PERTURBATION OF ANGLES BETWEEN LINEAR SUBSPACES*

SUN JI-GUANG (孙继广)
(*Computing Center, Academia Sinica, Beijing, China*)

Abstract

We consider in this note how the principal angles between column spaces $\mathscr{R}(A)$ and $\mathscr{R}(B)$ change when the elements in A and B are subject to perturbations. The basic idea in the proof of our results is that the non-zero cosine values of the principal angles between $\mathscr{R}(A)$ and $\mathscr{R}(B)$ coincide with the non-zero singular values of P_AP_B, the product of two orthogonal projections, and consequently we can apply a perturbation theorem of orthogonal projections proved by the author[4].

§ 1. Introduction

Let $\mathscr{X}$ and $\mathscr{Y}$ be given subspaces of the complex n-dimensional vector space $\mathbb{C}^n$, and assume that

$$p=\dim(\mathscr{X})\geqslant\dim(\mathscr{Y})=q\geqslant 1.$$

The principal angles $\theta_k\in[0, \pi/2]$ between $\mathscr{X}$ and $\mathscr{Y}$ are recursively defined for $k=1, \cdots, q$ by

$$\cos\theta_k=\max_{u\in\mathscr{X}}\max_{v\in\mathscr{Y}} u^Hv=u_k^Hv_k, \quad \|u\|_2=\|v\|_2=1,$$

subject to the constraints

$$u_j^Hu=0,\ v_j^Hv=0, \quad j=1, 2, \cdots, k-1.$$

In statistics the numbers $\cos\theta_1, \cdots, \cos\theta_k$ are called canonical correlation coefficients. Björck and Golub[1] pointed out that the principal angles are uniquely defined and have many important applications in statistics and numerical analysis.

Let $\mathscr{R}(A)$ be the column space of a complex $n\times p$ matrix A, and $\mathscr{R}(B)$ be the column space of a complex $n\times q$ matrix B. First order perturbation results for principal angles between $\mathscr{R}(A)$ and $\mathscr{R}(B)$ are given in [1], under the hypotheses of having a small change of A and B. The aim in this note is to remove the hypotheses and to derive perturbation bounds for principal angles. The main results are Theorem 3.1 and Theorem 3.2.

Notation. The symbol $\mathbb{C}^{n\times m}$ denotes the set of complex $n\times m$ matrices, and

$$\mathbb{C}_p^{n\times m}=\{A\in\mathbb{C}^{n\times m}\colon \operatorname{rank}(A)=p\}.$$

$\sigma(A)$ denotes the set of singular values of a matrix A, and $\sigma_+(A)$ the set of non-zero singular values of A. A^H is for conjugate transpose of A, and $I^{(n)}$ is the $n\times n$ identity matrix. $A^\dagger$ stands for the Moore-Penrose generalized inverse of a matrix A. $P_A=AA^\dagger$ is the orthogonal projection onto the column space $\mathscr{R}(A)$.

* Received October 14, 1985.

Besides, let $\|\ \|: \bigcup_{n,m=1}^{\infty} \mathbb{C}^{n\times m} \to \mathbb{R}_+$ (the set of nonnegative real numbers) be a family of unitarily invariant norms (u. i. n.), $\|\ \|_2$ be the spectral norm and $\|\ \|_F$ the Frobenius norm.

§ 2. Preliminaries

In this paragraph we will cite some results as the basis of our perturbation theorems for principal angles.

Theorem 2.1[1]. *Let $A\in\mathbb{C}_p^{n\times p}$ and $B\in\mathbb{C}_q^{n\times q}$, $p\geqslant q$. Assume that the columns of matrices U_A and U_B form unitary bases for two subspaces $\mathscr{R}(A)$ and $\mathscr{R}(B)$. Let the singular value decomposition* (SVD) *of the $p\times q$ matrix $U_A^H U_B$ be*

$$U_A^H U_B = UCV^H,\quad C=\operatorname{diag}(c_1, \cdots, c_q),$$

where $U^HU=V^HV=VV^H=I^{(q)}$. If we assume that $c_1\geqslant c_2\geqslant\cdots\geqslant c_q$, then the principal angles $\theta_1, \cdots, \theta_q$ associated with $\mathscr{R}(A)$ and $\mathscr{R}(B)$ are given by

$$\cos\theta_k = c_k,\quad k=1, \cdots, q.$$

In the following we use the symbol $\sigma(A, B)$ for the set $\{c_k\}_{k=1}^q$.

