SOLUTIONS OF ORTHOGONAL PROCRUSTES PROBLEMS UNDER PARTIALLY KNOWN COLUMNS

Jun Xu and David Zhang

DEPARTMENT OF COMPUTING, THE HONG KONG POLYTECHNIC UNIVERSITY

Abstract

The orthogonal Procrustes problem aims to find an orthogonal matrix $\mathbf{T}$ which transforms one given matrix $\mathbf{A}$ into another one $\mathbf{B}$ by minimizing the residual difference under Frobenius norm of matrix, i.e., $\min \|\mathbf{B} - \mathbf{TA}\|_F^2$. This problem can be applied in factor analysis, dictionary learning, camera calibration, and robotics, *etc*. In many real applications, some certain factors may have been partially known previously. The prior information could be formulated as partially known columns in $\mathbf{T}$ and would influence the solution of the whole orthogonal transformation matrix, and hence makes the original orthogonal Procrustes problem more difficult. In this paper, we studied the solution of orthogonal Procrustes problem under partially known columns, which includes the original orthogonal Procrustes problem as a special case.

Key words: Orthogonal Procrustes problem, partially known columns

1. Introduction

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ be two given matrices of observed data, we consider the problem of transforming the matrix $\mathbf{A}$ into the matrix $\mathbf{B}$ by an orthogonal matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$ so that the residual difference between $\mathbf{B}$ and $\mathbf{TA}$ is minimized under the least square sense. This is the classical orthogonal Procrustes problem [1] which can be formally stated as follows:

$$\min_{\mathbf{T}} \|\mathbf{B} - \mathbf{TA}\|_F^2 \quad \text{s.t.} \quad \mathbf{T}^\top \mathbf{T} = \mathbf{I}_{n \times n}, \tag{1}$$

where "$\top$" stands for the transpose operation. As Schönemann pointed out in [3], the solution of the orthogonal transformation matrix $\mathbf{T}$ for problem (1) is $\hat{\mathbf{T}} = \mathbf{UV}^\top$, where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are obtained by performing singular value decomposition (SVD) on $\mathbf{BA}^\top$ via $\mathbf{BA}^\top = \mathbf{U\Sigma V}^\top$.

Many researchers have studied this problem in different perspectives. Green [2] derived the solution of problem (1) when the matrices $\mathbf{A}$ and $\mathbf{B}$ are of full column rank. Schönemann [3] generalized the solution of problem (1) to matrices $\mathbf{A}$ and $\mathbf{B}$ of arbitrary rank. Schönemann further studied the two-sided orthogonal Procrustes problems [4]. Lissitz *et al.* [5] and Koschat [6] studied the weighted orthogonal Procrustes problems, in which the weights are arranged as a diagonal matrix. More details about weighted orthogonal

Procrustes problems can be found in Viklands's PhD thesis [7]. Higham [8] studied the symmetric Procrustes problem by replacing the orthogonal restriction of $\mathbf{T}^\top \mathbf{T} = \mathbf{I}_{n \times n}$ with the symmetric restriction of $\mathbf{T}^\top = \mathbf{T}$. Later, Watson *et al.* studied the solutions of orthogonal Procrustes problems under general matrix norms such as orthogonally invariant norms [9] and the commonly studied ℓ_1 norm [10]. The Procrustes problems under orthogonally invariant norms usually could not be solved in closed-form. Instead, iterative manners are proposed which involved techniques such as projected gradient methods [11] and the famous Newton's method. In [10], the weighted orthogonal Procrustes problem is also solved by projected gradient methods [11]. Berge [12] further discussed the rigid rotation problems which require the determinant of the orthogonal transformation matrix $\mathbf{T}$ to be 1 (i.e., eliminating the possibility of reflection or flip).

The classical orthogonal Procrustes problem [1] has wide applications in different research fields, such as factor analysis [2], rigid rotation [12] in robotics, dimensional reduction [13] in machine learning, camera calibration [14] and point matching [15] in computer vision, dictionary learning [16] in signal processing, *etc*. The extended two-sided orthogonal Procrustes problem has been applied into matrix perturbation problems [4].

