공간에서 다른 벡터 공간으로의 선형 사상이다.

이제 역행렬 (Inverse matrix)에 대한 이야기를 하자. 앞에서 우리는 역변환의 정의, 선형변환이 가역이기 위한 필요조건과 필요충분 조건을 모두 다뤘었다. 이는 모두 역행렬의 정의, 행렬이 가역이기 위한 필요조건과 필요충분 조건에도 그대로 대응된다.

선형변환 $T : V \to W$가 가역이기 위한 필요조건은 두 벡터 공간 V, W의 차원이 같은 것이다. 이때, V, W의 순서 기저를 각각 α, β라 하면, $A = [T]_\alpha^\beta$는 정사각 행렬이 된다. 즉, 행렬이 가역행렬이기 위한 필요조건은 정사각 행렬인 것이다. 선형변환에서 역변환의 정의는 합성변환이 항등변환인 것이었다. 행렬에서 역행렬의 정의는 행렬을 서로 곱했을 때 항등행렬이 나오는 것이다. 즉, 행렬 $A \in M_{n\times n}(\mathbb{R})$에 대하여, 다음의 등식을 만족하는 행렬 B를 역행렬이라고 한다:

$$AB = BA = \mathbf{I}.$$

이때 단위행렬 $\mathbf{I} \in M_{n\times n}(\mathbb{R})$은 주대각성분이 1이고 나머지는 모두 0인 행렬이다. 또한, 다음이 성립한다:

$$B = [T^{-1}]_\beta^\alpha = ([T]_\alpha^\beta)^{-1}.$$

구체적으로, 역행렬을 계산을 통해 구하는 방법에 대해서는 뒤에서 다룰 것이다. 또한, 역행렬과 관련된 유용한 정리들에 대해서도 이때 함께 다룰 것이다.

10 동형과 동형사상

Def. Isomorphic and Isomorphism. 두 벡터 공간 V, W에 대해, 가역인 선형변환 $T : V \to W$가 존재하면 V, W는 '동형' (isomorphic)이다. 이때, 가역인 선형변환을 V에서 W로의 '동형 사상' (isomorphism)이라고 한다.

Theorem 10.1. *순서기저가* β*인* n*차원 벡터공간* V*와* $\mathbb{R}^n$*은 동형이다. 이때, 동형 사상은 다음과 같다:*

$$\forall x \in V,\ \Phi_\beta(x) = [x]_\beta$$

Theorem 10.2. *n차원 벡터공간 V와 m차원 벡터 공간 W가 있다. V와 W의 순서기저를 각각 β, γ라 하자. 모든 선형 변환 집합 $L(V, W)$와 $M_{m\times n}(\mathbb{R})$은 동형이다. 이 때, 동형 사상은 다음과 같다:*

$$\forall T \in L(V, W), \ \Phi_\beta^\gamma(T) = [T]_\beta^\gamma.$$

사실 위 두 개의 정리를 엄밀하게 증명하려면, $2 \sim 3$ 시간은 더 이야기해야 한다. 그 과정도 나름의 의미가 있겠지만, 우리는 일단 위 정리를 받아들이고 넘어가는 선에서 마무리를 짓는다. 이제부터 우리는 표준 (순서)기저가 주어진 n차원의 벡터 공간 $\mathbb{R}^n$, $m \times n$ 행렬 $A : \mathbb{R}^n \to \mathbb{R}^m$에 대한 이야기로 초점을 맞출 것이다. 하지만, 이 논의들은 n차원의 일반적인 벡터 공간, $T : V \to W$에도 모두 적용되는 것이다. 왜냐하면, 이들은 서로 동형이기 때문이다.

11 선형 방정식

이제부터 선형대수학에서 매우 중요한 주제인 '선형 방정식' (linear equation)에 대해 이야기할 것이다. 사실, 공학 계열에 한정해서 선형대수학이 중요하다고 하는 이유는 '선형 방정식'과 '고윳값/고유벡터' 때문이라고 해도 과언이 아니다. 이는 중학교 (고등학교)에 배운 '연립일차방정식'을 확장한 것이다. 연립일차방정식은 크게 3가지 경우가 있는데, '해가 무수히 많은 경우', '해가 유일한 경우', '해가 없는 경우'이다. 우리는 '해가 유일한 경우'에 대해 이야기를 먼저 진행하고, 그 다음 '해가 무수히 많은 경우'에 대해 살펴볼 것이다. '해가 없는 경우'는 추후 '내적'을 배운 다음에 이야기한다. 이는 머신러닝의 선형 회귀에서도 가장 기본이 되는 개념이다.

선형 방정식은 다음과 같은 형태이다:

$$Ax = b$$

이때, 행렬 $A \in M_{m\times n}(\mathbb{R}), x \in \mathbb{R}^n, b \in \mathbb{R}^m$이다. 먼저, 가장 특수한 경우, 즉 A가 가역행렬이어서, 유일한 해 $x = A^{-1}b$를 갖는 경우에 대해 이야기할 것이다. 가역 행렬이므로, $m = n$은 당연하다.

먼저 연립일차방정식을 생각해보자.

$$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 12 \\ 3x_1 + 2x_2 - x_3 &= -4 \\ x_1 + x_2 + 2x_3 &= 10 \end{aligned}$$

위 방정식에서 바로 우리는 '해' (x_1, x_2, x_3의 값)를 구할 수는 없다. 그래서 우리는 다음과 같은 연산을 수행한다.

1. 연립방정식의 두 행의 위치를 교환
2. 연립방정식의 어느 행에 0이 아닌 실수값을 곱한다
3. 연립방정식의 어느 행과 다른 행을 더한다 (또는 뺀다).

위 3가지를 수행하는 목적은 '미지수의 개수'를 줄이기 위함이다. 미지수의 개수를 줄여서 $0 \cdot x_1 + 0 \cdot x_2 + ax_3 = b$와 같은 형태의 방정식을 하나 만들면, $x_3 = b/a$를 얻을 수 있고, 그 다음에는 이것을 대입하여, 다시 위의 3가지를 수행한다. 그러면 또 다른 해의 값을 하나 얻을 수 있고, 이러한 과정을 수행해 마지막 남은 미지수의 해까지 얻을 수 있다. 위 연립일차방정식에서 각각의 계수를 모아서 행렬로 만들면 $A \in M_{3\times 3}(\mathbb{R})$이 됨을 알 수 있다. 즉, 행렬의 행은 방정식의 개수, 열은 미지수의 개수와 대응된다. 행렬 A가 가역행렬이어서, $x = A^{-1}b$를 갖는 경우에 대해 먼저 이야기를 진행할 것이므로, 이는 곧 방정식의 개수와 미지수의 개수가 같은 연립일차방정식을 살펴보는 것이다. 그러면 우리가 연립일차방정식에서 수행하는 위 3개의 연산은 행렬의 행에 대해 수행하는 연산으로 볼 수 있다. 곧,

1. 행렬의 두 행의 위치를 교환
2. 행렬의 한 행에 0이 아닌 실수값을 곱한다
3. 행렬의 한 행과 다른 행을 더한다 (또는 뺀다)

이다. 이 3개의 연산들을 우리는 '기본행연산' (elementary row operation)이라고 부른다. 그러면 행렬 A에 기본행연산을 수행해서 우리가 얻고자 하는 행렬은 어떤 행렬일까? 이에 대한 질문의 답은 잠시 뒤에 하도록 하고, 먼저 위 기본 행연산이 '기본행렬과의 곱셈'으로 표현될 수 있음에 대해 이야기하자.

Def. 기본 행렬 기본 행렬은 항등 행렬에 기본행연산을 적용하여 얻은 행렬이다. 즉, 첫 번째 연산을 한 번 수행하여 얻은 기본행렬들, 두 번째 연산을 한 번 수행하여 얻은 기본행렬들, 세 번째 연산을 한 번 수행하여 얻은 기본행렬들이 있다.

Theorem 11.1. *임의의 행렬 A에 기본 행연산을 수행하여 얻은 행렬 B는 행렬 A와 기본 행렬 E와의 곱으로 표현된다. 즉, $B = EA$이다. 이때 E는 $\mathbf{I}$에 동일한 기본 행연산을 수행하여 얻은 행렬이다. 기본 행연산의 역연산을 수행하여 B에서 다시 원래의 행렬 A로 돌아갈 수 있다. 즉, $A = E^{-1}B$가 되게 하는 E^{-1}가 존재한다. 다시 말해서, 기본행렬은 항상 가역이다.*

이제 기본 행연산을 통해 어떻게 선형 방정식의 해를 찾는지를 보자. 먼저, 첨가 행렬을 정의한다. 이는 $[A|\mathbf{I}]$로 표기하며, 행렬 A에 단위행렬 $\mathbf{I}$를 '첨가'한 것이다. 이후, 기본 행연산을 유한 번 수행하여 $[I|B]$ 꼴로 만드는데, 이때 $B = A^{-1}$이다. 다음 행렬의 역행렬을 알려준 방식대로 직접 구해보자.

$A = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix}$ 이 방식을 통해 $Ax = b$의 A가 가역행렬일 때, $x = A^{-1}b$로 그 해를 구할 수 있다. 하지만, 우리는 언제 A가 가역행렬인지 더 이야기할 필요가 있다.

11.1 행렬의 랭크와 역행렬

행렬의 랭크에 대해 이야기하기전에, 행렬과 벡터의 곱셈 (선형 변환)에 대해 한 가지 사실을 짚고 넘어가자. 행렬 $A \in M_{m\times n}(\mathbb{R})$과 벡터 $x \in \mathbb{R}^n$의 곱셈은 다음과 같이 행렬의 열벡터 $(a_1, \cdots, a_n)$의 선형 결합과 동일하다.

$$Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

즉, 임의의 벡터 x와 행렬의 곱셈 결과 나오는 벡터 Ax는 행렬 A의 열공간 (-¿ $\mathbb{R}^m$의 부분공간)의 원소이다.

행렬 $A \in M_{m\times n}(\mathbb{R})$의 랭크는 다음과 같이 정의되는 선형 변환 $L_A : \mathbb{R}^n \to \mathbb{R}^m$의 랭크이다:

$$L_A(x) = Ax$$

다음은 매우 중요한 결과 중 하나이다.

Theorem 11.2. *정사각행렬 $A \in M_{n\times n}(\mathbb{R})$가 가역행렬인 것의 필요충분조건은 랭크가 n인 것이다.*

이는 Theorem 7.4.를 통해 확인할 수 있다. 왜 그런지 스스로 생각해보도록 하자.

Theorem 11.3. *두 벡터공간 V, W와 각각의 순서기저가 β, γ라 하자. 두 벡터공간 사이의 선형변환 $T : V \to W$의 행렬표현과 선형변환의 랭크는 같다. 즉, $rank(T) = rank([T]_\beta^\gamma)$이다.*

결국 선형변환의 랭크는 그 선형변환의 행렬표현의 랭크를 찾는 문제로 바꿔볼 수 있다. 임의의 행렬의 랭크를 보존해주는 연산으로서, 가역행렬과의 곱셈이 있다.

Theorem 11.4. *임의의 행렬 $A \in M_{m\times n}(\mathbb{R})$과 가역행렬 P, Q가 있다. 행렬의 곱셈이 잘 정의된다고 하자. 그러면, 다음이 성립한다.*

1. $rank(A) = rank(PA)$
2. $rank(A) = rank(AQ)$
3. $rank(A) = rank(PAQ)$

위 정리를 통해 우리는 임의의 행렬의 랭크를 구할 때, 가역행렬과의 곱셈을 수행할 수 있다. 앞에서 배웠던 '기본행렬'은 가역행렬이고 임의의 행렬과 기본행렬과의 곱셈은 행렬의 랭크를 보존함을 알 수 있다. 기본행렬과의 곱셈은 곧 기본행연산임을 기억하자. 따라서, 임의의 행렬의 랭크를 구할 때, 기본행연산을 수행하는 것은 행렬의 랭크를 보존해준다. 즉, 우리는 임의의 행렬의 랭크를 (유한 번의) '기본행연산'을 수행함으로서 구할 수 있다.

다음은 행렬의 랭크와 관련된 자주 사용되는 사실이다.

Theorem 11.5. *임의의 행렬의 랭크는 선형독립인 열 (열벡터)의 개수와 같다. 즉, 행렬의 랭크는 그 열에 의해 생성된 부분공간 (=열공간, Column space)의 차원과 같다.*

Proof. **¡1¿ Goal.** rank(A)가 A의 선형독립인 열벡터의 개수와 같음을 보인다.

¡2¿ 선형변환 $L_A : \mathbb{R}^n \to \mathbb{R}^m$를 $L_A(x) = Ax$로 정의하면,

$$\mathrm{rank}(A) = \mathrm{rank}(L_A) = \dim(\mathrm{Im}(L_A))$$

이다.

¡3¿ $\mathbb{R}^n$의 표준기저를 $\{e_1, \ldots, e_n\}$라 하자. 그러면 $L_A(e_i) = Ae_i$는 A의 i번째 열벡터 a_i와 같다.

¡4¿ 따라서

$$\mathrm{Im}(L_A) = \mathrm{span}\{L_A(e_1), \ldots, L_A(e_n)\} = \mathrm{span}\{a_1, \ldots, a_n\}$$

이고, 이는 곧 A의 열공간(Column space)이다.

¡5¿ Conclusion. rank(A) = dim(Im(L_A))는 열공간의 차원이므로, 이는 A의 선형독립인 열벡터의 개수와 같다. □

마지막으로, 다음은 임의의 행렬의 선형독립인 열벡터의 개수와 선형독립인 행벡터의 개수는 항상 같음을 알려준다.

Theorem 11.6. *임의의 행렬 $A \in M_{m\times n}(\mathbb{R})$의 랭크는 $A^t \in M_{n\times m}(\mathbb{R})$의 랭크와 같다.*

자, 그럼 다시 선형방정식 $Ax = b$의 이야기로 돌아가자. 우리는 여전히 이 방정식의 해가 유일한 경우, 즉 $x = A^{-1}b$를 유일한 해로 갖는 경우에 대해 이야기할 것이다. 먼저, A의 역행렬이 존재하는 필요충분조건을 우리는 Theorem 11.2.에서 배웠다. 열벡터의 개수가 n개 있는 정사각행렬의 랭크가 n이라는 것은, Theorem 11.5.에 의해 모든 열벡터가 선형독립이라는 것과 동일한 이야기이다. 즉, 행렬 A가 생성하는 열공간 (Column space)의 차원이 n이어야 함을 의미한다. Theorem 6.2.에 의해, 이는 곧 다음과 같은 영공간의 차원이 0임을, 즉 $\{0\}$만을 유일한 원소로 갖는 부분 공간이어야 함을 알 수 있다.

$$\{x : Ax = 0\}$$

다른 방식을 통해서도 생각해볼 수 있다. $Ax = 0$은 다음과 같은 행렬 A의 열벡터들의 선형 결합이고, 이들이 선형독립이라면 다음의 방정식의 해는 자명한 해 $x = 0$만 존재할 뿐이다:

$$Ax = x_1a_1 + x_2a_2 + \cdots + x_na_n = 0.$$

그러면 구체적으로 $Ax = b$의 해가 유일한 경우, x를 어떻게 구할 수 있을까? 앞에서 우린 가역행렬 A에 '유한 번' 기본행연산을 수행하여 그 역행렬 A^{-1}를 구할 수 있는 방법을 살펴보았다. 이번에는 이와 같은 방법이 아닌 다른 방법을 통해 해를 구하는 방법을 살펴볼 것이다. 이는 곧 '가우스 (요르단) 소거법'이다.

11.2 가우스 (요르단) 소거법

먼저, 선형 방정식 $Ax = b$의 첨가행렬 $[A|b]$를 정의하자. 이 행렬에 우리는 유한 번의 기본 행연산을 수행하여 '기약행사다리꼴' (reduced row echleon form) 행렬을 만들 수 있다. 이를 가우스 소거법이라 한다.

Def. 기약행사다리꼴 행렬 다음의 세 조건을 모두 만족하는 행렬을 기약행사다리꼴이라 한다.

