Linear Algebra for Statistics

Chapter 14

Instructor: Seokho Lee (이석호)

Hankuk University of Foreign Studies

Chapter 14 스펙트럼 분해

14.1 스펙트럼 분해(Spectral Decomposition)

한개의 행렬을 계수가 1인 행렬들의 합으로 분해하는 방법을 스펙트럼 분해라고 한다.

Figure 14.1: 힐베르트

Theorem (정리 14.1 스펙트럼 분해)

행렬 $A_{n\times n}$ 이 대칭행렬이고, 고유값이

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \tag{14.1}$$

일 때, 직교정규인 고유벡터를 열행렬로 하는 행렬

$$P = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \tag{14.2}$$

와 고유값으로 이루어진 대각행렬을

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n) \tag{14.3}$$

라 하면 행렬 A 는 다음과 같이 분해된다.

$$A = P\Lambda P^T = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T . \tag{14.4}$$

Proof.

행렬 A 가 대칭행렬이므로 직교정규인 고유벡터가 존재한다. 따라서 고유벡터를 열벡터로 하는 행렬

$$P = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$$

은 직교행렬이므로 A 를 대각화하면 $A = P\Lambda P^T$ 가 된다. 여기서

$$\begin{aligned}
P\Lambda P^T &= (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} \\
&= (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \begin{pmatrix} \lambda_1 \mathbf{v}_1^T \\ \lambda_2 \mathbf{v}_2^T \\ \vdots \\ \lambda_n \mathbf{v}_n^T \end{pmatrix} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \qquad (14.5)
\end{aligned}$$

임을 알 수 있다. □

Example

행렬 A 의 고유값이 $2, 5, -5$ 이고 고유벡터가 아래와 같을때, A 를 스펙트럼 분해하라.

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & -4 \end{pmatrix}, \quad \mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix}.$$

(풀이) 위의 고유벡터는 직교함을 알 수 있다. 하지만 길이가 1이 아니므로 이를 정규화하면

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \end{pmatrix}$$

이 된다. 따라서 행렬 A 는 다음과 같이 분해된다.

Example (continue)

$$
\begin{aligned}
A &= \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T \\
&= 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (1,0,0) + 5 \begin{pmatrix} 0 \\ \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{pmatrix} (0, \tfrac{3}{\sqrt{10}}, \tfrac{1}{\sqrt{10}}) + (-5) \begin{pmatrix} 0 \\ \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \end{pmatrix} (0, \tfrac{1}{\sqrt{10}}, -\tfrac{3}{\sqrt{10}}).
\end{aligned}
$$

즉,

$$
A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & -4 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4.5 & 1.5 \\ 0 & 1.5 & 0.5 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & -0.5 & 1.5 \\ 0 & 1.5 & -4.5 \end{pmatrix}.
$$

Example

다음 행렬 A 를 스펙트럼 분해하여 계수가 1인 3개의 행렬의 합으로 표현하라.

$$A = \begin{pmatrix} 2 & \sqrt{7} & 0 \\ \sqrt{7} & 8 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14.6}$$

14.2 특이값 분해 (Singular Value Decomposition)

특이값 분해는 스펙트럼 분해와는 달리 정방행렬이 아닌 경우에도 적용가능한 분해법으로 행렬근사(matrix approximation)나 최소제곱(least squares)법에 사용된다.

Theorem (정리 14.2 특이값 분해; SVD)

행렬 $A_{m\times n}$ $(m \geq n = Rank(A))$ 에 대하여 다음과 같은 분해를 특이값 분해(singular value decomposition; SVD)라고 한다.

$$A_{m\times n} = U_{m\times n} D_{n\times n} V^T_{n\times n}. \tag{14.7}$$

여기서, $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n), V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 이고

(1) $U^T U = V^T V = I_{n\times n}$

(2) $AA^T \mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad i = 1, 2, \ldots, n$

(3) $A^T A \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, 2, \ldots, n$

(4) *D 는 아래와 같이 정의된다* $(\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_n} > 0)$

$$D = diag(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$$

Proof.

이를 확인하기 위해 $A = UDV^T$ 라고 하면,

$$AA^T = UDV^TVDU^T = UD^2U^T = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}\begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix}$$

가 되어 행렬 AA^T 의 스펙트럼 분해와 일치한다. 따라서, λ_i 는 AA^T 의 고유값이고 $\mathbf{u}_i$ 는 직교하는 고유벡터이다. 같은 방법으로

$$A^TA = VDU^TUDV^T = VD^2V^T = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}\begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix}$$

이므로, λ_i 는 A^TA 의 고유값이기도 하며 $\mathbf{v}_i$ 는 직교하는 고유벡터이다. □

Remark 만일 λ_i 가 $A^T A$ (혹은 AA^T) 의 고유값이면 $\sqrt{\lambda_i}$ 는 행렬 A 의 특이값이다.

Remark 행렬 U 의 열벡터 $\mathbf{u}_i$ 를 i번째 왼쪽특이벡터(left singular vector), 행렬 V 의 열벡터 $\mathbf{v}_i$ 를 i번째 오른쪽특이벡터(right singular vector) 라고 한다.