In this paper, we study the solution of the orthogonal Procrustes problem when partial columns of the orthogonal transformation matrix $\mathbf{T}$ are known. This is a problem emerged in many practical applications. For examples, in camera calibration, some certain axis is usually fixed for modeling the calibration process [14], *etc*. In the point matching problem [15], if a part of the points have already been matched, the matched data could be used to guide the matching of the remaining points. In signal processing problems, dictionary learning [16] is widely used for image recovery. Some dictionary items can be learned from external clean images beforehand, and can be employed to guide the learning of the dictioanry items for the internal degraded image. The overall learned dictionary would integrate the advantages of both sides (i.e., external and internal) as well as be more adaptive to the degraded image being processed. We give a sufficiency and necessary conditions for the solution of this new emerged problem. Specifically, we presented a new proof on the sufficiency condition of the solution by using the generalized Kristof's Theorem [17], which is different from the proof in [3]. We also discuss the number of solutions when the matrices $\mathbf{A}$ and $\mathbf{B}$ are in nonsingular or singular conditions.

2. Definition of the Problem and Solution

Assume $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ be two given data matrices. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be the partially known columns. Eventhough the partially known columns may not be stacked together, we can stack them together into a matrix $\mathbf{X}$ by resorting the matrix columns. Define $\mathbf{P} \in \mathbb{R}^{n \times q}$ ($p + q = n$) as the unknown orthogonal matrix with the remaining columns $\mathbf{P}$. From Section 1, we can see that $[\mathbf{X}\ \mathbf{P}] = \mathbf{T} \in \mathbb{R}^{n \times n}$ is the whole orthonormal matrix which transforms the matrix $\mathbf{A}$ to best fit the matrix $\mathbf{B}$ under the least square sense. Instead of obtaining $\mathbf{T}$ in classical orthogonal Procrustes problem, we need to obtain partial columns of $\mathbf{T}$ under the restriction that partial columns are already given. This problem is more difficult than the original orthogonal Procrustes problem.

For simplicity, we assume $n \geq m$ and the other cases could be analyzed in a similar way. We formulate the orthogonal Procrustes problem with partially known columns as follows:

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{D}} \|\mathbf{B} - [\mathbf{X}\ \mathbf{P}]\mathbf{A}\|_F^2 \quad \text{s.t.} \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}, \mathbf{X}^\top \mathbf{P} = \mathbf{0}_{p \times q}. \tag{2}$$

Note that the original orthogonal Procrustes problem is a special case of the problem (2): when there is no

prior known column at all, i.e., $\mathbf{X} = \emptyset$, then we have $p = 0, q = n$ and the studied problem is reduced to the original orthogonal Procrustes problem. When the partially known columns $\mathbf{X} \neq \emptyset$, we crop the matrix $\mathbf{A}$ into two parts: $\mathbf{A}_X \in \mathbb{R}^{p \times m}$ and $\mathbf{A}_P \in \mathbb{R}^{q \times m}$ to interact with $\mathbf{X}$ and $\mathbf{P}$, respectively. Then we have

$$\|\mathbf{B} - [\mathbf{X}\ \mathbf{P}]\mathbf{A}\|_F^2 = \|\mathbf{B} - [\mathbf{X}\ \mathbf{P}][\mathbf{A}_X^\top\ \mathbf{A}_P^\top]^\top\|_F^2 = \|\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top - \mathbf{P}\mathbf{A}_P^\top\|_F^2. \tag{3}$$

Here, $\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$ is a known data matrix and we replace it with $\mathbf{B}^* = \mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$. In the following Results 1, we remove the notation "$*$" and directly use $\mathbf{B}$ to replace $\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$ as the finally known data matrix. Besides, we remove the subscript "P" in $\mathbf{A}_P$ for symbol simplicity.