1. 0이 아닌 성분을 가지는 행은 모든 성분이 0인 행보다 위에 위치한다.
2. 각 행의 처음으로 0이 아닌 성분은 그 성분을 포함한 열에서 유일하게 0이 아닌 성분이다.
3. 각 행의 처음으로 0이 아닌 성분은 1이고, 이전 행의 처음으로 0이 아닌 성분보다 오른쪽에 위치한다.

주어진 행렬에 대하여 그 행렬의 기약행사다리꼴은 '유일'하다. 기본 행연산을 어떻게 적용하냐, 어떤 순서로 적용해나가냐는 여러 가지 방법이 있지만 얻게 되는 최종적인 행렬 (기약행사다리꼴 행렬)은 모두 동일하다는 의미이다. 가우스 소거법은 다음과 같은 방법으로 '유한 번'의 기본행연산을 수행하여 기약행사다리꼴 행렬을 얻게 한다.

- '위의 행'에서 '아래의 행'으로 내려가면서 첨가행렬에 기본행연산을 수행해준다. 이때 '상삼각행렬'이 되도록 기본행연산을 수행해준다.
- 상삼각행렬을 얻으면 '아래의 행'에서 '위의 행'으로 기본행연산을 수행해주어, 기약행사다리꼴 행렬을 얻는다.

가우스 소거법은 주어진 행렬을 기약행사다리꼴 행렬로 변경하는 가장 효율적인 방법이다. 이후 기약행사다리꼴 행렬을 얻게 되면 선형방정식의 해는 자연스럽게 얻게 된다. 다음은 기약행사다리꼴과 관련하여 중요한 정리이다.

Theorem 11.7. *원래의 행렬 A에서 가우스 소거법을 통해 얻은 기약행사다리꼴 행렬을 R이라 하자. A와 R 사이에는 다음의 관계가 성립한다.*
1. $rank(A) = rank(R)$
2. *R에서 선형 독립인 열벡터는 A에서도 선형 독립이다. 예를 들어, R에서 1열과 3열이 선형 독립이면 A에서도 1열과 3열이 선형 독립이다. 또한, 나머지 선형 종속인 열벡터들을 선형 독립인 열벡터들의 선형 결합으로 표현할 때, 계수값들도 R과 A가 서로 동일하다.*

1번 내용은 Theorem 11.4.에 의해 자명하다. 하지만 이 정리는 한 단계 더 나아가서 기약행사다리꼴은 원래의 행렬의 어떤 열벡터들이 '선형독립'인지 알려주고 선형결합의 계수도 알려준다는 것이다. 그러한 점에서 의미가 있다. 이처럼, 기본행연산은 매우 강력하고 막강한 계산법이다. 스스로 갖고 있는 선형대수학 책을 활용하여 가우스 소거법을 연습해보길 바란다.

11.3 해가 무수히 많은 연립일차방정식

먼저 구체적인 연립일차방정식을 살펴보자.

$$x_1 + 2x_2 + x_3 = 7$$
$$x_1 - x_2 - x_3 = -4$$

이 방정식을 가우스 소거법을 통해 풀면, 해집합은 다음과 같다:

$$\{(1,1,4) + t(1,-2,3) : t \in \mathbb{R}\}$$

열벡터 표기가 원칙이지만, 편의상 행벡터 표기로 하였다. 실수 t의 개수는 무수히 많으므로 위 연립일차방정식의 해는 무수히 많다. 그러면 언제 이러할까?

먼저 $Ax = 0$에서 시작하자. $A \in M_{m\times n}(\mathbb{R})$이고 $\{x : Ax = 0\}$의 차원은 $n - rank(L_A)$이다 (Theorem 6.2.에 의해 성립한다). 만약 $rank(L_A) = n$이면, 영공간의 차원은 0이므로, 이는 영벡터만이 유일한 해라는 이야기이다. 하지만, 그 외의 경우 ($rank(L_A) < n$)에는 영공간의 차원이 0보다 크므로, 이는 영벡터가 아닌 해가 존재한다는 의미이다. 사실, $Ax = 0$은 해가 무수히 많거나, 유일하거나 딱 두 가지 상황만 존재한다. 예를 들어서 영벡터가 아닌 $x = (a_1, a_2, \cdots, a_n)$가 $Ax = 0$의 해라면 임의의 실수 c에 대해서, $cx = (ca_1, ca_2, \cdots, ca_n)$도 $Ax = 0$의 해가 되므로, 해가 무수히 많다. 예를 들어, $rank(L_A) = n-2$라고 해보자. 그러면 영공간의 차원은 2이다. 이 말은 $Ax = 0$을 만족하는 $\{x\}$중에서 선형독립인 벡터가 2개 존재한다는 의미와 동일하다. 그리고, 그 2개의 선형독립인 벡터들의 선형 결합으로 표현되는 모든 벡터가 $Ax = 0$의 해가 된다. 또 $m < n$이라고 해보자. 이때 $Ax = 0$은 항상 해가 무수히 많다. 왜냐하면 행렬의 랭크는 선형독립인 행벡터 (또는 선형독립인 열벡터)의 개수와 같은데 $m < n$이므로, $rank(A) \leq m < n$이므로, $n - rank(A) \geq 1$이다. 따라서 영공간의 차원이 1 이상이므로 무조건 해가 무수히 많을 수밖에 없다.

Theorem 11.8. *$A \in M_{m\times n}(\mathbb{R})$이고, $m < n$이면, $Ax = 0$의 해가 무수히 많다.*

이제 다음의 정리를 통해 우리는 $Ax = b$의 해집합으로 확장할 수 있다.

Theorem 11.9. *$Ax = 0$의 해가 무수히 많다고 하자. 그리고 그 해집합을 K라 하자.그러면, 동일한 행렬 A에 대해 $Ax = b$의 해집합은 다음과 같다. $Ax = b$의 임의의 한 해를 s라 하자. 즉, $As = b$이다.*

$$s + K = \{s + k : k \in K\}$$

직관적인 증명은 매우 단순하다. $A(s + k) = As + Ak = As = b$이므로, 집합 $s + K$는 $Ax = b$의 해집합이다. 엄밀한 증명은 조금 더 생각해야 되지만 이 정도 선에서 이해하고 넘어가자. 하지만 아직 한 가지 상황이 남아있다. 우리는 언제 연립일차방정식의 해가 없는지를 아직 이야기하지 않았다. 이를 위해, 연립일차방정식의 해가 존재하는 상황 (=유일한 경우 or 무수히 많은 경우)를 판단할 수 있는 한 가지 필요충분조건의 명제를 이야기하겠다.

Theorem 11.10. *$Ax = b$의 해가 존재하기 위한 필요충분조건은 $\mathrm{rank}(A) = \mathrm{rank}([A \mid b])$인 것이다.*

Proof. **¡1¿** A의 열벡터를 $a_1, \ldots, a_n$이라 하자. 그러면

$$\mathrm{col}(A) = \mathrm{span}\{a_1, \ldots, a_n\}$$

이고,

$$\mathrm{col}([A \mid b]) = \mathrm{span}\{a_1, \ldots, a_n, b\}$$

이다.

¡2¿ 따라서

$$\mathrm{rank}(A) = \dim \mathrm{col}(A) = \dim \mathrm{span}\{a_1, \ldots, a_n\}, \qquad \mathrm{rank}([A \mid b]) = \dim \mathrm{span}\{a_1, \ldots, a_n, b\}$$

이다.

¡3¿ **Claim.** $\mathrm{rank}(A) = \mathrm{rank}([A \mid b]) \iff b \in \mathrm{span}\{a_1, \ldots, a_n\}$. **Reason.** b를 추가해도 생성공간의 차원(=랭크)이 증가하지 않는다는 것은 b가 이미 기존 열벡터들의 선형결합이라는 뜻이기 때문이다.

¡4¿ 한편 $b \in \mathrm{span}\{a_1, \ldots, a_n\}$라는 것은 어떤 $x = (x_1, \ldots, x_n)^T$가 존재하여 $b = \sum_{i=1}^{n} x_i a_i$로 쓸 수 있다는 뜻이고, 이는 곧 $Ax = b$의 해 x가 존재한다는 것과 동치이다.

□

이는 바꿔 말하면 A의 랭크와 $A|b$의 랭크가 다르면, 해가 없다는 의미이다. 해가 없을 때, '근사해' (approximate solution)을 구하는데 이에 대해서는 나중에 다루도록 한다.

자, 그러면 우리는 언제 $Ax = b$의 해가 무수히 많은지 알 수 있다. 동일한 행렬 $A \in M_{m\times n}(\mathbb{R})$에 대해, $Ax = 0$의 해가 무수히 많을 경우 (즉, 영공간의 차원이 1 이상인 경우) $Ax = b$의 해가 무수히 많다. 이는 결국 행렬 A의 랭크가 n보다 작은 경우를 의미한다. 앞에서 임의의 행렬 A와 가우스 소거법을 적용한 기약행사다리꼴 행렬의 랭크는 서로 같고, 선형독립인 열벡터도 동일하다고 하였다. 즉, 임의의 연립일차방정식이 주어졌을 때 가우스 소거법을 적용하고, 기약행사다리꼴 행렬의 랭크를 판단하면 그 방정식이 해가 유일 ($m = n = rank(A)$)할지, 무수히 많을지 ($n > rank(A)$) 판단할 수 있다. 스스로 갖고 있는 선형대수학 책을 활용하여 연립일차방정식을 스스로 풀어보길 바란다.

12 행렬식

먼저, 행렬식 (det(A), $|A|$)은 정사각행렬 $A \in M_{n\times n}(\mathbb{R})$에 대해 정의되는 값이다. 행렬식이 중요한 가장 큰 이유는 다음과 같다.

Theorem 12.1. *임의의 정사각행렬 A가 가역행렬일 필요충분조건은 $\det(A) \neq 0$인 것이다.*

구체적인 계산은 다음과 같은 '여인수 전개' 방법을 통해 할 수 있다. $n = 2$인 경우와 $n > 2$인 경우를 나눠서 정의하겠다.

먼저, $A \in M_{2\times 2}(\mathbb{R})$의 행렬식은 다음과 같이 계산한다:

$$a_{11}a_{22} - a_{12}a_{21}$$

이때 a_{ij}는 행렬 A의 i행 j열 성분이다.

다음으로, $A \in M_{n\times n}(\mathbb{R})$ ($n > 2$)에 대해, i행과 j열을 '지워서' 얻은 $(n-1)\times(n-1)$ 크기의 행렬을 $\tilde{A}_{ij}$로 표기하자. 그러면 $|A|$는 다음과 같이 계산한다:

$$\sum_{j=1}^{n}(-1)^{1+j}a_{1j}|\tilde{A}_{1j}|$$

이는 1행에 대한 여인수 전개이며, 사실 임의의 행 또는 열에 대해 전개할 수 있다. i행에 대한 여인수 전개는 다음과 같다:

$$\det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}|\tilde{A}_{ij}|.$$

여기서 $C_{ij} = (-1)^{i+j}|\tilde{A}_{ij}|$를 (i, j)-여인수 (cofactor)라 한다.

12.1 행렬식의 라이프니츠 공식

행렬식에 대한 보다 일반적인 정의로 라이프니츠 (Leibniz) 공식이 있다:

$$\det(A) = \sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\prod_{i=1}^{n}a_{i,\sigma(i)}$$

여기서 S_n은 $\{1, 2, \cdots, n\}$의 모든 순열 (Permutation)의 집합이고, $\operatorname{sgn}(\sigma)$는 순열 σ의 부호 (짝순열이면 $+1$, 홀순열이면 -1)이다. 이 공식은 행렬식의 본질적인 정의를 제공하며, 행렬식이 '교대 다중선형 형식' (alternating multilinear form)이라는 사실과 직결된다.

12.2 행렬식의 성질

Theorem 12.2. *임의의 행렬 $A \in M_{n\times n}(\mathbb{R})$에 대하여 기본행연산이 행렬식에 미치는 영향은 다음과 같다:*

1. *A의 두 행을 교환하여 얻은 행렬을 B라 할 때, $|B| = -|A|$이다.*
2. *A의 한 행에 0이 아닌 실수 k를 곱한 행렬을 B라 하면, $|B| = k|A|$이다.*
3. *A의 한 행에 다른 행의 실수배를 더하여 얻은 행렬을 B라 하면, $|B| = |A|$이다.*

우리는 가우스 소거법의 절반 (위의 행에서 아래의 행으로 기본 행연산 수행) 과정을 통해 '상삼각행렬'을 만들 수 있고, 다음의 사실을 통해 행렬식을 얻을 수 있다.

Theorem 12.3. *상삼각행렬 (또는 하삼각행렬)의 행렬식은 대각성분의 곱과 같다. 예를 들어, 단위 행렬의 행렬식은 1이다.*

이는 앞에서 소개한 여인수 전개보다 더 효율적인 경우가 많고, 실제로 행렬의 크기가 살짝만 커지더라도 연산량을 상당히 줄여준다. 그 외의 유용한 행렬식의 추가적인 성질은 다음과 같다:

Theorem 12.4. *임의의 정사각행렬 $A, B \in M_{n\times n}(\mathbb{R})$와 스칼라 $c \in \mathbb{R}$에 대해:*

1. $\det(AB) = \det(A)\det(B)$ *(곱셈 공식)*
2. $\det(A^T) = \det(A)$ *(전치 불변)*
3. $\det(cA) = c^n \det(A)$
4. *A가 가역이면* $\det(A^{-1}) = \frac{1}{\det(A)}$
5. $\det(\mathbf{I}) = 1$

12.3 행렬식의 기하학적 의미

행렬식에는 기하학적 의미가 있다. 2×2 행렬 A의 두 열벡터가 이루는 평행사변형의 (부호가 있는) 넓이가 $\det(A)$이다. 일반적으로, $n \times n$ 행렬 A의 열벡터들이 이루는 n차원 평행육면체의 (부호가 있는) 부피가 $\det(A)$이다. 즉, $|\det(A)|$는 선형 변환 A가 부피를 얼마나 확대/축소하는지를 나타내는 값이다. $\det(A) < 0$이면 방향이 뒤집힘 (orientation reversal)을 의미한다.

12.4 수반행렬과 크라메르 공식

여인수 행렬 (cofactor matrix) C의 전치를 수반행렬 (adjugate matrix)이라 하고, $\mathrm{adj}(A) = C^T$로 표기한다.

Theorem 12.5. *가역행렬 A에 대하여, $A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$이다.*

이를 이용하여 크라메르 공식 (Cramer's rule)을 유도할 수 있다.

Theorem 12.6 (Cramer's rule)**.** *가역행렬 $A \in M_{n\times n}(\mathbb{R})$에 대해, $Ax = b$의 유일한 해의 i번째 성분은 다음과 같다:*

$$x_i = \frac{\det(A_i)}{\det(A)}$$

여기서 A_i는 A의 i번째 열을 b로 대체한 행렬이다.

크라메르 공식은 이론적으로 아름답지만, 실제 계산에서는 가우스 소거법보다 비효율적이다. n이 커지면 행렬식 계산의 복잡도가 매우 높아지기 때문이다. 하지만, 해의 특정 성분만 필요하거나, 행렬식이 이미 알려진 경우에는 유용할 수 있다.

12.5 Trace와의 관계

행렬의 대각합 (Trace)은 $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$로 정의된다. Trace와 행렬식은 고윳값과 다음과 같은 관계를 가진다:

Theorem 12.7. *행렬 $A \in M_{n\times n}(\mathbb{R})$의 고윳값이 $\lambda_1, \cdots, \lambda_n$ (중복 포함)이면:*

1. $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$ *(고윳값의 합)*
2. $\det(A) = \prod_{i=1}^{n} \lambda_i$ *(고윳값의 곱)*

또한, Trace의 중요한 성질로 $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ (순환성, cyclic property)가 있다. 이는 일반적으로 $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$로 확장된다.

13 고윳값과 고유벡터

선형대수학에서 선형방정식 $Ax = b$ 못지않게 매우 중요한 내용 중 하나인 고윳값과 고유벡터에 대해 살펴보자. 먼저 이는 '대각화 가능하다'라는 개념을 정의해야 한다. 물론 대각화 가능성과 별개로 고윳값/고유벡터 그 자체도 매우 중요한 의미들을 담고 있고 공학에서 무궁무진하게 활용된다. 예를 들어, 머신러닝에서는 '주성분 분석' (Principal component analysis)에 직접적으로 사용된다. 이러한 두 개념의 정의는 '대각화 가능'에서 유래한다.