Remark 정리 14.2에서 $\text{Rank}(A) = m \leq n$ 인 경우는

$$A_{m\times n} = U_{m\times m} D_{m\times m} V^T_{m\times n} \tag{14.8}$$

이 된다. 만일 행렬 $A_{m\times n}$ 의 계수가 $\text{Rank}(A) = r \leq \min(m, n)$ 인 경우는

$$A_{m\times n} = U_{m\times r} D_{r\times r} V^T_{r\times n}$$

로 표현된다. 어떠한 경우라도 정리 14.2의 성질을 만족한다.

Example

아래 행렬 A 에 대해 특이값분해를 실행하라.

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{14.9}$$

(풀이) 행렬 AA^T 의 고유값과 고유벡터를 구하면

$$AA^T = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$$

이므로 고유값은 $\lambda_1 = 5$, $\lambda_2 = 1$ 이고 고유벡터는 $\mathbf{u}_1 = (1, 0)^T$, $\mathbf{u}_2 = (0, 1)^T$ 이다. 따라서

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

이다.

Example (continue)

(풀이) 행렬 A^TA 의 고유값과 직교정규고유벡터를 구하면

$$A^TA = \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

이므로 고유값은 $0, 1, 5$ 임을 계산할 수 있다. $1, 5$ 에 대응하는 직교정규고유벡터를 구하면 $\mathbf{v}_1 = (2/\sqrt{5}, 0, -1/\sqrt{5})^T$, $\mathbf{v}_2 = (0, 1, 0)^T$ 임을 알 수 있으므로

$$V = \begin{pmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \\ -\frac{1}{\sqrt{5}} & 0 \end{pmatrix}$$

이다. 따라서 특이값분해는 아래와 같다.

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{5} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \\ -\frac{1}{\sqrt{5}} & 0 \end{pmatrix}^T = UDV^T.$$

Proposition (14.3 SVD의 전개)

행렬 $A_{m\times n}$ ($Rank(A) = m \le n$) 에 대하여, 특이값 분해를 하면 행렬 A 는 아래와 같이 m 개의 $m \times n$ 행렬의 합으로 표현 가능하다.

$$A = U_{m\times m} D_{m\times m} V_{n\times m}^T = \sum_{i=1}^{m} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T. \tag{14.10}$$

Example

Proposition 14.3 을 이용하여 식 (14.9) 의 행렬 A 에 대해 SVD 를 수행하면 다음과 같다.

$$\begin{aligned} A &= \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = U_{2\times 2} D_{2\times 2} V_{3\times 2}^T = \sum_{i=1}^{2} \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T \\ &= \sqrt{5} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left(\tfrac{2}{\sqrt{5}}, 0, -\tfrac{1}{\sqrt{5}} \right) + \sqrt{1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0, 1, 0) \\ &= \begin{pmatrix} 2 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

Remark 행렬 $A_{m\times n}$ $(\text{Rank}(A) = n \le m)$ 에 때하여, 특이값 분해를 하면 행렬 A 는 아래와 같이 n 개의 $m \times n$ 행렬의 합으로 표현된다.

$$A = U_{m\times n} D_{n\times n} V_{n\times n}^T = \sum_{i=1}^{n} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T. \tag{14.11}$$

Example

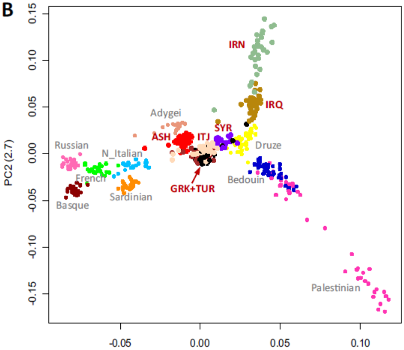
식 (14.11)을 사용하여 행렬 A 를 두 개의 행렬의 합으로 표시하라.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \tag{14.12}$$

14.3 행렬 분해의 활용

- 행렬의 계수 계산
- 역행렬이 존재하지 않는 행렬에 대한 일반화역행렬(generalized inverse)의 계산
- 주성분분석(principal component analysis) 등 고차원자료에 대한 차원축소(dimension reduction)
- 행렬의 근사(matrix approximation)
- 음성자료 및 영상 이미지 압축

Figure 14.2: 영상이미지 압축의 예



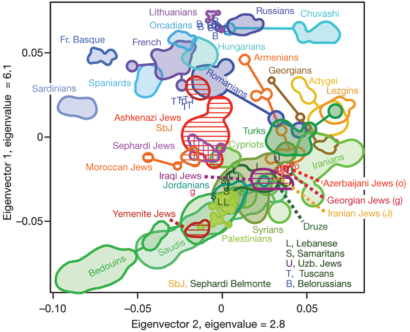


Figure 2 | PCA of west Eurasian high-density array data. Plot of kernel densities (Supplementary Note 2) for each population sample ($n > 10$) was estimated on the basis of PC1 and PC2 coordinates in Supplementary Fig. 3. Individuals from these samples were plotted by using PC1 and PC2 coordinates and were overlaid with the plot of kernel density.

Figure 14.3: 주성분분석을 통한 차원축소의 예