The Korean Journal of Applied Statistics
2023, Vol. 36, No. 3, 271–277

DOI: https://doi.org/10.5351/KJAS.2023.36.3.271
Print ISSN1225-066X / Online ISSN 2383-5818

A summary on the distance metric between two subspaces

Kipoong Kim[a], Choongrak Kim[1,b]

[a]Department of Statistics, Seoul National University; [b]Department of Statistics, Pusan National University

Abstract

In this paper, we introduced distance metric between two subspaces. For this, several matrix norms such as the spectral norm and the Frobenius norm are introduced. Further, the distance between two matrices based on the projection and principal angles are introduced. Finally, its application to the matrix perturbation theory with the famous Davis-Kahan theorem (1970) is illlustrated.

Keywords: eigenspace, Frobenius norm, perturbation theory, singular subspace, spectral norm

1. 서론

흔히 자료는 잡음(noise), 측정오차(measurement error) 등의 원인으로 교란(perturbation)되어 나타난다. p개의 변수를 n번 관측하여 $n \times p$ 행렬을 $\boldsymbol{Y}$라고 하면 이 행렬은 $\boldsymbol{Y} = \boldsymbol{B} + \boldsymbol{E}$로 표현할 수 있는데 여기서 $\boldsymbol{B}$는 알려져 있지 않은 참 행렬(true matrix), $\boldsymbol{E}$는 교란행렬(perturbation matrix)이다. 교란된 관측치를 이용하여 계산된 추정치는 교란의 크기에 의해 커다란 영향을 받게 된다. 이를 위해 관측행렬과 참 행렬간의 거리에 대한 측도가 필요하며, 이는 각 행렬이 속한 고유공간(eigenspaces) 또는 비정칙치 부공간(singular subspaces) 간의 거리측도를 통해 이루어진다. 공간의 거리측도는 통계학 분야에서도 많이 활용되고 있다. 가장 대표적인 예로 스펙트럴 군집화(spectral clustering)에서는 라플라스 행렬(Laplacian matrix)의 k개의 가장 작은 고유치에 해당되는 고유벡터를 이용하는데 이 경우 k를 정하는 방법으로 두 공간의 거리에 대한 Davis-Kahan (1970)의 정리를 이용한다. 이외에도 뾰족분포(spiky distribution)와 관련된 여러 추정문제에도 활용되고 있으며, 최근 인공지능분야의 후방탐색(background detection), 안면인식(face recognition)의 핵심이론으로 사용되는 강건주성분분석(robust principal component analysis) (Candes 등, 2011; Chandrasekaran 등, 2011)에도 교란이론이 중요하게 사용된다. 본 논문에서는 두 공간 간의 거리측도에 대한 Chen 등 (2021)의 연구결과를 소개하고 이를 교란이론에 적용하는 간단한 예를 살펴본 후 교란이론의 가장 핵심적 결과인 Davis-Kahan (1970)의 정리를 소개한다.

2. 행렬이론

2.1. 기호와 정의

이 절에서는 행렬에 대한 몇 가지 기호 및 행렬의 여러 가지 크기를 소개한다. $\boldsymbol{A}$는 $a_{ij}, i = 1, \ldots, m; j = 1, \ldots, n$를 ij 번째 원소로 갖는 $m \times n$ 행렬이라 하자. $\boldsymbol{A}$에 대하여 $\sigma_j(\boldsymbol{A})$는 j 번째 큰 비정칙치(singular value)를 나타내고 특히 $\sigma_{\max}(\boldsymbol{A})$와 $\sigma_{\min}(\boldsymbol{A})$는 각각 가장 큰 비정칙치와 가장 작은 비정칙치를 나타낸다고 하자. 또한, 정방행렬

This work was supported by a 2-year Research Grant of Pusan National University.