대각화 문제는 다음과 같은 질문에 대한 답을 제공해주는 문제이다. 먼저, 유한 차원 벡터 공간 V와 선형연산자 $T : V \to V$가 주어져 있다고 하자.

1. $[T]_\beta$가 대각행렬이 되도록 하는 V의 순서기저 β가 존재하는가?
2. β가 존재한다면 어떻게 찾을 수 있는가?

결국엔 특수한 기저를 찾고자 하는 질문이다.

먼저, 대각화 행렬에 대해 정의하겠다. 유한 차원 벡터공간 V와 선형연산자 $T: V \to V$가 주어져 있다고 하자.

- $[T]_\beta$가 대각행렬이 되도록 하는 순서기저 β가 존재하면, 선형연산자 T는 '대각화 가능'이라고 한다.
- 선형연산자 $L_A(x): Ax$가 대각화 가능일 때, 정사각행렬 A는 대각화 가능이라고 한다.

순서기저 $\beta = \{v_1, v_2, \cdots, v_n\}$에 대하여 $D = [T]_\beta$가 대각행렬이면 $T(v_j) = \sum_{i=1}^n D_{ij} v_i = D_{jj} v_j = \lambda_j v_j$ ($\lambda_j = D_{jj}$)가 성립한다. 역으로, β가 적절한 스칼라 $\lambda_1, \cdots, \lambda_n$에 대하여 $T(v_j) = \lambda_j v_j$가 되도록 하는 순서기저이면, $[T]_\beta$는 대각 성분이 $\lambda_1, \cdots, \lambda_n$인 대각행렬이다. 이러한 점을 바탕으로 우리는 고윳값과 고유벡터를 정의할 수 있다. 또한, β는 '순서 기저'이므로 영벡터를 원소로 가지면 안 됨은 당연하다.

Def. Eigenvalue, Eigenvector 벡터 공간 V와 선형 연산자 $T: V \to V$에 대하여, 영벡터가 아닌 벡터 $v \in V$와 어떤 스칼라 λ가 존재하여, $T(v) = \lambda v$를 만족할 때, 벡터 v를 T의 고유벡터, λ를 v에 대응하는 T의 고윳값이라고 정의한다.

행렬 $A \in M_{n\times n}(\mathbb{R})$에 대하여, 선형 연산자 $L_A : Ax$의 고유벡터를 행렬 A의 고유벡터, 고윳값을 행렬 A의 고윳값이라 정의한다.

Theorem 13.1. *유한 차원 벡터 공간 V의 선형연산자 T가 대각화 가능이기 위한 필요충분조건은 T의 고유벡터로 이루어진 V의 순서기저 β가 존재하는 것이다. 또한, T가 대각화 가능이고 β가 T의 고유벡터로 이루어진 순서기저이면, $[T]_\beta$는 대각행렬이다. 이때 대각성분은 고윳값이다.*

Theorem 13.2. *정사각행렬 $A \in M_{n\times n}(\mathbb{R})$이 대각화 가능이기 위한 필요충분조건은 A의 고유벡터로 이루어진 $\mathbb{R}^n$의 순서기저 β가 존재하는 것이다. 또한, $A = PDP^{-1}$로 표현이 가능하며 이때 P는 열벡터가 A의 고유벡터인 행렬, D는 대각성분이 A의 고윳값인 대각행렬이다.*

이제, 첫 번째 질문에 대한 답을 어느정도 할 수 있게 되었다. 'T의 고유벡터 집합 β가 V의 순서기저가 되면 $[T]_\beta$는 대각행렬, 다른 말로 T는 대각화 가능이다.' 그렇다면 T의 고유벡터가 언제 '기저'벡터가 될 수 있을까? 당연히 벡터 공간 V의 차원만큼 선형독립인 고유벡터가 존재해야 한다. 예를 들어, V가 n차원이라면, $T: V \to V$의 고유벡터 집합이 n개의 선형 독립인 벡터 집합이어야만 한다. 그렇다면, 언제 그럴 수 있을까? 이는 2번 질문과 연결된다.

Theorem 13.3. *행렬 $A \in M_{n\times n}(\mathbb{R})$에 대하여 스칼라 λ가 A의 고윳값이기 위한 필요충분조건은 $det(A - \lambda\mathbf{I}) = 0$인 것이다. 이때 $p_A(\lambda) := \det(\lambda\mathbf{I} - A)$를 A의 특성다항식 (characteristic polynomial)이라고 부른다. (근의 집합은 $\det(A - \lambda\mathbf{I})$와 동일하다.)*

Theorem 13.4. *행렬 $A \in M_{n\times n}(\mathbb{R})$의 고윳값 λ에 대하여 $v \neq 0$이 λ에 대응하는 고유벡터이기 위한 필요충분조건은 $(A - \lambda\mathbf{I})v = 0$인 것이다.*

선형연산자 $T: V \to V$와 '임의의' 순서기저 α에 대하여 행렬 $[T]_\alpha$의 고윳값, 고유벡터는 T의 고윳값, 고유벡터이다. 따라서 선형연산자의 고윳값, 고유벡터는 어떤 순서기저를 이용하여 행렬표현을 하든 그 행렬의 고윳값, 고유벡터와 동일하다. 즉, 우리는 '행렬'의 고윳값, 고유벡터를 찾는 문제만 고려해도 충분하다.

특성 다항식은 n차 다항식이고, 이는 결국 '최대' n개의 서로 다른 고윳값을 구할 수 있다는 의미이다. 다음 정리에 의해, 서로 다른 고윳값에 대응하는 고유벡터는 선형독립임이 보장된다.

Theorem 13.5. *선형연산자 T의 서로 다른 고윳값 λ_1, λ_2를 고려해보자. λ_1에 대응하는 고유벡터 집합이 선형독립이고, λ_2에 대응하는 고유벡터 집합이 선형독립이라고 하자. 그러면, 이 두 고유벡터 집합의 합집합도 선형독립이다.*

이 정리에 의해 만약 선형연산자 $T: V \to V$ (또는 정사각행렬 $A \in M_{n\times n}(\mathbb{R})$)이 '서로 다른' 고윳값 n개를 가진다면 반드시 선형독립인 고유벡터가 적어도 n개 존재함은 보장된다. 따라서, n차원 벡터 공간의 기저 집합이 될 수 있으므로 선형연산자 T (또는 정사각행렬 A)는 대각화 가능이다. 앞에서 $n \times n$ 행렬 A의 특성 다항식은 n차 다항식이라고 하였고 이 방정식은 서로 다른 해를 최대 'n'개 가질 수 있다. 이는 서로 다른 고윳값 n개를 가진다는 의미이므로 이때는 행렬 A가 대각화 가능이다. 하지만, 만약 서로 다른 해의 개수가 'n'개보다 적은 경우에는 어떠할까? 무조건 대각화가 불가능할까? 그렇지는 않다.

Theorem 13.6. *행렬 $A \in M_{n\times n}(F)$의 특성다항식이 어떤 체의 확장(예: $F = \mathbb{R}$이면 $\mathbb{C}$)에서 다음과 같이 인수분해된다고 하자:*

$$p_A(\lambda) = \det(\lambda\mathbf{I} - A) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i} \qquad (\lambda_1, \ldots, \lambda_s\text{는 서로 다름, } m_1 + \cdots + m_s = n).$$

각 고윳값 λ_i에 대응하는 고유공간의 차원(=기하적 중복도)이 대수적 중복도 m_i와 같다면, 행렬 A는 대각화 가능이다.

특성 다항식이 n개의 일차식으로 인수분해되지 않는다면 서로 다른 고윳값이 n개보다 적게 존재할 것이다. 그리고 이는 적어도 하나의 일차식이 2개 이상 존재한다는 의미이다. (= 중근을 가짐) 이때 일차식이 중복되는 개수를 '중복도' (=대수적 중복도)라고 한다. 그리고 대수적 중복도가 m $(1 \leq m \leq n)$인 고윳값에 대응하는 선형독립인 고유벡터가 m개 존재하면, 행렬 A는 대각화 가능인 것이다. 여기서 한 가지 용어를 정의하도록 하자.

Eigenspace : 벡터 공간 V와 선형연산자 T의 고윳값 λ에 대하여, 다음의 부분 공간 E_λ를 λ에 대응하는 T의 '고유공간' (Eigenspace) 이라고 한다:

$$E_\lambda := \{x \in V : T(x) = \lambda x\}$$

이는 곧 선형 변환 $T - \lambda \mathbf{I}_V$의 영공간 (null space)임을 알 수 있을 것이다. 행렬 $A \in M_{n\times n}(\mathbb{R})$의 고윳값 λ에 대응하는 고유공간도 이와 유사하게 정의된다. Theorem 13.6.은 이 고유공간의 차원에 대해 알려주는 것이다. λ의 (대수적) 중복도가 m이면, 고유공간 E_λ의 차원은 1이상 m이하라는 의미이다. 그리고 E_λ의 차원이 정확히 m일 때, 모든 고윳값에 대하여 이것이 성립할 때 선형 변환 T는 대각화 가능이다. 즉, 각각의 λ에 대응하는 선형독립인 고유벡터 집합으로 V의 순서기저를 만들 수 있다는 의미이다. 이 논리는 정사각행렬에 대해서도 동일하게 적용된다.

14 직합 : Direct sum

유한 차원 벡터공간 V와 선형연산자 $T : V \to V$가 주어져 있다고 하자. T가 대각화 가능하면 V를 여러 부분공간들의 합집합으로 표현할 때, 유용한 점을 제공해준다. 직합은 선형대수학 후반부에서 다룰 '스펙트럼 정리' (Spectral theorem)과도 직접적으로 연결된 개념으로 중요하다.

Direct sum: 벡터공간 V의 부분공간 $W, W_1, W_2, \cdots, W_k$가 있다고 하자. 다음을 만족하는 W를 $W_1, \cdots, W_k$의 '직합'이라고 하고 $W = W_1 \oplus W_2 \oplus \cdots \oplus W_k$로 표기한다:

$$W_i \subseteq W,\ (1 \leq i \leq k).$$

$$W = \sum_{i=1}^{k} W_i,\ W_j \cap \sum_{i \neq j} W_i = \{0\}\ (1 \leq j \leq k).$$

이때 '합' ($\sum$) 기호는 다음을 의미한다:

$$W = \{w_1 + w_2 + \cdots + w_k : w_i \in W_i,\ 1 \leq i \leq k\}.$$

다음의 정리는 직합의 정의와 필요충분관계인 명제들이다.

Theorem 14.1. *유한차원 벡터공간 V가 그 부분공간 $W_1, W_2, \cdots, W_k$의 직합이라고 하자.*

- *모든 $v \in V$마다 $v = v_1 + v_2 + v_3 + \cdots + v_k$ (각 $v_i \in W_i$)로 표현하는 방법이 유일하다.*
- *각 W_i의 순서기저 β_i에 대하여, $\beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$는 V의 순서기저이다. 역으로, $\beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$가 V의 순서기저가 되도록 하는 각 W_i의 순서기저 β_i가 존재한다.*

직합의 개념을 이용하여 우리는 대각화 가능을 다음과 같이 설명할 수 있다.

Theorem 14.2. *유한차원 벡터공간 V의 선형연산자 T가 대각화 가능이기 위한 필요충분조건은 V가 T의 고유공간의 직합인 것이다.*

15 내적 공간 : Inner product space

벡터 공간에 내적을 정의하면서 우리는 '길이', '거리', '직교함' 과 같은 기하학적 양을 도입할 수 있게 된다. 내적 공간은 미적분, 확률론 등 모든 수학에서 광범위하게 녹아있는 공간이다. 먼저 내적 (inner product)을 정의하자.

Def. 내적 실수 벡터 공간 $V/\mathbb{R}$에서 *내적* (inner product)은 함수

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

로서, 임의의 $x, y, z \in V$와 임의의 $a \in \mathbb{R}$에 대해 다음을 만족한다:

1. (첫 번째 성분에 대한 선형성) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$이고 $\langle ax, y \rangle = a\langle x, y \rangle$이다.
2. (대칭성) $\langle x, y \rangle = \langle y, x \rangle$이다.
3. (양의 정부호성) $\langle x, x \rangle \geq 0$이며, $\langle x, x \rangle = 0 \iff x = 0$이다.

내적이 정의된 벡터 공간을 내적 공간 (inner product space)이라고 한다. 우리가 흔히 봐왔던 내적인 점곱 (Dot product, 유클리드 내적)은 다음과 같이 정의된다. $x, y \in \mathbb{R}^n$에 대해서,

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$

이다. 우리는 같은 크기의 행렬들의 집합이 $M_{m\times n}(\mathbb{R})$ 벡터공간임을 앞에서 다루었다. 이 벡터 공간에도 우리는 내적을 정의할 수 있다. 참고로 대표적인 내적이 프로베니우스 내적 (Frobenius inner product)이다. 임의의 $A, B \in M_{m\times n}(\mathbb{R})$에 대해,

$$\langle A, B \rangle = tr(B^T A) = tr(A^T B)$$

와 같이 정의되고 이는 내적의 조건 $1, 2, 3$을 모두 만족한다. 이제 내적을 이용해서 정의되는 거리 (크기), 직교성을 살펴보도록 하겠다.

Def. Norm 벡터 공간 $V/\mathbb{R}$에서 *노름* (norm)은 함수 $\|\cdot\| : V \to \mathbb{R}$로서, 임의의 $x, y \in V$와 실수 $a \in \mathbb{R}$에 대하여 다음을 만족한다:

1. (양의 정부호성) $\|x\| \geq 0$이고 $\|x\| = 0 \iff x = 0$이다.
2. (절대 동차성) $\|ax\| = |a| \, \|x\|$이다.
3. (삼각 부등식) $\|x + y\| \leq \|x\| + \|y\|$이다.

노름이 주어진 벡터 공간을 normed space라고 한다. 또한 노름으로부터 거리 $d(x, y) := \|x - y\|$를 정의하면 d는 (수학적 의미에서의) 거리함수(metric)가 된다.

Theorem 15.1. *벡터공간 V가 Inner product space이면 Normed space이다. 하지만, 그 역은 성립하지 않는다.*

이는 바꿔 말하면 Inner product space보다 Normed space가 더 큰 집합에 속한다는 의미이다. Normed space가 Inner product space가 되기 위해서는 몇 가지 조건들이 추가적으로 더 필요하고 이에 대해서는 본 강의에서 다루지 않는다. Inner product space이면 Normed space이기 때문에 우리는 inner product를 이용해서 norm을 정의할 수 있다:

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

위와 같이 정의된 Norm은 앞서 Norm의 정의를 모두 만족한다. 1번과 2번 성질은 쉽게 확인되는데 3번 성질은 코시-슈바르츠 부등식 (Cauchy-Schwarz inequality)를 활용하여 확인된다. 코시-슈바르츠 부등식이란 다음을 의미한다.

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$$

코시-슈바르츠 부등식 또한 매우 자주 사용되는 공식으로 잘 기억해두자. 다음은 내적공간의 Norm과 관련된 몇 가지 유용한 공식들이다.

1. $\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2$
2. $\|x - y\|^2 = \|x\|^2 - 2\langle x, y \rangle + \|y\|^2$
3. $\langle x, y \rangle = \frac{1}{4}\|x + y\|^2 - \frac{1}{4}\|x - y\|^2$

Norm은 '거리 함수' (metric, distance function)으로 사용될 수 있다. 즉, 두 벡터 x, y 사이의 거리 $d(x, y)$는 내적 공간에서 $d(x, y) = \|x - y\|$로 정의할 수 있다. 즉, 내적 공간은 Norm과 metric을 유도하는 공간이다. 마지막으로, 행렬의 Frobenius inner product를 통하여 행렬의 norm을 정의할 수 있는데 $\|A\|_F^2 = tr(A^T A) = tr(AA^T)$ 이는 결국 행렬 모든 성분들의 제곱의 합이다. (참고로, 주대각성분의 합인 Trace는 '정사각 행렬'에 대해서만 정의하는 값이며, 임의의 두 행렬 $A, B \in M_{n\times d}(\mathbb{R})$에 대하여, $tr(AB^T) = tr(BA^T)$이다.)