Results 1: Let $\mathbf{A} \in \mathbb{R}^{q \times m}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$ ($n \geq m$) be two given data matrices, $\mathbf{X}$ is a given matrix consisting of partially known columns satisfying $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_{p \times p}$, then the sufficiency and necessary conditions of

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} \|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 \quad \text{s.t.} \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}, \mathbf{X}^\top \mathbf{P} = \mathbf{0}_{p \times q} \tag{4}$$

is $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{n \times q}$ and $\mathbf{V} \in \mathbb{R}^{q \times q}$ are the orthogonal matrices obtained by performing economy (a.k.a. reduced) SVD:

$$(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \tag{5}$$

Proof: Since $\mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}$, we have

$$\hat{\mathbf{P}} = \arg\min_{\mathbf{P}} \|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 = \arg\min_{\mathbf{P}} \|\mathbf{B}\|_F^2 + \|\mathbf{P}\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{B}^\top \mathbf{P}\mathbf{A}) = \arg\max_{\mathbf{P}} \mathrm{Tr}(\mathbf{A}\mathbf{B}^\top \mathbf{P}). \tag{6}$$

We can use Lagrange multiplier method and define the Lagrange function as:

$$\mathcal{L} = \mathrm{Tr}(\mathbf{A}\mathbf{B}^\top \mathbf{P}) - \mathrm{Tr}(\Gamma_1(\mathbf{P}^\top \mathbf{P} - \mathbf{I}_{q \times q})) - \mathrm{Tr}(\Gamma_2(\mathbf{P}^\top \mathbf{X})), \tag{7}$$

where Γ is the Lagrange multiplier. Take the derivative of $\mathcal{L}$ with respect to $\mathbf{P}$ and set it to $\mathbf{0}$ with proper dimensions, we can get

$$\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = \mathbf{B}\mathbf{A}^\top - \mathbf{P}(\Gamma_1 + \Gamma_1^\top) - \mathbf{X}\Gamma_2^\top = \mathbf{0}_{n \times q}. \tag{8}$$

Since $\mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}$ and $\mathbf{X}^\top \mathbf{P} = \mathbf{0}_{p \times q}$, by left multiplying both sides of the Eq. (8) by $\mathbf{X}^\top$, we have

$$\mathbf{X}^\top \mathbf{B}\mathbf{A}^\top = \Gamma_2^\top. \tag{9}$$

Put the Eq. (9) back into Eq. (8), we have

$$\mathbf{B}\mathbf{A}^\top - \mathbf{P}(\Gamma_1 + \Gamma_1^\top) - \mathbf{X}\mathbf{X}^\top \mathbf{B}\mathbf{A}^\top = 0. \tag{10}$$

Or equivalently,

$$(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{P}(\Gamma_1 + \Gamma_1^\top). \tag{11}$$

Right multiplying both sides of Eq. (11) by $\mathbf{P}^\top$, we have

$$(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top \mathbf{P}^\top = \mathbf{P}(\Gamma_1 + \Gamma_1^\top)\mathbf{P}^\top. \tag{12}$$

This shows that $(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\mathbf{P}^\top$ is a symmetric matrix of order $n \times n$. Then we perform economy (or reduced) singular value decomposition (SVD) on $(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. Since $(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\mathbf{P}^\top$ is symmetric, we have

$$(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\mathbf{P}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{P}^\top = \mathbf{P}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \tag{13}$$

and hence we have $\mathbf{U} = \mathbf{P}\mathbf{V}$ and equivalently $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$, since $\mathbf{V}\mathbf{V}^\top = \mathbf{I}_{q\times q}$. Note that we can also employ the property that "each symmetric matrix could be diagonalized" to obtain the same result. Hence, the necessary condition is proofed.