Theorem 2.2[2]. *Let $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_q$ and $\tilde\sigma_1\geqslant\tilde\sigma_2\geqslant\cdots\geqslant\tilde\sigma_q$ be the singular values of G and $\tilde G\in\mathbb{C}^{p\times q}$, respectively, $p\geqslant q$. Then for every unitarily invariant norm,*

$$\|\operatorname{diag}(\tilde\sigma_1-\sigma_1, \cdots, \tilde\sigma_q-\sigma_q)\| \leqslant \|\tilde G - G\|.$$

Theorem 2.3[4]. *Let $Z, W\in\mathbb{C}^{n\times m}$. If* $\operatorname{rank}(W)=\operatorname{rank}(Z)$, *then*

$$\|P_W - P_Z\| \leqslant \mu \min\{\|Z^\dagger\|_2, \|W^\dagger\|_2\}\|W-Z\|,$$

where μ is given in the following table:

$\|\ \|$	*arbitrary u. i. n.*	*Frobenius*	*spectral*
μ	2	$\sqrt{2}$	1

§ 3. Perturbation Theorems

Let $A\in\mathbb{C}_p^{n\times p}$, $B\in\mathbb{C}_q^{n\times q}$, $p\geqslant q$. Assume that the columns of U_A and U_B form unitary bases for $\mathscr{R}(A)$ and $\mathscr{R}(B)$, respectively. First we prove two lemmas.

Lemma 3.1. *Suppose that $\sigma(U_A^HU_B)=\{c_k\}_{k=1}^q$, $c_k=\cos\theta_k$, $k=1, \cdots, q$, $\frac{\pi}{2}\geqslant\theta_1\geqslant\cdots\geqslant\theta_q\geqslant0$. If (U_A, W_A) is an $n\times n$ unitary matrix and $\sigma(W_A^HU_B)=\{s_k\}_{k=1}^q$, $s_1\geqslant\cdots\geqslant s_q$, then*

$$s_k=\sin\theta_k,\quad k=1, \cdots, q. \tag{3.1}$$

Proof. From

$$(U_A^HU_B)^H(U_A^HU_B)+(W_A^HU_B)^H(W_A^HU_B)=I^{(q)}$$

it follows that

$$c_k^2+s_k^2=1,\quad k=1, \cdots, q.$$

Thus we get the relations (3.1). ∎

Lemma 3.2. *For the above mentioned A, B, U_A, U_B and W_A, we have*

$$\sigma_+(U_A^H U_B)=\sigma_+(P_A P_B) \tag{3.2}$$

and

$$\sigma_+(W_A^H U_B)=\sigma_+((I-P_A)P_B). \tag{3.3}$$

Proof. Assume that the SVD of $U_A^H U_B$ is

$$U_A^H U_B=U_1\begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}V_1^H, \tag{3.4}$$

where $U_1^H U_1=V_1^H V_1=V_1 V_1^H=I^{(q)}$, $C_1=\mathrm{diag}(c_1, \cdots, c_r)$, $c_1\geqslant\cdots\geqslant c_r>0$, $r\leqslant q$. Obviously,

$$\sigma_+(U_A^H U_B)=\{c_k\}_{k=1}^r. \tag{3.5}$$

On the other hand, from (3.4),

$$P_A P_B=U_A U_A^H U_B U_B^H=U_A U_1\begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}V_1^H U_B^H=U_2\begin{pmatrix} C_1 & 0 \\ 0 & 0 \end{pmatrix}V_2^H, \tag{3.6}$$

where $U_2=U_A U_1\in\mathbb{C}^{n\times q}$ and $V_2=U_B V_1\in\mathbb{C}^{q\times n}$ satisfy $U_2^H U_2=V_2^H V_2=I^{(q)}$. The decomposition (3.6) means that

$$\sigma_+(P_A P_B)=\{c_k\}_{k=1}^r. \tag{3.7}$$

Comparison of (3.5) and (3.7) gives (3.2).

Observe that

$$W_A W_A^H=I-U_A U_A^H=I-P_A,$$

hence utilizing a similar argument as above we get (3.3). ∎

Theorem 3.1. *Let* $A, \widetilde{A}\in\mathbb{C}^{n\times p}$, $B, \widetilde{B}\in\mathbb{C}^{n\times q}$, $p\geqslant q$. *Suppose that* $\sigma(A, B)=\{c_k\}_{k=1}^q$, $c_k=\cos\theta_k$, $k=1, \cdots, q$, $\frac{\pi}{2}\geqslant\theta_1\geqslant\cdots\geqslant\theta_q\geqslant 0$, $\sigma(\widetilde{A}, \widetilde{B})=\{\tilde{c}_k\}_{k=1}^q$, $\tilde{c}_k=\cos\tilde{\theta}_k$, $k=1, \cdots, q$, $\frac{\pi}{2}\geqslant\tilde{\theta}_1\geqslant\cdots\geqslant\tilde{\theta}_q\geqslant 0$. *Let*

$$C=\mathrm{diag}(c_1,\cdots, c_q), \qquad \widetilde{C}=\mathrm{diag}(\tilde{c}_1, \cdots, \tilde{c}_q), \tag{3.8}$$

and

$$S=\mathrm{diag}(s_1, \cdots, s_q), \qquad \widetilde{S}=\mathrm{diag}(\tilde{s}_1, \cdots, \tilde{s}_q), \tag{3.9}$$