다음은 직교성이다. 내적 공간에서 두 벡터 x, y가 서로 직교함 (orthogonal)은 다음과 같이 정의한다:

$$\langle x, y \rangle = 0.$$

임의의 벡터의 norm이 1이면 이 벡터를 우리는 단위 벡터 (unit vector)라고 부른다:

$$\|x\| = 1.$$

만약 두 벡터가 norm이 1이고 서로 직교하면 우리는 이들이 서로 '정규직교' (orthonormal)라고 한다. 0이 아닌 벡터에 자기 자신의 norm으로 나누어주면 (실수배) 이 벡터는 크기가 1인 정규벡터가 되고 이를 '정규화' (normalizing)라고 한다. 내적 공간은 기하학적 양인 '거리', '크기', '직교성' 등의 기하학적 개념들을 자연스럽게 정의해주는 공간임을 알 수 있다.

16 벡터의 외적과 역행렬 정리

조금은 뜬금없게 느껴질 수 있지만 여기서 유클리드 공간 벡터의 외적과 역행렬과 관련된 유용한 정리를 몇 개 살펴보자. 이것을 지금 이야기하는 이유는 내적을 배워야 이야기할 수 있는 내용들이기 때문이다.

먼저 유클리드 공간의 내적은 일반적으로 점곱 (dot product)이며, 두 벡터 $u, v \in \mathbb{R}^n$에 대해

$$\langle u, v\rangle = u^T v = v^T u$$

로 정의된다. 이 절에서 말하는 "외적"은 $\mathbb{R}^3$에서의 벡터곱 $(u \times v)$이 아니라, 두 벡터로부터 행렬을 만드는 *outer product*를 의미하며 uv^T로 쓴다. 일반적으로 $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$이면

$$uv^T \in M_{m\times n}(\mathbb{R})$$

이고, 특히 $u, v \in \mathbb{R}^n$이면 $uv^T \in M_{n\times n}(\mathbb{R})$이다. $u \neq 0$, $v \neq 0$이면 uv^T의 모든 열이 u의 스칼라배이므로 $\mathrm{rank}(uv^T) = 1$인 rank-1 행렬이 된다. 이를 이용하면 행렬의 곱셈을 "rank-1 행렬들의 합"으로 이해할 수 있다.

Theorem 16.1. *행렬 $A \in M_{n\times k}(\mathbb{R})$과 행렬 $B \in M_{k\times d}(\mathbb{R})$이 있다. 이 두 행렬의 곱 AB는 k개의 외적 행렬의 합으로 표현할 수 있다. 이때 a_i $(1 \leq i \leq k)$는 행렬 A의 i번째 열벡터이고, b_i $(1 \leq i \leq k)$는 행렬 B의 i번째 행벡터이다.*

$$AB = \sum_{i=1}^{k} a_i b_i$$

다음은 행렬 역변환의 보조 정리이다.

Theorem 16.2. *$A \in M_{n\times n}(\mathbb{R})$는 가역행렬이고, $u, v \in \mathbb{R}^n$라고 하자. (단, $u, v \neq 0$). 이때 $A + uv^T$가 가역행렬인 것과 $v^T A^{-1} u \neq -1$은 필요충분조건이다. 이 경우 다음과 같이 역행렬을 계산할 수 있다.*

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}$$

증명은 강의 시간을 참고한다. 다음은 푸시-스루 항등식 (Push-Through Identity)이다.

Theorem 16.3 (Push-through identity)**.** *행렬 $A \in M_{n\times d}(\mathbb{R})$와 $B \in M_{d\times n}(\mathbb{R})$에 대하여, $(\mathbf{I}_n + AB)$가 가역이면 $(\mathbf{I}_d + BA)$도 가역이며 다음이 성립한다:*

$$A(\mathbf{I}_d + BA)^{-1} = (\mathbf{I}_n + AB)^{-1}A.$$

특히 임의의 행렬 $D \in M_{n\times d}(\mathbb{R})$와 임의의 양의 실수 $\lambda > 0$에 대해,

$$D^T(\lambda\mathbf{I}_n + DD^T)^{-1} = (\lambda\mathbf{I}_d + D^T D)^{-1}D^T$$

가 성립한다.

이 항등식은 꽤 중요하며 증명은 강의 시간을 참고한다.

퀴즈 1:
행렬 A와 C에서 추출한 벡터의 외적의 가중합 (weighted sum)으로 행렬 ABC의 곱셈을 표현해보시오. 이때, 가중치는 행렬 B에서 추출한다. (행렬 곱셈은 잘 정의된다고 가정한다.)
힌트 : 행렬 곱셈 ABC의 임의의 i행 j열을 (정의대로) 식을 작성하면, 쉽게 보일 수 있다. 퀴즈 2:
행렬 $D \in M_{n\times d}(\mathbb{R})$이고, 각각의 열의 합은 모두 0이다. 행렬 $A \in M_{d\times d}(\mathbb{R})$에 대하여, DA 각 열의 합도 0임을 보이시오. 힌트 : 모든 성분이 1인 행벡터 $\mathbf{1}^T$를 활용하라.

17 그람-슈미츠 직교화

선형대수학 초반부에서 우리는 모든 벡터공간은 기저를 가짐을 배웠다. (**Theorem** 4.4. 참고) 이는 내적공간에도 동일하게 적용된다. 하지만, 내적 공간에서는 이보다 한 단계 더 나아가 다음과 같이 선언할 수 있다.

Theorem 17.1. *모든 유한 차원 내적 공간은 정규직교기저를 갖는다.*

여기서 정규직교기저란 정규직교하는 순서기저를 의미한다. 즉, $\beta = \{v_1, \cdots, v_n\}$이 V의 정규직교기저이면 이들의 크기는 모두 1이고 서로 직교하는 V의 기저벡터인 것이다. 예를 들어, $\mathbb{R}^n$의 표준 기저는 정규직교기저이다. 또한, 무한 차원 내적공간의 직교기저에 대한 이야기는 학부 선형대수학 수준을 많이 넘어서기 때문에 본 강의록에서는 다루지 않는다.

위 Theorem이 참인 이유는 우리가 임의의 기저 벡터를 정규 직교기저로 만들 수 있기 때문이다. 이 방법론이 '그람-슈미츠 직교화' (Gram-Schmidt orthogonalization)이다. 먼저 정규직교기저가 왜 중요한지 다음의 정리를 통해 살펴보자.

Theorem 17.2. *내적공간 V와 정규직교기저 $\beta = \{v_1, \cdots, v_n\}$가 있다. 그러면 임의의 $x \in V$에 대해 다음이 성립한다:*

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$$

Proof. **¡1¿ Goal.** 임의의 $x \in V$에 대해 $x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i$임을 보인다.

¡2¿ β가 기저이므로, 어떤 계수 $a_1, \ldots, a_n \in \mathbb{R}$가 유일하게 존재하여

$$x = \sum_{i=1}^{n} a_i v_i$$

로 쓸 수 있다.

¡3¿ 각 $j \in \{1, \ldots, n\}$에 대해 양변에 v_j와의 내적을 취하면

$$\langle x, v_j \rangle = \left\langle \sum_{i=1}^{n} a_i v_i,\ v_j \right\rangle = \sum_{i=1}^{n} a_i \langle v_i, v_j \rangle = \sum_{i=1}^{n} a_i \delta_{ij} = a_j$$

이다. (정규직교기저이므로 $\langle v_i, v_j \rangle = \delta_{ij}$.)

¡4¿ 따라서 $a_i = \langle x, v_i \rangle$이고, 이를 (2)에 대입하면 원하는 식이 성립한다.

□

우선, 유한 개의 벡터가 정규 직교이면 서로 선형독립이다. 이에 대한 증명은 어렵지 않으므로 스스로 해보자. 우리는 선형대수학 초반부에서 순서기저가 주어지면 임의의 벡터를 순서기저의 선형결합으로 '유일하게' 표현할 수 있음을 배웠고 이때 계수를 '좌표'라고 이름붙였다. 내적공간에서 정규직교기저가 주어지면 당연히 '기저'이므로 임의의 벡터를 순서기저의 선형결합으로 '유일하게' 표현할 수 있는데 더 나아가 '계수'까지도 구체적으로 그 값을 알 수 있다. 그 계수를 수학에서는 '푸리에 계수' (Fourier coefficient)라고 부르며 이것이 **Theorem** 16.2.의 내용이다.

이제 그람-슈미츠 직교화를 살펴보자. 먼저 두 선형독립인 벡터 $\{w_1, w_2\}$가 주어져 있다고 하자. 이를 이용해 정규직교기저 $\{v_1, v_2\}$를 얻고자 한다. 먼저, $w_1 = v_1$으로 둔다. 다음으로 v_2는 w_2가 $w_1 (= v_1)$과 직교하도록 만들어준다. 이는 w_2에서 w_1과 직교하는 부분만을 분리해주는 작업이라 생각할 수 있다. 즉, $v_2 = w_2 - cw_1$ ($c \in \mathbb{R}$)로 두고 $v_2 \perp w_1$이 되도록 c를 찾아야 한다. 직교의 정의는 내적이 0인 것이므로,

$$\langle v_2, w_1 \rangle = \langle w_2 - cw_1, w_1 \rangle = \langle w_2, w_1 \rangle - c\langle w_1, w_1 \rangle = 0.$$

즉, $c = \frac{\langle w_2, w_1 \rangle}{\|w_1\|^2}$이다. 따라서

$$v_1 = w_1,\ v_2 = w_2 - \frac{\langle w_2, w_1 \rangle}{\|w_1\|^2} w_1$$

이다. 이 과정을 귀납적으로 반복하면 임의의 선형독립인 유한 개의 벡터들을 정규직교인 벡터로 만들 수 있다. 이를 '그람-슈미츠 직교화'라고 한다.

Theorem 17.3 (Gram–Schmidt 직교화)**.** *내적공간 V와 선형독립인 순서집합 $\{w_1, w_2, \cdots, w_n\}$이 주어져 있다. 다음과 같이 벡터들을 정의하자:*

$$v_1 = w_1, \qquad v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\langle v_j, v_j \rangle} v_j \quad (2 \le k \le n).$$

그러면 $v_1, \cdots, v_n$은 모두 0이 아니고 서로 직교하며, $W := \mathrm{span}\{w_1, \cdots, w_n\}$의 직교기저가 된다. 더 나아가 각 k에 대해

$$u_k := \frac{v_k}{\|v_k\|}$$

로 정규화하면 $\{u_1, \cdots, u_n\}$은 W의 정규직교기저이다.

이제 우리는 임의의 (유한 차원) 내적 공간에서 항상 정규직교기저를 찾을 수 있다. 다음은 직교여공간 (orthogonal complement)이다.

Def. 직교여공간: 내적 공간 V의 부분집합 S에 대하여, S의 모든 원소와 수직인 V의 부분집합을 $S^\perp$이라고 하자. 이를 S의 직교여공간이라 한다.

$$S^\perp := \{x \in V : \forall y \in S,\ \langle x, y \rangle = 0\}$$

이때 $S^\perp$은 벡터공간이 되기 때문에 V의 '부분공간'이다. 직교여공간은 선형대수학 응용에서 꽤 유용한데 학부 선형대수학 수준에서는 직교여공간은 '정사영'과 직접적으로 연결되는 개념이다. 그리고 정사영은 머신러닝의 가장 기본적인 형태인 선형회귀의 '최소제곱법' 문제를 푸는 방법론이 된다.

18 정사영과 최소제곱법

정사영에 대한 엄밀한 정의는 선형대수학 극후반부에서 배울 것이다. 지금은 그냥 '수선의 발을 내림' 정도의 고등학교 수준으로만 생각해도 충분하다.

다음과 같은 상황을 생각해보자. 내적 공간 V와 그 부분 공간 W가 있다. 어떤 $x \in V$에 대해, x와 가장 거리가 가까운 y를 찾고자 한다. 즉, $\|x - y\|$가 가장 작은 $y \in W$를 찾고자 한다. 이는 다양한 상황에서 응용된다. 이때 필연적으로 등장하는 것이 직교이다. 즉 위 문제는 x와 직교하는 y를 찾는 문제로 환원된다. 이에 대해 이야기하도록 하겠다. 먼저, 다음의 정리는 중요한 관점을 제공해준다.

Theorem 18.1. *내적 공간 V의 부분 공간 W와 벡터 $x \in V$에 대하여, $x = u + z$인 유일한 $u \in W$, $z \in W^\perp$이 존재한다. 또한, W의 정규직교기저를 $v_1, \cdots, v_k$라 했을 때, 다음이 성립한다:*

$$u = \sum_{i=1}^{k} \langle x, v_i \rangle v_i.$$

또한, u는 x에 가장 가까운 벡터이다.

이때 u를 x의 정사영 벡터 (orthogonal projection) 라고 한다. 최소제곱법은 정사영이 응용되는 가장 대표적인 예시이다. 이에 대해 살펴보기 전에 정리 하나를 이야기하고 넘어가겠다.

Theorem 18.2. *n차원 내적 공간 V와 정규직교집합 $S = \{v_1, v_2, \cdots, v_k\}$가 주어져 있다고 하자. 그러면 다음이 성립한다.*

- *S를 확장하여 V의 정규직교기저 $\{v_1, \cdots, v_k, v_{k+1}, \cdots, v_n\}$을 얻을 수 있다.*
- *$W = span(S)$일 때, $W^\perp$의 정규직교기저는 $\{v_{k+1}, v_{k+2}, \cdots, v_n\}$이다.*
- *$V = W \oplus W^\perp$이다.*

앞에서 우리는 선형방정식 $Ax = b$의 해가 유일한 경우와 무수히 많은 경우들에 대해 살펴보았다. 최소제곱법은 $Ax = b$의 해가 없는 경우 어떻게 가장 근사적인 해 (approximated solution)를 찾을 것인가를 알려주는 방법이다. 즉, $A\tilde{x} \approx b$인 $\tilde{x}$를 찾는 방법이다. $A \in M_{m\times n}(\mathbb{R}), x \in \mathbb{R}^n, b \in \mathbb{R}^m$이라 하자. 최소제곱법 문제를 수식으로 표현하면 다음과 같다:

$$\tilde{x} = \arg\min_x \|Ax - b\|^2.$$

즉 b와 가장 거리가 가까운 Ax를 찾는 것이다. 앞서 우리는 행렬과 벡터의 곱셈은 행렬의 열벡터의 선형 결합으로 볼 수 있음을 배웠다. 즉, Ax는 행렬 A의 열벡터 $[a_1, a_2, \cdots, a_n]$의 선형결합, $\{x_1a_1 + x_2a_2 + \cdots + x_na_n\}$이다. 따라서 b와 가장 거리가 가까운 A의 열공간 (column space)의 원소를 찾는 문제로 해석할 수 있다. 이는 우리가 앞에서 본 정사영 개념과 정확히 대응된다. $\mathbb{R}^m$의 부분공간인 A의 열공간 $col(A)$이 있다. 우리는 $b \in \mathbb{R}^m$과 가장 거리가 가까운 $A\tilde{x} \in col(A)$를 찾고자 하는 것이다. **Theorem** 18.1.에 의해 $b = A\tilde{x} + (b - A\tilde{x})$로 표현하고 $A\tilde{x} \in col(A)$, $b - A\tilde{x} \in col(A)^\perp$을 만족하는 $b - A\tilde{x}$가 유일하게 존재함을 보장할 수 있다. 그리고 이때 $A\tilde{x}$는 $col(A)$의 원소 중에서 b와 가장 가까운 원소이다. 즉, $\mathbb{R}^m = col(A) \oplus col(A)^\perp$이다. $A\tilde{x}$와 $b - A\tilde{x}$는 서로 직교해야 하므로 내적값이 0이라는 사실을 이용하면 다음과 같은 값을 얻을 수 있다:

$$\tilde{x} = (A^TA)^{-1}A^Tb.$$

물론, $(A^TA) \in M_{n\times n}(\mathbb{R})$이 항상 가역행렬이란 보장은 없다. 다음은 이 행렬이 언제 가역행렬인지 알려준다.