[1] Corresponding author: Department of Statistics, Pusan National University, Busandaehak-ro, 63 beon-gil 2, Gumjeong-gu, Busan, 46288, Korea. E-mail: crkim@pusan.ac.kr

Published 30 June 2023 / journal homepage: http://kjas.or.kr

$(m = n)$의 경우 $\lambda_j(\boldsymbol{A})$는 j 번째 큰 고유치(eigenvalue)를 나타내고 특히 $\lambda_{\max}(\boldsymbol{A})$와 $\lambda_{\min}(\boldsymbol{A})$는 각각 가장 큰 고유치와 가장 작은 고유치를 나타낸다고 하자. 정방행렬의 대각원소 합을 나타내는 트레이스(trace)는 $tr(\boldsymbol{A})$로 나타낸다. 다음으로 행렬의 크기(노름, norm)에 대해 소개한다. $m \times n$ 행렬 $\boldsymbol{A}$에 대하여 $\|\boldsymbol{A}\|$는 스펙트럴 노름, 즉 가장 큰 비정칙치 $\sigma_{\max}(\boldsymbol{A})$를 나타내고, $\|\boldsymbol{A}\|_*$는 nuclear 노름, 즉 비정칙치들의 합을 나타내며, $\|\boldsymbol{A}\|_{\mathrm{F}}$는 Frobenius 노름, 즉 $\|\boldsymbol{A}\|_{\mathrm{F}} := \sqrt{\sum_i \sum_j a_{ij}^2} = \sqrt{tr(\boldsymbol{A}^T\boldsymbol{A})}$를 나타낸다.

2.2. 행렬의 기본 이론 및 성질

행렬의 기본 이론 및 성질로서 크기불변노름(unitarily invariant norm)과 고유치 및 비정칙치의 교란 상한값(perturbation bounds for eigenvalues and singular values)에 대해 소개한다.

Definition 1. *$m \times n$ 행렬 $\boldsymbol{A}$에 대해 행렬 노름 $|||\cdot|||$이*

$$|||\boldsymbol{A}||| = \left|\left|\left|\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{V}\right|\right|\right|$$

을 만족하면 크기불변(unitarily invariant)이라 한다. 단, 여기서 $\boldsymbol{U}$는 $m \times m$ 정규직교행렬(orthonormal matrix), $\boldsymbol{V}$는 $n \times n$ 직교행렬이다.

크기불변노름은 다음과 같은 성질을 가지는데 증명은 Stewart and Sun (1990)을 참고하기 바란다.

Lemma 1. *임의의 크기불변노름 $|||\cdot|||$에 대하여 다음과 같은 성질들이 성립된다.*

$$|||\boldsymbol{AB}||| \le |||\boldsymbol{A}||| \cdot \|\boldsymbol{B}\|, \qquad |||\boldsymbol{AB}||| \le |||\boldsymbol{B}||| \cdot \|\boldsymbol{A}\|,$$
$$|||\boldsymbol{AB}||| \ge |||\boldsymbol{A}|||\,\sigma_{\min}(\boldsymbol{B}), \qquad |||\boldsymbol{AB}||| \ge |||\boldsymbol{B}|||\,\sigma_{\min}(\boldsymbol{A}).$$

다음으로 고유치 및 비정칙치의 교란 상한값에 대한 따름정리를 소개한다. 이 정리에 대한 증명은 Tao (2012)를 참고하기 바란다.

Lemma 2. (고유치에 대한 Weyl 부등식) *대칭인 $n \times n$ 실수행렬 $\boldsymbol{A}$와 $\boldsymbol{E}$에 대하여 $\boldsymbol{A}$와 $\boldsymbol{A} + \boldsymbol{E}$의 i $(1 \le i \le n)$번째 큰 고유치는 다음과 같은 부등식을 갖는다.*

$$|\lambda_i(\boldsymbol{A}) - \lambda_i(\boldsymbol{A} + \boldsymbol{E})| \le \|\boldsymbol{E}\|.$$