where $s_k=\sin\theta_k$, $\tilde{s}_k=\sin\tilde{\theta}_k$, $k=1, \cdots, q$. *Then for every unitarily invariant norm,*

$$\|\widetilde{C}-C\|,\ \|\widetilde{S}-S\|\leqslant\delta(A, \widetilde{A})+\delta(B, \widetilde{B}); \tag{3.10}$$

for the Frobenius norm,

$$\|\widetilde{C}-C\|_F,\ \|\widetilde{S}-S\|_F\leqslant\delta_F(A, \widetilde{A})+\delta_F(B, \widetilde{B}); \tag{3.11}$$

and for the spectral norm,

$$\|\widetilde{C}-C\|_2,\ \|\widetilde{S}-S\|_2\leqslant\delta_2(A, \widetilde{A})+\delta_2(B, \widetilde{B}). \tag{3.12}$$

Where

$$\delta(X, \widetilde{X})=2\|X\|\,\|X^\dagger\|_2\cdot\frac{\|\widetilde{X}-X\|}{\|X\|}, \tag{3.13}$$

$$\delta_F(X, \widetilde{X})=\sqrt{2}\|X\|_F\|X^\dagger\|_2\cdot\frac{\|\widetilde{X}-X\|_F}{\|X\|_F} \tag{3.14}$$

and

$$\delta_2(X, \widetilde{X})=\|X\|_2\|X^\dagger\|_2\cdot\frac{\|\widetilde{X}-X\|_2}{\|X\|_2}. \tag{3.15}$$

Proof. First assume that the columns of U_A, $U_{\widetilde{A}}$, U_B and $U_{\widetilde{B}}$ form unitary bases for $\mathscr{R}(A)$, $\mathscr{R}(\tilde{A})$, $\mathscr{R}(B)$ and $\mathscr{R}(\tilde{B})$, respectively. By the hypotheses, we have

$$\sigma(U_A^H U_B)=\{c_k\}_{k=1}^q, \quad \sigma(U_{\tilde{A}}^H U_{\tilde{B}})=\{\tilde{c}_k\}_{k=1}^q.$$

Let W_A and $W_{\tilde{A}}$ be such that (U_A, W_A) and $(U_{\tilde{A}}, W_{\tilde{A}})$ are $n\times n$ unitary matrices. Then from Lemma 3.1

$$\sigma(W_A^H U_B)=\{s_k\}_{k=1}^q, \quad \sigma(W_{\tilde{A}}^H U_{\tilde{B}})=\{\tilde{s}_k\}_{k=1}^q.$$

By Lemma 3.2 and Theorem 2.2 we get

$$\|\tilde{C}-C\|\leqslant\|P_{\tilde{A}}P_{\tilde{B}}-P_AP_B\|\leqslant\|P_{\tilde{A}}-P_A\|+\|P_{\tilde{B}}-P_B\| \tag{3.16}$$

and

$$\|\tilde{S}-S\|\leqslant\|(I-P_{\tilde{A}})P_{\tilde{B}}-(I-P_A)P_B\|\leqslant\|P_{\tilde{A}}-P_A\|+\|P_{\tilde{B}}-P_B\|. \tag{3.17}$$

Utilizing Theorem 2.3, from (3.16) and (3.17) we obtain inequalities (3.10)—(3.12) at once. ∎

Observe that if

$$|\cos\tilde{\theta}-\cos\theta|\leqslant\delta, \quad |\sin\tilde{\theta}-\sin\theta|\leqslant\delta, \qquad \theta, \tilde{\theta}\in[0, \pi/2],$$

then

$$|\tilde{\theta}-\theta|\leqslant\frac{\pi}{2}\delta.$$

Hence from (3.11) and (3.12) we can deduce the following theorem.

Theorem 3.2. *Assuming the hypotheses of Theorem 3.1, then we have*

$$\sqrt{\sum_{k=1}^q(\tilde{\theta}_k-\theta_k)^2}\leqslant\frac{\pi}{2}(\delta_F(A, \tilde{A})+\delta_F(B, \tilde{B})) \tag{3.18}$$

and

$$|\tilde{\theta}_k-\theta_k|\leqslant\frac{\pi}{2}(\delta_2(A, \tilde{A})+\delta_2(B, \tilde{B})), \quad k=1, \cdots, q, \tag{3.19}$$

where $\delta_F(\ ,\ )$ *and* $\delta_2(\ ,\ )$ *are defined by* (3.14) *and* (3.15), *respectively.*

References

[1] A. Björck, G. H. Golub, Numerical methods for computing angles between linear subspaces, *Math. Comp.*, **27** (1973), 579—594.

[2] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Quart. J. Math. Oxford*, **11** (1960), 50—59.

[3] G. W. Stewart, On the perturbation of pseudo-inverses, projections and linear least squares problems, *SIAM Rev.*, **19** (1977), 634—662.

[4] Sun Ji-guang, The stability of orthogonal projections, *J. Graduate School*, **1** (1984), 123—133.