Theorem 18.3. *임의의 행렬 A에 대하여, $rank(A) = rank(A^T) = rank(A^TA) = rank(AA^T)$이다.*

Theorem 18.4. *행렬 $A \in M_{m\times n}(\mathbb{R})$의 랭크가 n이라고 하자. 즉, $rank(A) = n$이다. 그러면 A^TA의 랭크도 n이고, 따라서 가역행렬이다.*

19 좌표 변환 행렬

선형대수학 전체에 걸쳐 제일 중요한 개념 중 하나는 '기저'이다. 벡터 공간 $V/\mathbb{R}$이 주어져 있고, 서로 다른 두 순서기저 집합 β, β'이 주어져 있다고 하자. 그러면, 동일한 선형 변환 (= 선형 연산자)$T : V \to V$는 서로 다른 두 행렬 표현을 가질 것이다. 이를 각각 $A = [T]_\beta, B = [T]_{\beta'}$이라 하자. 두 행렬은 다르다. (즉, 다른 성분들을 가지고 있을 것이다.) 하지만, 두 행렬 사이에는 모종의 관계가 숨어 있다. 선형대수학에서는 이 두 행렬이 서로 '닮음'이라고 부른다. 이를 알아보기 위해 먼저 '좌표 변환 행렬'을 정의하자.

Def. 좌표 변환 행렬 **(Change of coordinate matrix)** : 두 순서기저 β, β'이 주어져 있을 때 $Q = [I_V]_{\beta'}^{\beta}$를 β' 좌표를 β 좌표로 변환하는 행렬이라고 한다. 이 행렬은 항상 가역행렬이고, Q^{-1}는 β 좌표를 β' 좌표로 변환해주는 행렬이다.

$$\forall v \in V,\ [v]_\beta = [\mathbf{I}_V(v)]_\beta = [\mathbf{I}_V]_{\beta'}^{\beta}[v]_{\beta'} = Q[v]_{\beta'}.$$

Theorem 19.1. *유한 차원 벡터 공간 V와 선형 연산자 $T : V \to V$, 두 순서기저 β, β'이 주어져 있다. Q가 β' 좌표를 β 좌표로 변환하는 행렬이라고 하면, 다음이 성립한다:*

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q.$$

즉, 같은 선형 변환을 서로 다른 기저로 보면 행렬은 다르지만, 이 두 행렬은 닮음 행렬이다.

즉, 동일한 선형 변환을 서로 다른 순서기저에 대한 행렬표현을 할 때, 행렬의 모양은 다르지만 두 행렬 사이에는 모종의 관계가 성립한다는 것을 보여준다. 위 관계를 선형대수학에서는 '닮음'이라고 한다. 이는 수학에서 중요한 한 가지 아이디어를 제공해준다. 복잡해 보이는 행렬 (선형변환)을 다른 기저를 선택하면 단순한 행렬로 표현할 수 있다는 점이다. 이는 앞서 고윳값/고유벡터 관련 내용들에서 대각화 가능한 행렬 (선형변환)에 대해 그 고유벡터들을 기저로 선택하면, 대각행렬이라는 매우 단순한 행렬로 표현할 수 있다는 것과도 연결된다. 대각화 가능한 행렬 $A \in M_{n\times n}(\mathbb{R})$은 다음이 성립한다:

$$A = PDP^{-1}.$$

P는 A의 고유벡터를 열벡터로 하는 행렬, D는 A의 고윳값을 대각성분으로 하는 대각행렬이다. A와 D는 닮음 행렬이고, 이는 어떤 동일한 선형 변환을 서로 다른 기저에서 바라본 결과이다. 즉, 선형변환을 고유벡터 기저 집합에 대한 행렬표현으로 보면, 단순히 실수배 (scaling)를 수행하고 있는 것임을 알려주는 것이다. 이는 선형 변환을 원래의 기저 집합으로 바라봤을 때에는 볼 수 없는 직관을 제공해준다.

20 대칭행렬, 직교행렬 그리고 양의 (준)정부호 행렬

먼저, 대칭행렬 (= 자기수반행렬, 에르미트 행렬) 의 정의는 $A = A^T$인 행렬 A를 의미한다. 전치 행렬과 같아야 하므로, 대칭행렬은 정사각행렬이다. ($A \in M_{n\times n}(\mathbb{R})$) 전치와 관련된 기본 연산 공식은 외워둔다.

$$(A^T)^T = A$$
$$(A^T)^{-1} = (A^{-1})^T$$
$$(A+B)^T = A^T + B^T$$
$$(AB)^T = B^T A^T$$

2번 공식은 A가 가역행렬일 때 성립하는 것은 당연하다. '전치 연산' (Transpose)는 꽤나 심오한 의미를 가지고 있다. 이에 대해 살펴보자. 먼저 벡터 공간 (혹은 내적 공간) V에서 실수 $\mathbb{R}$로 가는 선형 변환 $T : V \to \mathbb{R}$을 선형 범함수 (linear functional)라고 한다. 이러한 선형 범함수의 집합 $L(V, \mathbb{R})$을 V의 쌍대 공간 (dual space)이라고 하며, $V^\star$로 표기한다. V가 유한 차원이면, $V^\star$도 V와 같은 차원이며, 따라서 둘은 동형이다. n차원 벡터 공간 V와 m차원 벡터 공간 W, 그리고 각각의 순서기저 β, γ가 있을 때, 선형 변환 $T : V \to W$와 행렬 $A = [T]_\beta^\gamma$ 사이에는 일대일 대응 관계가 성립함을 배웠었다. 그렇다면, A^T와 연결되는 선형 변환은 어떻게 될까? 여기에서 쌍대 공간이 등장한다.

Theorem 20.1. *임의의 선형 변환 $T : V \to W$에 대하여, 함수 $T^t : W^\star \to V^\star$를 다음과 같이 정의하자.*

$$\forall g \in W^\star, T^t(g) = g \circ T.$$

그러면, T^t는 선형 변환이고, $[T^t]_{\gamma^\star}^{\beta^\star} = ([T]_\beta^\gamma)^T$이다.

즉, 선형 변환 (혹은 행렬)의 전치는 쌍대 공간에서 쌍대 공간으로 가는 선형 변환을 의미한다.

Remark 20.1 (전치(dual map)와 수반(adjoint)의 관계)**.** *위에서 정의한 T^t는 쌍대공간 사이의* dual map*이다. 한편 V, W가 내적공간이면, Riesz 표현정리에 의해 $V \simeq V^\star$, $W \simeq W^\star$로 식별할 수 있고, 이 식별 아래에서 T^t는 내적에 대한 수반 연산자 T^*와 같은 역할을 한다. 특히 V, W에서 정규직교기저를 택하면 T^*의 행렬표현은 T의 행렬표현의 전치(복소수의 경우 켤레전치)가 된다.*

수학에서 꽤 중요한 정리 중 하나인 Riesz-representation theorem에 대해 언급하고 대칭 행렬에 대한 이야기로 넘어가겠다. 우선, 내적공간 V가 주어져 있고, $y \in V$에 대해서 $g(x) = \langle x, y\rangle$이란 함수를 생각해보자. 함수 $g : V \to \mathbb{R}$은 선형 변환이다. 이는 내적의 정의에 의해 자명하다. 더 나아가, V가 유한 차원이면 $V \to \mathbb{R}$로의 모든 선형변환은 모두 내적으로 표현된다.

Theorem 20.2. *유한 차원 내적 공간 V와 선형 변환 $f : V \to \mathbb{R}$이 있다. 모든 $x \in V$에 대해서 $f(x) = \langle x, y\rangle$을 만족하는 벡터 $y \in V$가 유일하게 존재한다.*

이는 Riesz-representation theorem의 특수한 예시이다. 이 이론에 대한 일반적인 서술은 함수해석학에서 다루므로, 본 강의에서는 다루지 않겠다.

대칭 행렬의 기본적인 속성들을 살펴보고, 이 행렬이 왜 중요한 행렬인지 이야기하겠다. $A, B \in M_{n\times n}(\mathbb{R})$가 대칭행렬이면, 다음의 행렬들도 대칭행렬이다.

$$A + B, A - B, A^n.$$

Theorem 20.3. *$A \in M_{n\times n}(\mathbb{R})$이 대칭 행렬인 것과 A가 대각화 가능하고, 그 고유벡터 집합이 $\mathbb{R}^n$의 정규직교기저인 것은 필요충분조건이다.*

이는 스펙트럼 정리의 단순한 버전이다. 스펙트럼 정리 전체는 다음 챕터에서 다룰 것이다. 결국, 대칭행렬은 대각화 가능하고 그 고유벡터가 서로 직교한다는 것을 알려준다. 그리고 사실 유한 차원의 내적 공간에서 이러한 성질을 갖는 행렬 (또는 선형변환)은 대칭 행렬뿐이다. (필요 충분 조건이므로.) 위 정리는 수학과 공학의 무수히 많은 분야에서 응용된다.
다음은 직교 행렬이다. 직교 행렬의 정의는 다음과 같다:

$$A \in M_{n\times n}(\mathbb{R}),\ A^TA = AA^T = \mathbf{I}.$$

즉, 바꿔말하면 $A^{-1} = A^T$인 정사각 행렬 A를 직교 행렬이라고 한다. 직교 행렬은 정의 자체에서 가역 행렬임을 전제로 하고 있음을 기억하자. 또한, 각 열벡터가 서로 정규 직교이면 그 행렬은 직교 행렬임을 알 수 있다. (바꿔 말하면, $\mathbb{R}^n$의 정규 직교 기저가 된다.)
다음은 직교 행렬의 유용하고도 중요한 성질들이다.

Theorem 20.4. *직교 행렬 $A \in M_{n\times n}(\mathbb{R})$과 $\mathbb{R}^n$의 표준 내적과 그에 의해 유도된 norm에 대해 다음이 성립한다:*

1. $\|Ax\| = \|x\|$.
2. $\langle Ax, Ay\rangle = \langle x, y\rangle$.

즉, 직교행렬은 (표준) 내적과 노름을 보존해주는 행렬이다. 대칭 행렬이면서 직교인 행렬은 꽤나 막강한 성질을 가지고 있다.

Theorem 20.5. *행렬 A가 대칭이고, 직교행렬인 것과 그 고윳값이 모두 절댓값이 1이고, 고유벡터 집합이 정규 직교 기저인 것은 필요충분조건이다.*

즉, **Theorem** 20.3.에서 직교 행렬이란 조건이 추가되면 모든 고윳값의 절댓값이 1이라는 사실도 얻을 수 있다. 필요충분조건이므로, 그 역도 성립한다.
다음은 양의 (준)정부호 행렬이다.
Def. Positive (semi)-definite matrix 임의의 행렬 $A \in M_{n\times n}(\mathbb{R})$가 임의의 영벡터가 아닌 $x \in \mathbb{R}^n$에 대해, $x^TAx > 0$이면 A는 양의 정부호 행렬 (Positive definite matrix)이라고 한다.
만약, $x \in \mathbb{R}^n$에 대해, $x^TAx \geq 0$이면 양의 준정부호 행렬 (Positive semi-definite matrix)라고 한다.

Theorem 20.6. *대칭행렬 A에 대해, 이 행렬이 양의 정부호 행렬인 것과 A의 모든 고윳값이 양수인 것은 필요충분조건이다. 또한, 이 행렬이 양의 준정부호 행렬인 것과 A의 모든 고윳값이 0 이상인 것은 필요충분조건이다. 또한, A가 양의 준정부호 행렬인 것과 $A = B^TB$인 정사각행렬 B가 존재하는 것과 필요충분조건이다.*

양의 정부호 행렬은 항상 가역행렬이다. (정의를 통해 확인해볼 수 있다.) 또한, **Theorem** 20.6.에서 3번째 내용을 반대로 생각하면 임의의 정사각행렬 B에 대해, B^TB는 항상 대칭이고, 양의 준정부호 행렬임을 알 수 있다. 이는 여러 상황들에서 꽤 유용하게 적용된다.
마지막으로, 대칭이고 양의 정부호 행렬인 M은 벡터 공간의 내적을 정의할 수 있다. 예를 들어, 벡터 공간 V의 다음은 내적이 될 수 있다: (= 내적의 정의를 만족한다)

$$\langle x, y\rangle = x^TMy.$$

사실, dot product는 $M = \mathbf{I}$인 특수한 내적으로 볼 수 있다. 이러한 내적은 꽤나 다양한 응용분야에서 널리 쓰인다. 통계학 또는 머신러닝에서의 Mahalanobis distance, Precondition matrix, Riemannian metric, (직교좌표계가 아닌) 새로운 좌표계로의 좌표변환 등등에서 Dot product가 아닌 위와 같은 내적을 기반으로 분석을 수행하면, 그 직관이 더 잘 와닿는 경우들이 많이 있다. 이에 대해서는 추후 기회가 되면 공부해보길 바란다.

21 정사영과 스펙트럼 정리

먼저 직합에 대한 내용을 잘 복습해두길 바란다. (Section 14) 벡터공간 $V/\mathbb{R}$이 있을 때, 두 부분공간 W_1, W_2가 있고 $V = W_1 \oplus W_2$이면 임의의 $x \in V$에 대해 $x = x_1 + x_2$인 $x_1 \in W_1, x_2 \in W_2$가 유일하게 존재한다.
벡터 공간 V에서 선형변환 $T : V \to V$가 사영 (Projection)이라는 것은 다음을 만족한다는 것이다:

$$T^2 = T.$$

즉, 한 번 사영을 한 뒤 $T(v)$, 다시 연이어서 사영을 해도 결과가 바뀌지 않는다. 사영도 선형변환이므로 영공간 $N(T)$과 상공간 $R(T)$을 생각할 수 있다. 이때 매우 중요한 사실이 성립한다:

$$V = R(T) \oplus N(T).$$

정사영(Orthogonal projection)은 여기서 한 단계 더 나아가 임의의 벡터를 해당 부분공간에 '수직으로' 투영하는 것이다. 내적공간 V의 벡터 $v \in V$에 대해 다음을 만족하는 W의 벡터 w가 유일하게 존재한다:

$$\langle v - w, x \rangle = 0, \quad \forall x \in W.$$

이를 $w = \text{proj}_W(v)$로 표기하며, 정사영 선형변환 $T(v) = w$ 역시 $T^2 = T$를 만족할 뿐만 아니라 추가로 직교성을 갖추어 $R(T)^{\perp} = N(T)$와 $N(T)^{\perp} = R(T)$를 만족한다.

행렬 관점에서, 부분공간 W의 정규직교기저를 열벡터로 하는 행렬을 $U \in M_{n\times k}(\mathbb{R})$라 하자 ($U^T U = \mathbf{I}_k$). 그러면 정사영 행렬은 $P = UU^T$로 주어진다.

Theorem 21.1. *행렬 P가 정사영 행렬인 것의 필요충분조건은 대칭이고 멱등적인 것이다. 즉, $P = P^T = P^2$이다.*

정규직교기저가 아닐 경우, 임의의 기저 행렬 A에 대해 정사영 행렬은 $P = A(A^T A)^{-1} A^T$이다. 이는 머신러닝의 선형 회귀와 최소제곱법(Least squares)의 수학적 기반이다.

Theorem 21.2 (스펙트럼 정리, Spectral Theorem)**.** *유한 차원 내적 공간 V의 자기수반(self-adjoint) 선형 연산자 T (행렬표현은 대칭행렬 $A = A^T$)는 오직 실수인 고윳값 $\lambda_1, \cdots, \lambda_k$를 갖는다. λ_i에 대응하는 고유공간을 W_i, W_i로의 정사영을 P_i라 하면 다음이 성립한다:*

1. *$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$ (고유공간들의 직교 직합 분해).*
2. *$\mathbf{I} = P_1 + P_2 + \cdots + P_k$ (항등 연산자의 분해, Resolution of identity).*
3. *$T = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_k P_k$ (스펙트럼 분해).*

스펙트럼 정리는 ”모든 실수 대칭행렬은 직교행렬에 의해 대각화 가능하며, 그 고윳값은 모두 실수이다” 라는 선형대수학의 가장 아름다운 결론 중 하나이다.