Lemma 3. (비정칙치에 대한 Weyl 부등식) *$m \times n$ 실수행렬 $\boldsymbol{A}$와 $\boldsymbol{E}$에 대하여 $\boldsymbol{A}$와 $\boldsymbol{A} + \boldsymbol{E}$의 i $(1 \le i \le \min\{m, n\})$번째 큰 비정칙치는 다음과 같은 부등식을 갖는다.*

$$|\sigma_i(\boldsymbol{A} + \boldsymbol{E}) - \sigma_i(\boldsymbol{A})| \le \|\boldsymbol{E}\|.$$

위의 두 가지 Lemma는 다음과 같이 해석할 수 있다. 만약 오차항의 교란이 크지 않다면 고유치나 비정칙치들의 변동 역시 크지 않다고 할 수 있다.

3. 두 부공간의 거리측도

3.1. 부공간의 거리 및 각도

고유치공간(eigenspaces)과 비정칙 부공간(singular subspaces)에 대한 교란이론을 설명하기 위해 두 부공간에 대한 거리개념을 나타낼 수 있는 측도(metric)에 대해 논의할 필요가 있다.

3.1.1. 기호 및 정의

n 차원의 부공간 $\mathcal{U}^\star$와 $\mathcal{U}$의 차수는 r $(1 \le r \le n)$이라고 하면 이들 부공간은 $n \times r$ 행렬 $\boldsymbol{U}^\star$와 $\boldsymbol{U}$로 표현할 수 있다. 즉, 이들 행렬의 열들은 정규직교기저(orthonormal basis)을 이룬다. 이제 편의상 두 개의 $n \times (n-r)$ 행렬 $\boldsymbol{U}_\perp^\star$와 $\boldsymbol{U}_\perp$을 정의하는데 이들은 $[\boldsymbol{U}^\star, \boldsymbol{U}_\perp^\star]$와 $[\boldsymbol{U}, \boldsymbol{U}_\perp]$이 각각 $n \times n$ 직교행렬이 되도록 한다. 즉, $\boldsymbol{U}_\perp^\star$와 $\boldsymbol{U}_\perp$는 각각 $\boldsymbol{U}^\star$와 $\boldsymbol{U}$의 직교여사행렬(orthogonal complement matrix)이 된다.

3.1.2. 거리측도 및 주각

두 공간의 거리측도(distance metric)와 주각(principal angles)에 대한 개념을 논의하고 소개한다. 두 개의 부공간 $\mathcal{U}$와 $\mathcal{U}^\star$간의 거리로서 직관적으로 어떤 노름 $|||\cdot|||$에 대하여 $|||\boldsymbol{U} - \boldsymbol{U}^\star|||$을 생각할 수 있다. 하지만 이는 명백한 단점이 있다. 왜냐하면 어떤 $r \times r$ 회전행렬(rotation matrix) $\boldsymbol{R}$이 존재하여 $\boldsymbol{U}\boldsymbol{R}$이 $\mathcal{U}$의 정규직교기저가 될 수 있다. 따라서, 두 부공간 $\mathcal{U}$와 $\mathcal{U}^\star$은 동일함에도 불구하고 회전행렬이 사용되면 $|||\boldsymbol{U} - \boldsymbol{U}^\star||| \neq 0$이 되는 단점이 있으므로 $|||\boldsymbol{U} - \boldsymbol{U}^\star|||$을 두 공간의 거리측도로 사용할 수 없다. 따라서, 이상의 논의를 바탕으로 다음과 같은 거리측도를 생각할 수 있다.

(i) 최적 회전에 의한 거리
회전행렬 $\boldsymbol{R}$에 의한 단점을 보완하기 위해 다음과 같은 거리를 생각할 수 있으며 이러한 회전행렬을 최적 회전(optimal rotation)이라 부른다.

$$\mathrm{dist}_{|||\cdot|||}(\boldsymbol{U}, \boldsymbol{U}^\star) := \min_{\boldsymbol{R}} \left|\left|\left|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right|\right|\right|, \tag{3.1}$$

여기서, $|||\cdot|||$는 노름을 나타내며 주로 스펙트럴 노름이나 Frobenius 노름 등이 사용된다.