Now we proof the sufficiency condition. If $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$, then it is easy to check that $\hat{\mathbf{P}}^\top\hat{\mathbf{P}} = \mathbf{I}_{q\times q}$. To proof that $\mathbf{X}^\top\hat{\mathbf{P}} = \mathbf{0}_{p\times q}$, we can resort the Eq. (5) for help. Since $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_{p\times p}$, by left multiplying both sides of Eq. (5) by $\mathbf{X}^\top$, we have

$$\mathbf{X}^\top(\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{X}^\top\mathbf{B}\mathbf{A}^\top - \mathbf{X}^\top\mathbf{X}\mathbf{X}^\top\mathbf{B}\mathbf{A}^\top = \mathbf{0}_{p\times n}. \tag{14}$$

It means that $\mathbf{X}^\top\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \mathbf{0}_{p\times p}$ and hence $\mathbf{X}^\top\mathbf{U} = \mathbf{0}_{p\times p}$. Then $\mathbf{X}^\top\hat{\mathbf{P}} = \mathbf{X}^\top\mathbf{U}\mathbf{V}^\top = \mathbf{0}_{p\times q}$.

Besides, since

$$\|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 = \|\mathbf{B}\|_F^2 + \|\mathbf{P}\mathbf{A}\|_F^2 - 2\mathrm{Tr}(\mathbf{B}^\top\mathbf{P}\mathbf{A}), \tag{15}$$

Until now, if we want to proof that $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$ is the solution of problem (3), $\mathrm{Tr}(\mathbf{B}^\top\hat{\mathbf{P}}\mathbf{A})$ has to be a maximum if $\|\mathbf{B} - \hat{\mathbf{P}}\mathbf{A}\|_F^2$ is to be a minimum, as long as $\mathbf{P}$ satisfying the conditions in Eq. (3). Note that by cyclic perturbation which retains the trace unchanged and due to $\mathbf{X}^\top\hat{\mathbf{P}} = \mathbf{0}_{p\times q}$, we have

$$\begin{aligned} \mathrm{Tr}(\mathbf{B}^\top\hat{\mathbf{P}}\mathbf{A}) &= \mathrm{Tr}(\mathbf{B}\mathbf{A}^\top\hat{\mathbf{P}}^\top) \\ &= \mathrm{Tr}((\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\hat{\mathbf{P}}^\top) \\ &= \mathrm{Tr}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{V}\mathbf{U}^\top) \\ &= \mathrm{Tr}(\boldsymbol{\Sigma}). \end{aligned} \tag{16}$$

Now we need to proof that $\mathrm{Tr}(\boldsymbol{\Sigma}) \geq \mathrm{Tr}(\mathbf{B}^\top\mathbf{P}\mathbf{A})$ for every $\mathbf{P}$ satisfying that $\mathbf{P}^\top\mathbf{P} = \mathbf{I}_{q\times q}, \mathbf{X}^\top\mathbf{P} = \mathbf{0}_{p\times q}$. Since $\mathrm{Tr}(\mathbf{B}^\top\mathbf{P}\mathbf{A}) = \mathrm{Tr}((\mathbf{I}_{n\times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\mathbf{P}^\top) = \mathrm{Tr}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{P}^\top) = \mathrm{Tr}(\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{P}^\top\mathbf{U}) \leq \mathrm{Tr}(\boldsymbol{\Sigma})$. The last inequality can be obtained by using a generalization version [17] of the Kristof's Theorem [18]. The equality is obtained at $\mathbf{V}^\top\mathbf{P}^\top\mathbf{U} = \mathbf{I}_{q\times q}$, i.e., $\mathbf{P} = \mathbf{U}\mathbf{V}^\top = \hat{\mathbf{P}}$. This completes the proof. ∎

Note that if the partially known prior were not present, the solution is clearly the solution of the original orthogonal Procrustes problem, i.e., $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are the orthogonal matrices obtained by performing economy (a.k.a. reduced) SVD: $\mathbf{B}\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. The difference between the solutions of the original orthogonal Procrustes problem and its partially known prior version quantify the effect on the residual of requiring $\mathbf{P}$ to be orthogonal to the external prior $\mathbf{P}^\top\mathbf{X}$.

3. Uniqueness of Solution $\hat{\mathbf{P}}$

Now we discuss the uniqueness of the solution $\hat{\mathbf{P}}$. Our discussion is according to the rankness of the $\boldsymbol{\Sigma}$ generated in the SVD. We first give a lemma describing the rank of $\mathbf{X}\mathbf{X}^\top$ where $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the partially known orthogonal matrix.