22 케일리-해밀턴 정리와 최소다항식

이 절에서는 행렬과 다항식의 깊은 관계를 다루며, 이는 조르당 표준형과 행렬 함수 이론으로 나아가는 교두보 역할을 한다.

22.1 케일리-해밀턴 정리 (Cayley-Hamilton Theorem)

Theorem 22.1 (Cayley-Hamilton)**.** *임의의 정사각행렬 $A \in M_{n\times n}(\mathbb{R})$의 특성다항식을 $p_A(\lambda) = \det(\lambda \mathbf{I} - A)$라 하자. 그러면 다음이 성립한다:*

$$p_A(A) = \mathbf{0}.$$

즉, 모든 정사각행렬은 자기 자신의 특성 방정식을 만족한다.

수반행렬(adjugate matrix)을 이용한 고전적이고 우아한 대수적 증명은 다음과 같다.

Proof (케일리-해밀턴 정리). **¡1¿ Goal.** 임의의 $n \times n$ 행렬 A에 대하여 $p_A(A) = \mathbf{0}$임을 보인다.

¡2¿ Assume. 특성다항식을 $p_A(\lambda) = \det(\lambda \mathbf{I} - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1 \lambda + c_0$라 하자.

¡3¿ 행렬 $\lambda \mathbf{I} - A$의 수반행렬(adjugate matrix)을 $B(\lambda)$라 하면, 행렬식의 역행렬 기본 성질 $M \operatorname{adj}(M) = \det(M)\mathbf{I}$에 의해 다음 항등식이 성립한다:

$$(\lambda \mathbf{I} - A) B(\lambda) = \det(\lambda \mathbf{I} - A)\mathbf{I} = p_A(\lambda)\mathbf{I}.$$

¡4¿ Claim. $B(\lambda)$는 λ에 대한 다항식 행렬이다. **Reason.** $B(\lambda)$의 각 성분은 $\lambda \mathbf{I} - A$의 $(n-1)\times(n-1)$ 소행렬식이므로, λ에 대한 최대 $n-1$차의 다항식이다.

¡5¿ 따라서 $B(\lambda)$를 다음과 같이 전개하여 쓸 수 있다:

$$B(\lambda) = B_{n-1}\lambda^{n-1} + B_{n-2}\lambda^{n-2} + \cdots + B_1 \lambda + B_0 \quad (B_i \in M_{n\times n}(\mathbb{R})).$$

¡6¿ 위 식을 원래의 항등식에 대입하면:

$$(\lambda \mathbf{I} - A)(B_{n-1}\lambda^{n-1} + \cdots + B_0) = (\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0)\mathbf{I}.$$

¡7¿ 양변의 λ 거듭제곱 계수를 서로 비교하면 다음 행렬 연립방정식을 얻는다:

$$\begin{aligned} B_{n-1} &= \mathbf{I} \\ B_{n-2} - AB_{n-1} &= c_{n-1}\mathbf{I} \\ &\vdots \\ B_0 - AB_1 &= c_1\mathbf{I} \\ -AB_0 &= c_0\mathbf{I} \end{aligned}$$

¡8¿ 위 식들의 양변 왼쪽에 차례대로 $A^n, A^{n-1}, \cdots, A, \mathbf{I}$를 곱한다.

¡9¿ **Conclusion.** 곱해진 모든 식의 좌변과 우변을 각각 더하면, 좌변은 연쇄적으로 상쇄되어(Telescoping sum) $\mathbf{0}$이 되고, 우변은 $A^n + c_{n-1}A^{n-1} + \cdots + c_0\mathbf{I} = p_A(A)$가 된다. 따라서 $p_A(A) = \mathbf{0}$이다.

□

이 정리에 의해, 가역행렬 A의 역행렬 A^{-1}은 상수항 $c_0 \neq 0$를 이용해 A의 $(n-1)$차 다항식으로 환원할 수 있다.

22.2 최소다항식 (Minimal Polynomial)

Definition 22.1. *행렬 $A \in M_{n\times n}(\mathbb{R})$에 대해, $q(A) = \mathbf{0}$을 만족하는 최고차항의 계수가 1인 다항식 (monic polynomial) 중 차수가 가장 낮은 것을 A의 최소다항식이라 하고 $m_A(\lambda)$로 표기한다.*

Theorem 22.2. *행렬 A의 최소다항식 $m_A(\lambda)$는 유일하게 존재하며, 다음 성질을 갖는다:*

1. *$m_A(\lambda)$는 특성다항식 $p_A(\lambda)$를 나눈다.*
2. *$m_A(\lambda)$와 $p_A(\lambda)$는 대수적 중복도를 제외하면 완전히 같은 근 (고윳값)을 갖는다.*
3. *행렬 A가 대각화 가능할 필요충분조건은 $m_A(\lambda)$가 서로 다른 일차식의 곱으로 완전히 인수분해되는 것이다. (즉, 중근이 없는 것이다.)*

23 조르당 표준형 (Jordan Canonical Form)

대각화가 불가능한 행렬(기하적 중복도가 대수적 중복도보다 작은 경우)에 대해서도 '거의 대각행렬'에 가까운 대수적 표준 형태를 얻을 수 있다.

Definition 23.1. *행렬 N이 어떤 양의 정수 k에 대해 $N^k = \mathbf{0}$을 만족하면, N을 멱영행렬 (nilpotent matrix)이라 한다. 멱영행렬의 고윳값은 오직 0뿐이다.*

Definition 23.2. *크기 k인 고윳값 λ의 조르당 블록 (Jordan block) $J_k(\lambda)$은 대각성분이 모두 λ이고 바로 윗대각선의 성분이 모두 1이며 나머지는 0인 $k \times k$ 행렬이다.*

Theorem 23.1 (조르당 표준형 정리)**.** *특성다항식이 일차식의 곱으로 완전히 인수분해되는 임의의 정사각행렬 A에 대하여, 가역행렬 P가 존재하여 다음이 성립한다:*

$$P^{-1}AP = J = \begin{bmatrix} J_{k_1}(\lambda_{i_1}) & & \\ & \ddots & \\ & & J_{k_s}(\lambda_{i_s}) \end{bmatrix}$$

이때 블록 대각행렬 J를 A의 조르당 표준형이라 하며, 블록의 순서를 제외하면 유일하다.

조르당 표준형은 임의의 행렬을 $\lambda\mathbf{I} + N$ (스칼라 행렬 + 멱영행렬)의 형태로 분해하여 분석할 수 있게 해주며, 행렬 지수 함수 및 일반적인 행렬 함수를 정의할 때 필수적인 대수적 역할을 한다.

24 특잇값 분해 (Singular Value Decomposition)

정사각행렬에 대한 스펙트럼 분해를 임의의 $m \times n$ 직사각행렬로 완벽히 확장한 선형대수학의 '왕관의 보석'이다.

Theorem 24.1 (특잇값 분해 정리)**.** *랭크가 r인 임의의 행렬 $A \in M_{m\times n}(\mathbb{R})$은 다음과 같이 분해할 수 있다:*

$$A = U\Sigma V^T$$

여기서 직교행렬 $U \in M_{m\times m}(\mathbb{R})$와 직교행렬 $V \in M_{n\times n}(\mathbb{R})$, 그리고 주대각성분이 특잇값 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$이고 나머지는 0인 행렬 $\Sigma \in M_{m\times n}(\mathbb{R})$가 존재한다.

여기서 V의 열벡터들은 $A^T A$의 고유벡터(우측 특이벡터)들이고, U의 열벡터들은 AA^T의 고유벡터(좌측 특이벡터)들이다. 특잇값 σ_i는 $A^T A$의 양의 고윳값들의 제곱근이다.

이 분해를 통해 행렬 A를 랭크 1 행렬들의 합으로 나타낼 수 있으며($A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$), $k < r$에서 절단(Truncation)하면 $\|A - \hat{A}_k\|$ 오차를 최소화하는 최적의 랭크 k 근사 행렬을 얻는다 (Eckart-Young Theorem). 이는 머신러닝의 주성분 분석(PCA)과 데이터 압축의 수학적 토대이다.

25 행렬 노름과 극분해 (Matrix Norms and Polar Decomposition)

행렬 공간에 크기와 거리를 부여하는 노름(Norm)과, 특잇값 분해(SVD)와 밀접하게 연관된 기하학적 분해법인 극분해(Polar Decomposition)를 다룬다.

25.1 행렬 노름 (Matrix Norms)

행렬 공간 $M_{m\times n}(\mathbb{R})$에 주어진 노름 $\|\cdot\|$은 벡터 노름의 3가지 공리를 만족하며, 추가로 정사각행렬의 곱셈에 대하여 부분 준동형성 **(Sub-multiplicativity)**을 갖는 경우가 많다:

$$\|AB\| \le \|A\|\|B\|.$$

1. 유도된 노름 (Induced Norms / Operator Norms)
주어진 벡터 노름 $\|\cdot\|_p$에 대하여, 행렬 A가 벡터의 길이를 최대로 얼마나 늘리는지를 측정하는 작용소 노름이다:

$$\|A\|_p := \max_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p=1} \|Ax\|_p.$$

- $\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |A_{ij}|$ (최대 절대 열 합 노름)
- $\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}|$ (최대 절대 행 합 노름)
- $\|A\|_2 = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^T A)}$ (스펙트럼 노름, Spectral norm). 행렬의 가장 큰 특잇값과 같으며 행렬의 조건수 계산에 핵심적으로 쓰인다.

2. 섀튼 노름과 성분별 노름 (Schatten and Entry-wise Norms)
행렬의 특잇값 분포 전체를 요약하는 섀튼 p-노름은 $\|A\|_{\mathcal{S}_p} = (\sum \sigma_i^p)^{1/p}$으로 정의된다.

- 프로베니우스 노름 **(Frobenius Norm, $p = 2$):** $\|A\|_F := \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\mathrm{tr}(A^T A)} = \sqrt{\sum_i \sigma_i^2}$. 계산이 용이하고 미분 가능성이 우수하여 딥러닝 손실함수에 가장 널리 쓰인다.
- 핵 노름 **(Nuclear Norm, $p = 1$):** $\|A\|_* := \sum_i \sigma_i$. L1 노름이 벡터의 희소성을 유도하듯, 핵 노름은 행렬의 '랭크'를 최소화하는 타이트한 볼록 이완(Convex relaxation)으로 작용하여 넷플릭스 추천 시스템 등 행렬 완성(Matrix completion) 알고리즘에 쓰인다.

25.2 극분해 (Polar Decomposition)

극분해는 복소수의 극형식 $z = re^{i\theta}$ ($r \ge 0$, $|e^{i\theta}| = 1$)를 행렬로 완벽하게 일반화한 것으로, 임의의 행렬을 형태를 보존하는 '회전(또는 반사)'과 축 방향의 '스케일링'으로 분해한다. SVD가 $A = U_S \Sigma V_S^T$일 때, 극분해는 $A = (U_S V_S^T)(V_S \Sigma V_S^T) = U_{polar} P$로 연결된다.

Theorem 25.1 (극분해 정리)**.** *임의의 정사각행렬 $A \in M_{n\times n}(\mathbb{R})$에 대하여 다음을 만족하는 직교행렬 U와 대칭 양의 준정부호 행렬 P가 존재한다:*

$$A = UP.$$

이때 P는 유일하며 $P = (A^T A)^{1/2}$이다. 만약 A가 가역행렬이면 U도 유일하고, P는 양의 정부호 대칭행렬이 된다.

Proof (가역행렬인 경우의 극분해 증명). ¡1¿ **Goal.** 가역행렬 A에 대하여 $A = UP$를 만족하는 유일한 직교행렬 U와 대칭 양의 정부호 행렬 P가 존재함을 보인다.

¡2¿ **Show.** 존재성

¡3¿ $A^T A$는 대칭행렬이고 A가 가역이므로, 임의의 $x \neq 0$에 대해 $x^T (A^T A)x = \|Ax\|^2 > 0$이다. 즉, $A^T A$는 양의 정부호이다.

¡4¿ 스펙트럼 정리에 의해 양의 정부호 대칭행렬은 유일한 양의 정부호 제곱근을 가진다. 이를 $P := (A^T A)^{1/2}$라 하자.

¡5¿ P 역시 양의 정부호이므로 가역이다. 행렬 U를 $U := AP^{-1}$로 정의한다.

¡6¿ **Claim.** U는 직교행렬이다. **Reason.** $U^TU = (P^{-1})^TA^TAP^{-1} = P^{-1}P^2P^{-1} = \mathbf{I}$이기 때문이다.

¡7¿ 따라서 $A = UP$로 분해됨을 보였다.

¡8¿ **Show.** 유일성

¡9¿ **Assume.** 또 다른 분해 $A = U_1P_1$이 존재한다고 가정하자 (U_1 직교, P_1 대칭 양의 정부호).

¡10¿ $A^TA = (U_1P_1)^T(U_1P_1) = P_1U_1^TU_1P_1 = P_1^2$이다.

¡11¿ 대칭 양의 정부호 행렬의 제곱근은 유일하므로 $P_1 = (A^TA)^{1/2} = P$이어야 한다.

¡12¿ $P_1 = P$이므로, $U_1P = UP$ 양변에 P^{-1}를 곱하면 $U_1 = U$이다.

¡13¿ **Conclusion.** 가역행렬의 극분해는 유일하게 존재한다.

□

26 행렬 분해 요약: LU, QR, Cholesky

특잇값 분해(SVD) 이외에도 수치 선형대수학에서 핵심적인 역할을 하는 분해법을 요약한다.

- **LU** 분해: $PA = LU$. 일반적인 정사각행렬을 가우스 소거법을 통해 하삼각행렬(L)과 상삼각행렬(U)로 분해한다. 선형방정식 풀이에 최적화되어 있다.
- **QR** 분해: $A = QR$. 그람-슈미츠 직교화를 행렬로 표현한 것으로, 직교행렬(Q)과 상삼각행렬(R)로 분해한다. 수치적으로 안정적인 최소제곱법 해법 및 고윳값 계산(QR 알고리즘)의 근간이다.
- **Cholesky** 분해: $A = LL^T$. 대칭 양의 정부호 행렬에만 적용 가능한 특수한 LU 분해로, 연산량이 절반이며 수치적으로 극히 안정적이다. 가우시안 프로세스 등 공분산 분해에 널리 쓰인다.

27 복소 벡터 공간과 유니터리 행렬

실수 행렬이라도 복소 고윳값을 가질 수 있으므로 복소 벡터 공간 $\mathbb{C}^n$으로의 확장이 대수학의 완성을 위해 필수적이다. 복소 내적에서는 켤레 대칭성($\langle x, y\rangle = \overline{\langle y, x\rangle}$)이 성립하며, 전치(Transpose) 대신 켤레 전치($A^* = \bar{A}^T$)를 사용한다.

- 에르미트 행렬 **(Hermitian matrix)**: $A^* = A$. 실수 대칭행렬의 확장이며, 양자역학의 관측가능량(Observable)을 대변한다. 고윳값은 항상 실수이다.
- 유니터리 행렬 **(Unitary matrix)**: $A^*A = \mathbf{I}$. 직교행렬의 복소 확장이며, 길이와 내적을 보존하는 등거리 변환이다. 양자 상태의 시간 진화를 묘사한다.
- 슈어 분해 **(Schur decomposition)**: 임의의 정사각 복소행렬 A에 대해, 유니터리 행렬 U가 존재하여 $U^*AU = T$(상삼각행렬)가 성립한다.

28 이차 형식과 쌍선형 형식

이차 형식(Quadratic form)은 대칭행렬 A에 의해 스칼라 함수 $Q(x) = x^TAx$로 표현된다. 주축 정리(Principal Axis Theorem)에 의해, 직교 좌표 변환 $x = Py$ (P는 직교행렬)를 통해 교차항(Cross terms)이 전혀 없는 표준형 $Q = \lambda_1y_1^2 + \cdots + \lambda_ny_n^2$로 변환할 수 있다. 실베스터의 관성 법칙(Sylvester's Law of Inertia)에 따라, 합동 변환(P^TAP)을 가해도 양의 고윳값, 음의 고윳값, 0인 고윳값의 개수(관성, Signature)는 절대 변하지 않는다.