(ii) 정사영 행렬에 의한 거리
부공간 $\mathcal{U}$에 대한 정사영 행렬은 $\boldsymbol{U}\boldsymbol{U}^\top$로서 유일하게 존재하고 이는 어떤 회전행렬에 의해서도 영향을 받지 않는다. 왜냐하면 임의의 $r \times r$ 회전행렬에 대하여 $\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{U}\boldsymbol{R}\boldsymbol{R}^\top\boldsymbol{U}^\top$가 되기 때문이다. 이러한 사실로 부터 두 부공간 $\mathcal{U}$와 $\mathcal{U}^\star$의 거리를 다음과 같이 정의할 수 있다.

$$\mathrm{dist}_{\mathrm{p},|||\cdot|||}(\boldsymbol{U}, \boldsymbol{U}^\star) := \left|\left|\left|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right|\right|\right|, \tag{3.2}$$

여기서 $|||\cdot|||$는 노름을 나타내며 주로 스펙트럴 노름이나 Frobenius 노름 등이 사용되고 p는 정사영(projection)을 나타낸다.

(iii) 주각에 의한 거리
행렬 $\boldsymbol{U}^\top\boldsymbol{U}^\star$의 비정칙치를 $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r \ge 0$ 이라 하자. 만약 $\|\boldsymbol{U}^\top\boldsymbol{U}^\star\| \le \|\boldsymbol{U}\|\|\boldsymbol{U}^\star\| = 1$일 경우 모든 비정칙치 $\{\sigma_i\}_{i=1}^r$은 구간 $[0, 1]$에 존재한다. 이 경우, 두 부공간의 주각(principal angles 또는 canonical angles)은 다음과 같이 정의된다.

$$\theta_i := \arccos(\sigma_i), \quad i = 1, \ldots, r.$$

따라서,

$$0 \le \theta_1 \le \cdots \le \theta_r \le \pi/2$$

이 됨은 명백하다. 가장 간단한 경우로 $r = 1$인 경우를 살펴보자. 이때 주각 θ_1은 두 단일벡터(unit vector) $\boldsymbol{U}$와 $\boldsymbol{U}^\star$가 이루는 각도라는 사실은 잘 알려져 있다. 이러한 성질을 일반화시키면 두 부공간 $\mathcal{U}$와 $\mathcal{U}^\star$의 거리는 다음과 같이 정의할 수 있다.

$$\mathrm{dist}_{\sin,|||\cdot|||}(\boldsymbol{U}, \boldsymbol{U}^\star) := |||\sin\boldsymbol{\Theta}||| := \mathrm{diag}\left([\sin\theta_1, \ldots, \sin\theta_r]\right). \tag{3.3}$$

3.1.3. 거리측도간의 관계

위에서 소개한 (3.1), (3.2), (3.3) 거리측도들은 서로 밀접한 관련이 있다. 우선 다음 Lemma를 통해 $\mathrm{dist}_{\mathrm{p},|||\cdot|||}(\cdot,\cdot)$와 $\mathrm{dist}_{\sin,|||\cdot|||}(\cdot,\cdot)$ 사이의 관계를 살펴본다.

Lemma 4. *만약 $2r \le n$이라면, $\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}$의 비정칙치는 다음과 같이 주어진다.*

$$\underbrace{\sin\theta_r, \sin\theta_r, \sin\theta_{r-1}, \sin\theta_{r-1}, \ldots, \sin\theta_1, \sin\theta_1}_{2r}, \underbrace{0, 0, \ldots, 0}_{n-2r}.$$

즉, Lemma 4는 정사영 행렬에 의한 거리와 주각에 의한 거리의 관계를 나타내고 있다. 이 Lemma는 스펙트럼 노름과 Frobenius 노름 하에서 다음과 같이 표현할 수 있다.