Lemma 1: Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be orthogonal matrix with $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_{p \times p}$, then $\mathrm{rank}(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top) \geq n - p$.

Proof: We firstly proof that $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top) = p$. The upper bound of the $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top)$ can be determined by $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top) \leq \min\{\mathrm{rank}(\mathbf{X}), \mathrm{rank}(\mathbf{X}^\top)\} = p$. The lower bound of the $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top)$ can be determined by Sylvester's inequality as $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top) \geq \mathrm{rank}(\mathbf{X}) + \mathrm{rank}(\mathbf{X}^\top) - p = 2p - p = p$. Hence, we have $\mathrm{rank}(\mathbf{X}\mathbf{X}^\top) = p$. Then, $\mathrm{rank}(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top) \geq \mathrm{rank}(\mathbf{I}_{n \times n}) - \mathrm{rank}(\mathbf{X}\mathbf{X}^\top) \geq n - p$. ∎

The rank of $\boldsymbol{\Sigma}$ largely depends on the rank of $\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top$, $\mathbf{B}$ and $\mathbf{A}$. Note that the rank of $\mathbf{B}$ and $\mathbf{A}$ are not larger than m and q, respectively. The rank of $\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top$ is at least $n - q = p$ (since $\mathrm{rank}(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top) \geq \mathrm{rank}(\mathbf{I}_{n \times n}) - \mathrm{rank}(\mathbf{X}\mathbf{X}^\top) \geq n - p = q$). From above observations, we can see that the rank of $\boldsymbol{\Sigma}$ can be equal to or less than q.

Results 2: If $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top$ is nonsingular, then $\mathrm{rank}(\boldsymbol{\Sigma}) = q$, and $\boldsymbol{\Sigma}$ may have distinct or multiple non-zero singular values. In the SVD of $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, the singular vectors in $\mathbf{U}$ and $\mathbf{V}$ can be determined up to orientation. Hence, we can reformulate the SVD as

$$(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}^* \mathbf{K}_u \boldsymbol{\Sigma} \mathbf{K}_v (\mathbf{V}^*)^\top, \tag{17}$$

where $\mathbf{U}^* \in \mathbb{R}^{n \times q}$ and $\mathbf{V}^* \in \mathbb{R}^{q \times q}$ are arbitrarily orientated singular vectors of $\mathbf{U}$ and $\mathbf{V}$, respectively. $\boldsymbol{\Sigma} \in \mathbb{R}^{q \times q}$ are diagonal matrix with singular values are arranged in weak descending order along the diagonal, i.e., $\boldsymbol{\Sigma}_{11} \geq \boldsymbol{\Sigma}_{22} \geq ... \geq \boldsymbol{\Sigma}_{qq} \geq 0$. The $\mathbf{K}_u$ and $\mathbf{K}_v$ are diagonal matrices with $+1$ or -1 as diagonal elements in arbitrary distribution. If we fix $\mathbf{K}_u$, then $\mathbf{K}_v$ is uniquely determined to meet the requirement that the diagonal elements of $\boldsymbol{\Sigma}$ should be nonnegative. And the orientations of the singular vectors of $\mathbf{U}^*$ is fixed, then the $\mathbf{U} = \mathbf{U}^* \mathbf{K}_u$ is determined, so does the orientations of the singular vectors of $\mathbf{V}^*$ and $\mathbf{V}^\top = \mathbf{K}_v (\mathbf{V}^*)^\top$. In this case, the solution of $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top = \mathbf{U}^* \mathbf{K}_u \mathbf{K}_v (\mathbf{V}^*)^\top$ is unique. The case that the $\boldsymbol{\Sigma}$ have multiple singular values also has unique solution of $\hat{\mathbf{P}}$, which can be discussed in a similar way.