29 쌍대 공간과 몫공간

쌍대 공간 **(Dual space):** 벡터 공간 $V/\mathbb{R}$에서 스칼라 $\mathbb{R}$로의 모든 선형 변환(선형 범함수, Linear functional)의 집합을 쌍대 공간 V^*라 한다. 선형 변환 $T : V \to W$의 행렬표현의 전치(Transpose) 행렬은, 근본적으로 쌍대 공간 사이의 사상 $T^t : W^* \to V^*$을 의미한다.

몫공간 **(Quotient space):** 부분공간 W에 대해, 평행이동된 잉여류(coset) $v + W$들의 집합을 몫공간 V/W라 한다. $\dim(V/W) = \dim(V) - \dim(W)$가 성립하며, 이는 현대 대수학의 제1 동형정리($V/\operatorname{Ker}(T) \cong R(T)$) 및 차원 정리의 근간이다.

30 텐서곱 기초

두 벡터 공간 V, W에 대해, 형식적인 기호 $v \otimes w$의 이중 선형결합으로 이루어진 텐서곱(Tensor product) 공간 $V \otimes W$를 정의한다. 행렬 관점에서는 크로네커 곱(Kronecker product)으로 표현된다. 양자역학의 다체계 얽힘(Entanglement)이나 딥러닝의 텐서 네트워크 이론에 필수적이다.

31 선형대수학의 응용 요약

- 주성분 분석 **(PCA)**: 데이터 공분산 행렬의 스펙트럼 분해를 통해 가장 큰 고윳값들에 대응하는 고유벡터로 데이터를 사영하여 최적의 차원 축소를 이룬다.
- 릿지 회귀 **(Ridge Regression)**: 역행렬이 존재하지 않거나 조건수가 매우 클 때, 대각선에 $\lambda \mathbf{I}$를 더해 정칙화(Regularization)하여 수치적 안정을 꾀한다.
- 페이지랭크 **(PageRank)**: 마르코프 체인 확률 행렬의 가장 큰 고윳값 1에 대응하는 고유벡터(정상 분포)를 구하여 네트워크 노드의 중요도를 판별한다.

32 행렬 함수와 행렬 미적분학 (The Grand Theory of Matrix Functions and Calculus)

이 절에서는 스칼라 함수 $f : \mathbb{R} \to \mathbb{R}$를 *행렬*로 확장하여 $f(A)$를 정의하고, 그 성질과 미분(섭동) 공식을 체계적으로 정리한다. 특히 본 강의록에서는 실수 대칭행렬 $A \in \mathrm{Sym}(n)$에 초점을 맞춘다. 이 경우 스펙트럼 정리로 인해 모든 이야기가 "고윳값에 f를 적용한 뒤 다시 합성한다"라는 한 줄로 정리되지만, *정의의 엄밀성(잘 정의됨)*과 *필수 정리들의 논리적 연결*을 놓치면 이후(로그, 제곱근, 단조성, 미분)에서 큰 혼란이 생긴다.

핵심 메시지**.** 대칭행렬 A는 직교대각화 $A = Q\Lambda Q^T$를 가지며, 적절한 함수 f에 대해

$$f(A) = Q\, f(\Lambda)\, Q^T, \qquad f(\Lambda) = \mathrm{diag}\big(f(\lambda_1), \dots, f(\lambda_n)\big)$$

로 정의한다. 이후의 모든 성질(대칭성, 양의정부호성, 노름/trace/det 공식, 프레셰 도함수)은 이 정의에서 나온다.

32.1 동기**:** 왜 $f(A)$가 필요한가**?**

행렬 함수는 다음과 같은 문제들을 *하나의 언어*로 통일한다.

- 선형 **ODE:** $\dot{x} = Ax$의 해는 $x(t) = e^{tA}x(0)$로 주어진다.
- **SPD(**대칭 양의정부호**)** 행렬의 기하**:** 공분산 행렬 $\Sigma \succ 0$의 "로그" $\log \Sigma$, "제곱근" $\Sigma^{1/2}$, "역제곱근" $\Sigma^{-1/2}$는 통계/최적화/정보기하의 기본 연산이다.
- 정칙화와 장벽함수**:** $\phi(X) = -\log\det(X)$는 $X \succ 0$ 영역의 대표적인 장벽(barrier)이며, 그 미분은 행렬 역함수와 직접 연결된다.

32.2 스펙트럼 정리 복습**:** 스펙트럼 분해와 스펙트럴 프로젝터

Theorem 32.1 (실수 대칭행렬의 스펙트럼 분해)**.** *$A \in \mathrm{Sym}(n)$이면 직교행렬 Q와 실수 대각행렬 $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$가 존재하여*

$$A = Q\Lambda Q^T$$

이다. 또한 서로 다른 고윳값들을 $\mu_1, \dots, \mu_k$라 하고 각 고유공간으로의 직교사영*을 $P_1, \dots, P_k$라 하면*

$$A = \sum_{j=1}^{k} \mu_j P_j, \qquad P_i P_j = \delta_{ij} P_i, \qquad \sum_{j=1}^{k} P_j = \mathbf{I}$$

가 성립한다.

Remark 32.1. *P_j는 "A의 고유공간으로의 직교사영"이므로* 기저 선택과 무관하게 *A로부터 유일하게 결정된다. 이 유일성이 곧 $f(A)$의* **well-definedness***의 핵심이다.*

32.3 다항식 함수 $p(A)$: 가장 안전한 출발점

Definition 32.1 (다항식의 행렬 대입). *다항식* $p(t) = \sum_{m=0}^{M} c_m t^m$*와 정사각행렬* A*에 대해*

$$p(A) := \sum_{m=0}^{M} c_m A^m$$

로 정의한다.

Proposition 32.2 (대칭행렬에 대한 다항식 함수의 스펙트럼 표현). $A \in \mathrm{Sym}(n)$*이고* $A = Q\Lambda Q^T$ *이면 임의의 다항식* p*에 대하여*

$$p(A) = Q\, p(\Lambda)\, Q^T, \qquad p(\Lambda) = \mathrm{diag}\big(p(\lambda_1), \ldots, p(\lambda_n)\big)$$

이다. 특히 $p(A) \in \mathrm{Sym}(n)$*이며,* $p(A)$*는* A*와 교환한다.*

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$ with $Q^T Q = \mathbf{I}$.

¡2¿ Goal. $p(A) = Qp(\Lambda)Q^T$를 보인다.

¡3¿ Λ가 대각행렬이므로 $\Lambda^m = \mathrm{diag}(\lambda_1^m, \ldots, \lambda_n^m)$이고 따라서 $p(\Lambda) = \sum_{m=0}^{M} c_m \Lambda^m$이다.

¡4¿ $A^m = (Q\Lambda Q^T)^m = Q\Lambda^m Q^T$가 모든 $m \geq 0$에 대해 성립한다. **Reason.** $Q^T Q = \mathbf{I}$이므로 곱을 전개하면 중간의 $Q^T Q$가 소거된다.

¡5¿ 따라서

$$p(A) = \sum_{m=0}^{M} c_m A^m = \sum_{m=0}^{M} c_m (Q\Lambda^m Q^T) = Q\Big(\sum_{m=0}^{M} c_m \Lambda^m\Big)Q^T = Qp(\Lambda)Q^T.$$

¡6¿ Conclusion. 원하는 식이 성립한다.

□

32.4 일반 함수 f에 대한 정의: 실수 대칭행렬의 함수해석

이제 f가 다항식이 아닐 때 $f(A)$를 정의한다. 대칭행렬에서는 스펙트럼 정리 덕분에 다음 정의가 가장 자연스럽다.

Definition 32.2 (스펙트럼(고윳값) 기반 정의). $A \in \mathrm{Sym}(n)$*의 서로 다른 고윳값을* $\mu_1, \ldots, \mu_k$*라 하고 대응하는 직교사영을* $P_1, \ldots, P_k$*라 하자. 어떤 함수* f*가* $\{\mu_1, \ldots, \mu_k\}$*에서 정의되어 있으면 다음으로* $f(A)$*를 정의한다:*

$$f(A) \ := \ \sum_{j=1}^{k} f(\mu_j)\, P_j.$$

Proposition 32.3 (well-definedness: 고유벡터 선택과 무관함). *위 정의로 주어진* $f(A)$*는* A*의 직교대각화* $A = Q\Lambda Q^T$ *에서* Q*의 선택(특히 중복 고윳값의 경우 고유공간 안에서의 기저 선택)에 의존하지 않는다.*

Proof. **¡1¿ Assume.** $A \in \mathrm{Sym}(n)$이고 서로 다른 고윳값이 $\mu_1, \ldots, \mu_k$이며 P_j는 각 고유공간으로의 직교사영이다.

¡2¿ Goal. $f(A) = \sum_j f(\mu_j)P_j$가 직교대각화의 선택과 무관하게 유일함을 보인다.

¡3¿ 각 P_j는 "μ_j-고유공간으로의 직교사영"이라는 기하학적 정의로부터 유일하다. **Reason.** 직교사영은 상공간과 영공간(또는 W와 $W^{\perp}$)이 주어지면 유일하게 결정된다.

¡4¿ 따라서 $\sum_j f(\mu_j)P_j$는 A에 의해 결정되는 P_j들과 스칼라 $f(\mu_j)$만으로 표현되며, 특정한 고유벡터(=특정 Q)를 선택할 여지가 없다.

¡5¿ Conclusion. $f(A)$는 잘 정의된다.

□

Remark 32.2 (대각화 표현과의 일치). $A = Q\Lambda Q^T$ *이고* Λ*의 대각원소가* $\lambda_1, \ldots, \lambda_n$*일 때, 위 정의는*

$$f(A) = Q\,\mathrm{diag}\big(f(\lambda_1), \ldots, f(\lambda_n)\big)Q^T$$

와 동치이다. (중복 고윳값은 같은 값이 반복해서 나타난다.)

32.5 동치인 다른 정의들: 멱급수와 다항식 근사

위 스펙트럼 정의는 대칭행렬에 대해 가장 강력하지만, 해석학적 관점에서 자주 쓰이는 두 정의와도 연결된다.

Proposition 32.4 (멱급수 정의의 정당화 (대칭행렬 버전))**.** *$f(t) = \sum_{m=0}^{\infty} c_m t^m$이 수렴반경 $R > 0$의 멱급수이고, $A \in \mathrm{Sym}(n)$이 $\|A\|_2 < R$를 만족하면 급수 $\sum_{m=0}^{\infty} c_m A^m$는 절대수렴하며 그 합은 스펙트럼 정의의 $f(A)$와 일치한다.*

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$이고 $\|A\|_2 = \max_i |\lambda_i| < R$.

¡2¿ Goal. $\sum_{m\ge 0} c_m A^m = Qf(\Lambda)Q^T$임을 보인다.

¡3¿ 앞의 명제로 $A^m = Q\Lambda^m Q^T$이므로 부분합 $S_M := \sum_{m=0}^{M} c_m A^m$는

$$S_M = Q\Big(\sum_{m=0}^{M} c_m \Lambda^m\Big)Q^T$$

이다.

¡4¿ 각 대각성분에 대해 $\sum_{m=0}^{\infty} c_m \lambda_i^m$가 절대수렴한다. **Reason.** $|\lambda_i| < R$이므로 스칼라 멱급수가 수렴한다.

¡5¿ 따라서 대각행렬 급수 $\sum_{m=0}^{\infty} c_m \Lambda^m$가 성분별로 수렴하며 그 합이 $f(\Lambda)$이다.

¡6¿ 직교행렬 Q는 노름을 보존하므로($\|QXQ^T\|_2 = \|X\|_2$) S_M은 수렴하며 극한은 $Qf(\Lambda)Q^T$이다.

¡7¿ Conclusion. 멱급수로 정의한 합은 스펙트럼 정의와 일치한다.

□

Remark 32.3 (연속함수의 다항식 근사)**.** *스펙트럼이 포함된 구간 $[a, b]$에서 f가 연속이면(예: $f = \sqrt{\cdot}$, $\log$ 등), 바이어슈트라스 근사정리에 의해 $[a, b]$에서 f를 균등하게 근사하는 다항식열 p_m이 존재한다. 이때 $f(A)$를 $p_m(A)$의 극한으로 정의해도 스펙트럼 정의와 일치한다.*

32.6 기본 성질: 대칭성, 유사불변성, 스펙트럼 매핑

Theorem 32.5 (기본 성질)**.** *$A \in \mathrm{Sym}(n)$이고 f, g가 $\mathrm{spec}(A)$에서 정의되어 있다고 하자.*

(a) $f(A) \in \mathrm{Sym}(n)$.

(b) $f(A)$는 A와 교환한다: $Af(A) = f(A)A$.

(c) (함수 대수 성질) $(f + g)(A) = f(A) + g(A)$, $(fg)(A) = f(A)g(A)$.

(d) (직교 유사불변성) 임의의 직교행렬 U에 대해 $f(UAU^T) = Uf(A)U^T$.

(e) (스펙트럼 매핑) A의 고유벡터 v가 고윳값 λ를 가지면, v는 $f(A)$의 고유벡터이고 고윳값은 $f(\lambda)$이다. 따라서 $\mathrm{spec}(f(A)) = \{f(\lambda) : \lambda \in \mathrm{spec}(A)\}$.

Proof. **¡1¿ Assume.** $A = \sum_{j=1}^{k} \mu_j P_j$는 스펙트럼 분해이며 $f(A) = \sum_{j=1}^{k} f(\mu_j)P_j$로 정의한다.

¡2¿ Show. (a) P_j는 대칭 행렬이므로($P_j^T = P_j$) 선형결합인 $f(A)$도 대칭이다.

¡3¿ Show. (b) $AP_j = \mu_j P_j$이고 $P_j A = \mu_j P_j$이므로

$$Af(A) = \sum_j f(\mu_j)AP_j = \sum_j f(\mu_j)\mu_j P_j = \sum_j f(\mu_j)P_j A = f(A)A.$$

¡4¿ Show. (c) $P_i P_j = \delta_{ij} P_j$를 사용하면

$$f(A)g(A) = \sum_{i,j} f(\mu_i)g(\mu_j)P_i P_j = \sum_j f(\mu_j)g(\mu_j)P_j = (fg)(A)$$

이고, 덧셈은 자명하다.

¡5¿ Show. (d) $UAU^T = \sum_j \mu_j (UP_j U^T)$이며 $UP_j U^T$는 μ_j-고유공간으로의 직교사영이므로 $f(UAU^T) = \sum_j f(\mu_j)UP_j U^T = Uf(A)U^T$이다.

¡6¿ Show. (e) $Av = \lambda v$이면 v는 λ-고유공간에 속하므로 대응하는 프로젝터 P_λ에 대해 $P_\lambda v = v$이고 다른 고윳값 프로젝터는 0이 된다. 따라서 $f(A)v = \sum_j f(\mu_j)P_j v = f(\lambda)v$이다.

¡7¿ Conclusion. 모든 항이 증명되었다.

□

32.7 노름, trace, det: 고윳값으로 환원되는 공식들

대칭행렬 함수의 많은 양은 고윳값으로 완전히 환원된다.

Proposition 32.6 (스펙트럼 노름). *$A \in \mathrm{Sym}(n)$이면*

$$\|f(A)\|_2 = \max_{\lambda \in \mathrm{spec}(A)} |f(\lambda)|.$$

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$.

¡2¿ $f(A) = Qf(\Lambda)Q^T$이고, 직교유사변환은 2-노름을 보존하므로 $\|f(A)\|_2 = \|f(\Lambda)\|_2$이다.

¡3¿ 대각행렬의 2-노름은 대각성분의 절댓값 최댓값이므로 $\|f(\Lambda)\|_2 = \max_i |f(\lambda_i)|$이다.

¡4¿ Conclusion. 결론이 성립한다.

□

Proposition 32.7 (trace와 determinant). *$A \in \mathrm{Sym}(n)$이고 $A = Q\Lambda Q^T$ 이면*

$$\mathrm{tr}\big(f(A)\big) = \sum_{i=1}^{n} f(\lambda_i).$$

또한 $f(\lambda_i)$들이 모두 정의되어 있을 때

$$\det\big(f(A)\big) = \prod_{i=1}^{n} f(\lambda_i).$$

Proof. **¡1¿** $f(A) = Qf(\Lambda)Q^T$이고 trace와 det는 유사변환 불변량이므로 $\mathrm{tr}(f(A)) = \mathrm{tr}(f(\Lambda))$, $\det(f(A)) = \det(f(\Lambda))$이다.