Lemma 5. *임의의 $1 \le r \le n$에 대하여 다음이 성립한다.*

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| = \|\sin\boldsymbol{\Theta}\| = \left\|\boldsymbol{U}_\perp^\top\boldsymbol{U}^\star\right\| = \left\|\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\right\|;$$
$$\frac{1}{\sqrt{2}}\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|_F = \|\sin\boldsymbol{\Theta}\|_F = \left\|\boldsymbol{U}_\perp^\top\boldsymbol{U}^\star\right\|_F = \left\|\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\right\|_F.$$

다음으로, 위에서 언급한 두 노름 하에서 $\mathrm{dist}_{|||\cdot|||}(\cdot,\cdot)$와 $\mathrm{dist}_{\mathrm{p},|||\cdot|||}(\cdot,\cdot)$ 간의 관계는 다음과 같다.

Lemma 6. *임의의 $r \times r$ 회전행렬 $\boldsymbol{R}$ $(1 \le r \le n)$에 대하여 다음이 성립한다.*

$$\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\| \le \min_{\boldsymbol{R}}\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\| \le \sqrt{2}\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|;$$
$$\frac{1}{\sqrt{2}}\left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|_F \le \min_{\boldsymbol{R}}\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\|_F \le \left\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right\|_F.$$

위의 결과는 $\min_{\boldsymbol{R}}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|$와 $\min_{\boldsymbol{R}}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}$의 상한경계가 $\boldsymbol{U} = [1, 0]^\top$와 $\boldsymbol{U}^\star = [0, 1]^\top$일때 얻을 수 있음을 쉽게 보일 수 있다. 위의 결과로부터 $|||\cdot|||$가 스펙트럴 노름 또는 Frobenius 노름일때 $\mathrm{dist}_{|||\cdot|||}(\cdot,\cdot)$와 $\mathrm{dist}_{\mathrm{p},|||\cdot|||}(\cdot,\cdot)$는 $\sqrt{2}$를 제외하고 서로 같음을 알 수 있다. 실제 두 공간 $\boldsymbol{U}, \boldsymbol{U}^\star$의 거리측도로 다음의 두 가지가 주로 사용된다.

$$\mathrm{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) := \min_{\boldsymbol{R}}\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\|;$$
$$\mathrm{dist}_F(\boldsymbol{U}, \boldsymbol{U}^\star) := \min_{\boldsymbol{R}}\left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\|_F.$$

3.2. 교란이론에의 응용

3.2.1. 설정 및 기호

$\boldsymbol{M}^\star$ 와 $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E}$를 두 개의 $n \times n$ 실수대칭행렬이라고 하면 $M^\star$와 M의 고유분해(spectral decomposition)에 의해 다음과 같이 표현할 수 있다.

$$\boldsymbol{M}^\star = \sum_{i=1}^n \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top} = \begin{bmatrix} \boldsymbol{U}^\star & \boldsymbol{U}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^\star & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\star\top} \\ \boldsymbol{U}_\perp^{\star\top} \end{bmatrix},$$
$$\boldsymbol{M} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^\top \\ \boldsymbol{U}_\perp^\top \end{bmatrix},$$

여기서 $\{\lambda_i\}$ (또는 $\{\lambda_i^\star\}$)는 $\boldsymbol{M}$ (또는 $\boldsymbol{M}^\star$)의 고유치를 나타내고, $\boldsymbol{u}_i$ (또는 $\boldsymbol{u}_i^\star$)는 고유치 λ_i (또는 $\lambda_i^\star$)에 해당되는 고유벡터를 나타내고,

$$\begin{aligned} \boldsymbol{U} &:= [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r] \in \mathbb{R}^{n\times r}, \qquad & \boldsymbol{U}_\perp &:= [\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{n\times(n-r)}, \\ \boldsymbol{\Lambda} &:= \operatorname{diag}([\lambda_1, \ldots, \lambda_r]), \qquad & \boldsymbol{\Lambda}_\perp &:= \operatorname{diag}([\lambda_{r+1}, \ldots, \lambda_n]) \end{aligned}$$

로 표현하며, 마찬가지로 $\boldsymbol{U}^\star, \boldsymbol{U}_\perp^\star, \boldsymbol{\Lambda}^\star$, 그리고 $\boldsymbol{\Lambda}_\perp^\star$도 같은 의미를 나타낸다.