If $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top$ is singular, then $0 \leq \mathrm{rank}(\boldsymbol{\Sigma}) = r < q$, and there is $q - r$ (at least one) zero singular values. The previous discussion on non-degenerative case still can be applied to the singular vectors related to the nonzero singular values, and this part is still unique. However, the singular vectors related to the zero singular values could be in arbitrary orientations as long as they satisfy the orthogonal conditions that $\mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}_{q \times q}$. Note that $\mathbf{U} \in \mathbb{R}^{n \times q}$, so $\mathbf{U}\mathbf{U}^\top$ no longer equals to the identity matrix of order n. From Eq. (12), we can get

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{P}^\top = \mathbf{P}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \tag{18}$$

Right multiplying each side by $\mathbf{PV}$ and then left multiplying each side by $\mathbf{U}^\top$, we can get

$$\boldsymbol{\Sigma} = \mathbf{U}^\top \mathbf{P}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \mathbf{P}\mathbf{V} \tag{19}$$

Hence, we can define a diagonal matrix $\mathbf{D} = \mathbf{U}^\top \mathbf{P}\mathbf{V} \in \mathbb{R}^{q \times q}$, the diagonal elements of which are

$$\mathbf{D}_{ii} = \begin{cases} 1 & \text{if } 1 \leq i \leq r; \\ \pm 1 & \text{if } r < i \leq q. \end{cases}$$

Thus, we obtain that $\mathbf{P} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where $\mathbf{D}$ is defined above. Hence, once we get the solution of $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$ in problem (3), the final solution for $\mathbf{P}$ when $\mathrm{rank}(\boldsymbol{\Sigma}) < q$ is not unique since the matrix $\mathbf{D}$ is not uniquely determined. In fact, since the number of $\mathbf{D}$ with different diagonal combinations is 2^{q-r}, the number of solutions for $\mathbf{P}$ is 2^{q-r} given fixed $\mathbf{U}$ and $\mathbf{V}$.

4. Concluding Remarks

In this paper, we studied the classical orthogonal Procrustes problem and gave the solution of this problem under partially known guidance, which included the original orthogonal Procrustes problem as a special case (i.e., when there is no guidance at all). Since the orthogonal Procrustes problem had been generalized to two-sided version [4] and weighted version [5, 6], we are highly motivated to study these generalized problems under partially known guidance.

References

[1] John R. Hurley and Raymond B. Cattell. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7(2):258–262, 1962.

[2] Bert F. Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17(4):429–440, 1952.

[3] Peter H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[4] Peter H. Schönemann. On two-sided orthogonal procrustes problems. *Psychometrika*, 33(1):19–33, 1968.

[5] Robert W. Lissitz, Peter H. Schönemann, and James C. Lingoes. A solution to the weighted procrustes problem in which the transformation is in agreement with the loss function. *Psychometrika*, 41(4):547–550, 1976.

[6] Martin A. Koschat and Deborah F. Swayne. A weighted procrustes criterion. *Psychometrika*, 56(2):229–239, 1991.

[7] Thomas Viklands. Algorithms for the weighted orthogonal procrustes problem and other least squares problems. *PhD Thesis*, 2006.

[8] Nicholas J. Higham. The symmetric procrustes problem. *BIT Numerical Mathematics*, 28(1):133–143, 1988.

[9] G. A. Watson. The solution of orthogonal procrustes problems for a family of orthogonally invariant norms. *Advances in Computational Mathematics*, 2(4):393–405, 1994.

[10] Nickolay T. Trendafilov and GA. Watson. The ℓ_1 oblique procrustes problem. *Statistics and Computing*, 14(1):39–51, 2004.

[11] Moody T. Chu and Kenneth R. Driessel. The projected gradient method for least squares matrix approximations with spectral constraints. *SIAM Journal on Numerical Analysis*, 27(4):1050–1060, 1990.

[12] Jos M. F. ten Berge. The rigid orthogonal procrustes rotation problem. *Psychometrika*, 71(1):201–205, 2006.

[13] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

[14] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[15] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, Apr 1991.

[16] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[17] Jos M. F. Ten Berge. A generalization of kristof's theorem on the trace of certain matrix products. *Psychometrika*, 48(4):519–523, 1983.

[18] Walter Kristof. A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, 7(3):515 – 530, 1970.