¡2¿ 대각행렬에 대해 $\mathrm{tr}(f(\Lambda)) = \sum_i f(\lambda_i)$, $\det(f(\Lambda)) = \prod_i f(\lambda_i)$가 자명하다.

¡3¿ Conclusion. 원하는 식이 성립한다.

□

32.8 대표 예시 I: 행렬 지수 exp(A)

Definition 32.3 (행렬 지수 함수). *임의의 정사각행렬 A에 대해*

$$\exp(A) := e^A := \sum_{m=0}^{\infty} \frac{A^m}{m!}$$

로 정의한다.

Proposition 32.8 (대칭행렬의 지수는 SPD). *$A \in \mathrm{Sym}(n)$이면 e^A는 대칭* 양의정부호 *행렬이며, 고윳값은 e^{λ_i}이다.*

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$.

¡2¿ Claim. $e^A = Qe^\Lambda Q^T$이며 $e^\Lambda = \mathrm{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})$이다. **Reason.** 멱급수 정의와 앞의 "멱급수 정당화" 명제를 적용한다.

¡3¿ e^Λ의 대각성분은 모두 > 0이므로 $e^\Lambda \succ 0$이고, 직교유사변환은 양의정부호성을 보존하므로 $e^A \succ 0$이다.

¡4¿ Conclusion. e^A는 SPD이고 고윳값은 e^{λ_i}이다.

□

Proposition 32.9 (교환하는 행렬의 지수법칙). *$A, B \in \mathbb{R}^{n\times n}$이 $AB = BA$를 만족하면 $e^{A+B} = e^A e^B$이다.*

Remark 32.4. *교환하지 않으면 일반적으로 $e^{A+B} \neq e^A e^B$이며, 이 오차는 리 대수의 BCH(Baker–Campbell–Hausdorff) 전개로 측정된다.*

Proposition 32.10 (대칭행렬 버전의 야코비 공식). *$A \in \mathrm{Sym}(n)$이면*

$$\det(e^A) = e^{\mathrm{tr}(A)}.$$

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$.

¡2¿ 위의 trace/det 공식에 의해 $\det(e^A) = \prod_i e^{\lambda_i} = e^{\sum_i \lambda_i}$이다.

¡3¿ 또한 $\mathrm{tr}(A) = \mathrm{tr}(\Lambda) = \sum_i \lambda_i$이다.

¡4¿ Conclusion. 따라서 $\det(e^A) = e^{\mathrm{tr}(A)}$이다.

□

32.9 대표 예시 II: 제곱근, 절댓값, 부호 (대칭행렬)

Definition 32.4 (PSD 제곱근). *$A \in \mathrm{Sym}(n)$이 $A \succeq 0$라 하자. 스펙트럼 분해 $A = Q\Lambda Q^T$에서 $\Lambda = \mathrm{diag}(\lambda_i)$이고 $\lambda_i \geq 0$이므로*

$$A^{1/2} := Q\,\mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})Q^T$$

로 정의한다.

Theorem 32.11 (PSD 제곱근의 존재와 유일성). *$A \succeq 0$인 대칭행렬에 대해*

(a) $A^{1/2} \succeq 0$이며 $(A^{1/2})^2 = A$이다.

(b) $B \succeq 0$이고 $B^2 = A$이면 $B = A^{1/2}$이다. (유일성)

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$ with $\Lambda = \mathrm{diag}(\lambda_i)$, $\lambda_i \geq 0$.

¡2¿ Show. (a) 정의에 의해 $A^{1/2} = Q\,\mathrm{diag}(\sqrt{\lambda_i})Q^T$이고 $(A^{1/2})^2 = Q\,\mathrm{diag}(\lambda_i)Q^T = A$이다. 또한 대각성분이 ≥ 0이므로 $A^{1/2} \succeq 0$이다.

¡3¿ Show. (b) **Assume.** $B \succeq 0$이고 $B^2 = A$.

¡4¿ 그러면 $AB = BA$이다. **Reason.** $AB = B^2B = B^3 = BB^2 = BA$.

¡5¿ A, B는 둘 다 대칭이고 교환하므로 동시에 직교대각화 가능하다. 즉, 어떤 직교행렬 U에 대해 $U^TAU = \mathrm{diag}(\lambda_i)$, $U^TBU = \mathrm{diag}(\mu_i)$ (모두 실수)로 쓸 수 있다.

¡6¿ $B^2 = A$를 대각성분으로 비교하면 $\mu_i^2 = \lambda_i$이고, $B \succeq 0$이므로 $\mu_i = \sqrt{\lambda_i}$이다.

¡7¿ 따라서 $B = U\,\mathrm{diag}(\sqrt{\lambda_i})U^T = A^{1/2}$이다.

¡8¿ Conclusion. 유일성이 성립한다.

□

Definition 32.5 (절댓값과 부호). *$A \in \mathrm{Sym}(n)$에 대해*

$$|A| := \left(A^2\right)^{1/2}, \qquad \mathrm{sign}(A) := Q\ \mathrm{diag}(\mathrm{sign}(\lambda_1), \dots, \mathrm{sign}(\lambda_n))\,Q^T$$

로 정의한다. 여기서 $\mathrm{sign}(0) := 0$으로 둔다.

Remark 32.5. *A가 가역이면 $\mathrm{sign}(A)$는 직교행렬이며(고윳값이 ± 1), 분해 $A = \mathrm{sign}(A)\,|A|$는 대칭행렬의 "polar-like" 분해로 볼 수 있다.*

32.10 대표 예시 III: 로그와 분수 거듭제곱 (SPD)

이제부터는 $A \succ 0$ (대칭 양의정부호)인 경우를 다룬다. 이때 $\lambda_i > 0$이므로 $\log \lambda_i$, λ_i^α가 잘 정의된다.

Definition 32.6 (로그와 거듭제곱). *$A \in \mathrm{Sym}(n)$이 $A \succ 0$이고 $A = Q\Lambda Q^T$라 하자.*

$$\log A := Q\,\mathrm{diag}(\log \lambda_1, \dots, \log \lambda_n)Q^T, \qquad A^\alpha := Q\,\mathrm{diag}(\lambda_1^\alpha, \dots, \lambda_n^\alpha)Q^T \quad (\alpha \in \mathbb{R})$$

로 정의한다.

Proposition 32.12 (역관계: $\exp(\log A) = A$). *$A \succ 0$이면 $\exp(\log A) = A$이고, $S \in \mathrm{Sym}(n)$이면 $\log(\exp S) = S$이다.*

Proof. **¡1¿ Assume.** $A = Q\Lambda Q^T$ with $\lambda_i > 0$.

¡2¿ 정의에 의해 $\log A = Q\,\mathrm{diag}(\log \lambda_i)Q^T$이므로 $\exp(\log A) = Q\,\mathrm{diag}(e^{\log \lambda_i})Q^T = Q\,\mathrm{diag}(\lambda_i)Q^T = A$이다.

¡3¿ 또한 $S = Q\,\mathrm{diag}(\sigma_i)Q^T$이면 $\exp S = Q\,\mathrm{diag}(e^{\sigma_i})Q^T \succ 0$이고 $\log(\exp S) = Q\,\mathrm{diag}(\log(e^{\sigma_i}))Q^T = S$이다.

¡4¿ Conclusion. 두 항등식이 성립한다.

□

32.11 Loewner 순서와 단조성: 언제 $A \preceq B \Rightarrow f(A) \preceq f(B)$인가?

Definition 32.7 (Loewner(PSD) 순서)**.** *$A, B \in \mathrm{Sym}(n)$에 대해 $A \preceq B$는 $B - A \succeq 0$을 의미한다.*

Theorem 32.13 (교환하는 경우: 스칼라 단조성으로 충분)**.** *$A, B \in \mathrm{Sym}(n)$이 $AB = BA$를 만족한다고 하자. $A \preceq B$이고 $f : \mathbb{R} \to \mathbb{R}$가 어떤 구간 $I \supset \mathrm{spec}(A) \cup \mathrm{spec}(B)$에서 단조증가이면*

$$f(A) \preceq f(B)$$

가 성립한다.

Proof. **¡1¿ Assume.** $A, B \in \mathrm{Sym}(n)$, $AB = BA$.

¡2¿ Claim. A, B는 동시에 직교대각화 가능하다. **Reason.** 대칭행렬의 각 고유공간은 자신과 교환하는 대칭행렬에 의해 불변이므로, A의 고유공간들 위에서 B를 다시 직교대각화하면 된다.

¡3¿ 따라서 어떤 직교행렬 Q가 존재하여 $A = Q\,\mathrm{diag}(a_i)Q^T$, $B = Q\,\mathrm{diag}(b_i)Q^T$로 쓸 수 있다.

¡4¿ $A \preceq B$이면 $B - A = Q\,\mathrm{diag}(b_i - a_i)Q^T \succeq 0$이므로 모든 i에 대해 $b_i - a_i \geq 0$, 즉 $a_i \leq b_i$이다.

¡5¿ f가 단조증가이므로 $f(a_i) \leq f(b_i)$이고 따라서 $\mathrm{diag}(f(b_i) - f(a_i)) \succeq 0$이다.

¡6¿ 그러므로 $f(B) - f(A) = Q\,\mathrm{diag}(f(b_i) - f(a_i))Q^T \succeq 0$, 즉 $f(A) \preceq f(B)$.

¡7¿ Conclusion. 증명되었다.

□

Remark 32.6 (비가환의 난점과 operator monotone)**.** *A, B가 교환하지 않으면 스칼라 단조증가만으로는 $f(A) \preceq f(B)$가 보장되지 않는다. 이때 필요한 개념이* **operator monotone(=matrix monotone)** *함수이며, 대표적으로*

$$f(t) = t^{\alpha}\ (0 \leq \alpha \leq 1), \qquad f(t) = \log t$$

는 $(0, \infty)$에서 operator monotone이다(뢰브너–하인츠 부등식, Löwner theory). 이 부분은 함수해석학/행렬해석학의 깊은 주제이므로 여기서는 결과를 소개하는 데 그친다.

32.12 행렬 미적분: 프레셰 도함수와 분할차분 공식 (대칭행렬)

최적화/통계/역문제에서는 $f(A)$ 자체보다도 *섭동 E에 대한 1차 변화량*이 중요하다.

Definition 32.8 (프레셰(Fréchet) 도함수)**.** *함수 f의 A에서의 프레셰 도함수는*

$$f(A + E) = f(A) + L_f(A)[E] + o(\|E\|)$$

를 만족하는 선형 사상 $L_f(A) : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$이다.

Proposition 32.14 (다항식의 프레셰 도함수)**.** *$p(t) = \sum_{m=0}^{M} c_m t^m$이면*

$$L_p(A)[E] = \sum_{m=1}^{M} c_m \sum_{j=0}^{m-1} A^j E A^{m-1-j}.$$

Proof. **¡1¿ Goal.** $p(A + E) - p(A)$의 E에 대한 1차항을 계산한다.

¡2¿ $(A + E)^m = A^m + \sum_{j=0}^{m-1} A^j E A^{m-1-j} + O(\|E\|^2)$. **Reason.** 곱을 전개하면 E가 정확히 한 번 등장하는 항들의 합이 위의 합이고, E가 두 번 이상 등장하는 항들은 $O(\|E\|^2)$이다.

¡3¿ 따라서 $p(A + E) - p(A) = \sum_{m=1}^{M} c_m \sum_{j=0}^{m-1} A^j E A^{m-1-j} + O(\|E\|^2)$이다.

¡4¿ Conclusion. 정의에 의해 선형 1차항이 $L_p(A)[E]$이다.

□

Theorem 32.15 (대칭행렬에서의 분할차분(divided differences) 공식)**.** *$A \in \mathrm{Sym}(n)$이고 $A = Q\Lambda Q^T$라 하자. f가 $\mathrm{spec}(A)$를 포함하는 구간에서 미분 가능하면, 임의의 섭동 E에 대해*

$$L_f(A)[E] = Q\Big(f^{[1]}(\Lambda)\ \circ\ (Q^T E Q)\Big)Q^T$$

이다. 여기서 $\circ$는 아다마르(Hadamard) 곱이고, $f^{[1]}(\Lambda) \in \mathbb{R}^{n\times n}$는

$$\big(f^{[1]}(\Lambda)\big)_{ij} := \begin{cases} \dfrac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}, & i \neq j, \\ f'(\lambda_i), & i = j \end{cases}$$

로 정의된다.

Remark 32.7. *이 공식은 "고유값에만 의존하는 스칼라 함수" $\Phi(A) = \operatorname{tr} f(A) = \sum_i f(\lambda_i)$의 미분을 특히 단순화한다: 임의의 대칭 E에 대해*

$$D\Phi(A)[E] = \operatorname{tr}\big(f'(A)\,E\big).$$

예를 들어 $A \succ 0$에서 $\Phi(A) = \log\det(A) = \operatorname{tr}(\log A)$이면 $D\Phi(A)[E] = \operatorname{tr}(A^{-1}E)$가 된다.

32.13 수치적 계산: 고유분해 vs 반복법/근사

대칭행렬은 수치적으로 가장 다루기 좋은 클래스이다.

- **직접 계산(고유분해):** $A = Q\Lambda Q^T$를 구한 뒤 $f(A) = Qf(\Lambda)Q^T$. 대칭행렬의 고유분해는 일반 행렬보다 안정적이다.
- **대규모 희소/저랭크:** Lanczos/Krylov 부분공간으로 극단 고윳값이나 $f(A)v$만을 근사한다. (예: $e^A v$, $\log(A)v$)
- **반복법(예: 제곱근/역제곱근):** SPD A의 제곱근은 뉴턴 반복 $X_{k+1} = \frac{1}{2}(X_k + AX_k^{-1})$로 구할 수 있고, 적절한 스케일링 하에서 역제곱근은 Newton–Schulz 형태 $X_{k+1} = \frac{1}{2}X_k(3\mathbf{I} - AX_k^2)$로 근사할 수 있다. (초기값과 스케일링 조건이 수렴을 좌우한다.)

33 가역행렬의 동치 조건 총정리 (The Invertible Matrix Theorem)

긴 선형대수학 여정의 대미로, 정사각행렬 $A \in M_{n\times n}(\mathbb{R})$가 "가역(Invertible)이다" 라는 단 한 줄의 명제가 내포하는 16가지의 위대한 동치 조건들을 총정리한다.

Theorem 33.1. *다음 명제들은 모두 상호 논리적 동치(Logically equivalent)이다:*

1. *A는 가역행렬이다.*
2. *역행렬 A^{-1}가 존재한다.*
3. *행렬식이 0이 아니다 ($\det(A) \neq 0$).*
4. *풀 랭크 행렬이다 ($\operatorname{rank}(A) = n$).*
5. *A의 열벡터(또는 행벡터)들이 선형독립이다.*
6. *A의 열공간(또는 행공간)이 전체 벡터 공간 $\mathbb{R}^n$과 같다.*
7. *동차 선형방정식 $Ax = 0$은 오직 자명해 $x = 0$만을 갖는다. ($\operatorname{Ker}(A) = \{0\}$)*
8. *임의의 벡터 $b \in \mathbb{R}^n$에 대해 $Ax = b$는 유일한 해를 갖는다.*
9. *가우스 소거법을 통한 A의 기약행사다리꼴(RREF)은 단위행렬 $\mathbf{I}_n$이다.*
10. *A는 유한 개의 기본 행렬(Elementary matrix)들의 곱으로 표현된다.*
11. *선형 변환 $T(x) = Ax$는 단사(Injective) 함수이다.*
12. *선형 변환 $T(x) = Ax$는 전사(Surjective) 함수이다.*
13. *A의 고윳값 중 0인 것이 존재하지 않는다.*
14. *A의 모든 특잇값(Singular value)이 엄밀히 양수이다 ($\sigma_i > 0$).*
15. *A^TA 및 AA^T는 대칭 양의 정부호(Positive definite) 행렬이다.*
16. *A의 극분해 $A = UP$에서 고유한 성분 P가 대칭 양의 정부호 행렬이다.*

— The End —