3.2.2. 예제

일반적으로 실수 대칭 행렬의 고유벡터(고유공간)은 작은 교란에도 크게 변할 수 있다. 이를 이해하기 위해 다음과 같은 간단한 예제를 생각해 보자 (Hsu, 2016).

$$\boldsymbol{M}^\star = \begin{bmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} -\epsilon & \epsilon \\ \epsilon & \epsilon \end{bmatrix}, \quad \boldsymbol{M} = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix},$$

여기서 $0 < \epsilon < 1$은 임의로 작은 값이 될 수 있고, $M^\star$와 M의 고유벡터가 각각 다음과 같이 주어짐을 쉽게 보일 수 있다:

$$\boldsymbol{u}_1^\star = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{그리고} \quad \boldsymbol{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

따라서, 스펙트럴 노름과 Frobenius 노름을 사용할 경우 각각 두 공간의 거리는 다음과 같이 주어진다.

$$\left\|\boldsymbol{u}_1\boldsymbol{u}_1^\top - \boldsymbol{u}_1^\star\boldsymbol{u}_1^{\star\top}\right\| = \frac{1}{\sqrt{2}}, \quad \text{그리고} \quad \left\|\boldsymbol{u}_1\boldsymbol{u}_1^\top - \boldsymbol{u}_1^\star\boldsymbol{u}_1^{\star\top}\right\|_F = 1,$$

여기서 두 값 모두 ϵ의 크기 또는 교란 $\boldsymbol{E}$의 크기에 상관없이 상당히 커질 수 있다. 아울러, 교란의 크기 ϵ이 $M^\star$의 인접한 고유치의 차이(eigengap)와 밀접한 관계(즉, $\lambda_1(M^\star) - \lambda_2(M^\star) = 2\epsilon$)가 있으며, 이것은 고유공간 교란에 영향을 미치는 인접한 고유치 차이가 매우 중요한 역할을 한다는 것을 암시한다.

3.2.3. Davis-Kahan $\sin\Theta$ *정리*

고유공간 교란 이론의 핵심적 결과는 Davis and Kahan (1970)의 정리에 나타나는데, 이는 교란행렬의 크기와 고유치 차이를 이용하여 두 공간의 거리의 상한값을 제시한 것이다. 여기서는 임의의 대칭 행렬 $\boldsymbol{A}$에 대해 $\boldsymbol{A}$의 고유치 집합을 $\texttt{eigenvalues}(\boldsymbol{A})$로 표시한다.

Theorem 1. *(Davis-Kahan* $\sin\boldsymbol{\Theta}$ *정리). 실수* $\alpha, \beta \in \mathbb{R}$*와 고유치 차이* $\Delta > 0$*에 대하여*

$$\begin{aligned} &\texttt{eigenvalues}\left(\boldsymbol{\Lambda}^\star\right) \subseteq [\alpha, \beta], \\ &\texttt{eigenvalues}\left(\boldsymbol{\Lambda}_\perp\right) \subseteq (-\infty, \alpha - \Delta] \cup [\beta + \Delta, \infty) \end{aligned} \tag{3.4}$$

을 만족하면

$$\begin{aligned} \operatorname{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) &\le \sqrt{2}\|\sin\boldsymbol{\Theta}\| \le \frac{\sqrt{2}\left\|\boldsymbol{E}\boldsymbol{U}^\star\right\|}{\Delta} \le \frac{\sqrt{2}\|\boldsymbol{E}\|}{\Delta}; \\ \operatorname{dist}_F(\boldsymbol{U}, \boldsymbol{U}^\star) &\le \sqrt{2}\|\sin\boldsymbol{\Theta}\|_F \le \frac{\sqrt{2}\left\|\boldsymbol{E}\boldsymbol{U}^\star\right\|_F}{\Delta} \le \frac{\sqrt{2r}\|\boldsymbol{E}\|}{\Delta} \end{aligned}$$

이 된다.

Theorem 1은 일반적으로 Davis-Kahan $\sin\boldsymbol{\Theta}$ 정리라고 불리며, 이는 부공간 사이의 $\sin\boldsymbol{\Theta}$ 거리와 관련되어 있기 때문이다. 두 상한값 모두 교란 크기에 비례하고, 인접한 고유치 차이 Δ에 반비례한다. 예를 들어, $\|\boldsymbol{E}\|$를 잡음 크기로 보고, 고유치 차이를 신호 강도로 해석하면, **Theorem 1**은 신호 대 잡음비가 감소함에 따라 고유공간 교란이 점진적으로 작아진다고 할 수 있다.

4. 결론

본 연구에서는 두 부공간의 거리측도에 대해 소개하였다. 이를 위해 여러 가지 행렬의 노름을 소개하고 주어진 노름하에서 여러 가지 행렬간의 거리측도를 소개하였으며 이들간의 관계를 설명하였다. 이를 이용하여 행렬의 교란이론에 적용하고 교란이론의 핵심적 결과인 Davis-Kahan (1970) 정리를 소개하였다. 두 부공간의 거리측도는 서론에서 소개한 것처럼 통계학의 다양한 분야 뿐만 아니라 후방탐색, 안면인식 등 인공지능의 중요한 분야에 많이 활용되고 있으므로 통계적 학습론의 기초 필수 이론이라고 생각된다.

References

Candès EJ, Li X, Ma, Y, and Wright J (2011). Robust principal component analysis?, *Journal of the ACM*, **58**, 1–37.

Chandrasekaran V, Sanghavi S, Parrilo PA, and Willsky AS (2011). Rank-Sparsity incoherence for matrix decomposition, *SIAM Journal on Optimization*, **21**, 572–596.

Chen Y, Chi Y, Fan J, and Ma C (2021). Spectral methods for data science: A statistical perspective, *Foundations and Trends® in Machine Learning*, **14**, 566–806.

Davis C and Kahan WM (1970). The rotation of eigenvectors by a perturbation. III, *SIAM Journal on Numerical Analysis*, **7**, 1–46.

Hsu D (2016). COMS 4772: Advanced machine learning, *Lecture notes, Columbia University*.

Li RC (1998). Relative perturbation theory: II. Eigenspace and singular subspace variations, *SIAM Journal on Matrix Analysis and Applications*, **20**, 471–492.

Stewart GW and Sun JG (1990). *Matrix Perturbation Theory*, Academic Press, Boston.

Tao T (2012). *Topics in Random Matrix Theory*, American Mathematical Society, Providence, R.I.

Received May 23, 2023; Revised May 30, 2023; Accepted May 30, 2023

두 부공간의 거리측도에 대한 요약

김기풍[a], 김충락[1,b]

[a]연세의료원 내과학교실; [b]부산대학교 통계학과

요 약

본 논문에서는 두 부공간의 거리측도에 대해 소개하였다. 이를 위해 여러 가지 행렬의 노름을 소개하고 주어진 노름하에서 여러 가지 행렬간의 거리측도를 소개하였으며 이들간의 관계를 설명하였다. 이를 이용하여 행렬의 교란이론에 적용하고 교란이론의 핵심적 결과인 Davis-Kahan (1970) 정리를 소개하였다. 두 부공간의 거리측도는 통계학의 다양한 분야 뿐만 아니라 후방탐색, 안면인식 등 인공지능의 중요한 분야에 많이 활용되고 있다.

주요용어: 고유공간, 교란이론, 비정칙치 부공간, 스펙트럴 노름, 프로베니우스 노름

본 연구는 부산대학교 2년 과제 연구비에 의하여 수행되었음.

[1]교신저자: (46288) 부산시 금정구 부산대학로 63번길 2, 부산대학교 통계학과. E-mail: crkim@pusan.ac.kr