SOLUTIONS OF ORTHOGONAL PROCRUSTES PROBLEMS UNDER PARTIALLY KNOWN GUIDANCE

JUN XU AND DAVID ZHANG

DEPARTMENT OF COMPUTING, THE HONG KONG POLYTECHNIC UNIVERSITY

Abstract

The orthogonal Procrustes problem aims to find an orthogonal matrix which transforms one given matrix into another by minimizing their Frobenius matrix norm. This problem can be applied in applications such as permutation theory, machine learning, and camera calibration, *etc*. In real cases, the permutation matrix may have been partially known, the dictionaries can be partially learned from external data, and the calibration of camera should be done under some fixed priors. This prior information makes the original orthogonal Procrustes problem more difficult. In this paper, we consider the solution of this problem under partially known priors, which includes the original orthogonal Procrustes problem as a special case with no such prior.

Key words: orthogonal Procrustes problem, partially known guidance

1. Introduction

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ be two given matrices from observed data, we consider the problem of transforming the matrix $\mathbf{A}$ into $\mathbf{B}$ by an orthogonal matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$ so that the error between the matrix $\mathbf{B}$ and the $\mathbf{TA}$ 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 "$\top$" means transpose operation. The solution of the transformation matrix $\mathbf{T}$ for problem (1) is $\hat{\mathbf{T}} = \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are obtained by performing singular value decomposition (SVD) on $\mathbf{BA}^\top$ via $\mathbf{BA}^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$.

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 to matrices $\mathbf{A}$ and $\mathbf{B}$ of arbitrary rank and further studied the two-sided orthogonal Procrustes problems [4]. Higham [5] studied the symmetric Procrustes problem by replacing the orthogonal restriction on $\mathbf{T}$ with symmetric $\mathbf{T}^\top = \mathbf{T}$. Later, Watson *et al.* studied the solutions of orthogonal Procrustes problems under general matrix norms such as orthogonally invariant norms [6] and the special ℓ_1 norm [7]. The solutions for the generalized problems are not in closed-form

and need projected gradient methods [8] or the famous Newton's method. In [7], the more general weighted orthogonal Procrustes problem is also solved by projected gradient methods [8]. Recently, Berge [9] further discussed the rigid rotation problem which requires the determinant of $\mathbf{T}$ is 1 (i.e., eliminating the possibility of reflection). Recently, Viklands studied the weighted orthogonal Procrustes problem in his/her PhD thesis and local and global solutions are presented therein.

The classical orthogonal Procrustes problem [1] has many applications including factor analysis [2], rigid rotation [9], machine learning [10], computer vision [11, 12], optical imaging , and robotics. The extended two-sided orthogonal Procrustes problem has also been used in matrix perturbation problem [4].

In this paper, we study the solution of the orthogonal Procrustes problem when partial columns of the orthogonal transformation matrix are known. This is possibly emerged in practical applications. For examples, in the point matching problem [12], if some points have already been matched, the matched data can be used to guide the matching of the remaining points. In the machine learning models such as dictionary learning [13], if the external data is employed to guide the learning of the internal data, the dictionary learned will be more adapted to the data being processed. Besides, we proof the sufficiency condition of the solution by using the generalized Kristof's Theorem [14], which is clearly different from the proof in [3].

2. Definition of the Problem and Solution

Assume $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ be two given data matrices. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{P} \in \mathbb{R}^{n \times q}$ $(p + q = n)$, where $\mathbf{X}$ is the partially known guidance employed to guide the solutions of $\mathbf{P}$. From the Section 1, we can see that $[\mathbf{X} \ \mathbf{P}] = \mathbf{T} \in \mathbb{R}^{n \times n}$ is the orthonormal matrix which transform the matrix $\mathbf{A}$ to best fit the matrix $\mathbf{B}$ under the least square sense. Instead of obtaining $\mathbf{T}$ in classical orthogonal Proctustes problem, we need to obtain part of $\mathbf{T}$ with partially known guidance in the guidance version. For simplicity, we assume $n \geq m$ and the other cases can be analyzed in a similar way. We formulate the orthogonal Procrustes problem with partially known priors as:

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

Note that the original orthogonal Procrustes problems is a special case of the problem discussed in this paper. When there is no guidance at all, i.e., $\mathbf{X} = \emptyset$, we have $p = 0, q = n$ and the studied problem is reduced to the classical orthgonal Procrustes problem.

When the partially known guidance $\mathbf{X} \neq \emptyset$, 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

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

The $\mathbf{B} - \mathbf{X}\mathbf{A}_X^\top$ 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 use $\mathbf{B}$ as the finally known data matrix.

Results 1: Let $\mathbf{A} \in \mathbb{R}^{q \times m}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$ be two given data matrices, given partially known prior of $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_{p \times p}$. Then the sufficiency and necessary conditions of

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

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:

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

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 0, we can get

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

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 the Eq. (7) by $\mathbf{X}^\top$, we have

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

Put the results back to Eq. (7), 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 Eq. (10) 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 $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top$ and get $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. Since $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top\mathbf{P}^\top$ is symmetric, we have

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

and hence we have $\mathbf{U} = \mathbf{P}\mathbf{V}$ and equivalently $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$. Note that we can also employ the property of symmetric matrix that every symmetric matrix can be diagonalized to obtain this results. The necessary condition is proofed.

Now we proof the sufficiency condition. If $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$, then $\hat{\mathbf{P}}$ satisfies that $\hat{\mathbf{P}}^\top\hat{\mathbf{P}} = \mathbf{I}_{q \times q}$ and $\mathbf{X}^\top\hat{\mathbf{P}} = \mathbf{0}_{p \times q}$. The first is obvious and now we consider the second one. From the Eq. (4), since $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_{p \times p}$, 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 $\mathbf{X}^\top\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{0}_{p \times p}$ 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

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

Until now, if we want to proof that $\hat{\mathbf{P}} = \mathbf{UV}^\top$ is the solution of problem (3), $\text{Tr}(\mathbf{B}^\top \hat{\mathbf{P}} \mathbf{A})$ 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

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

Now we need to proof that $\text{Tr}(\boldsymbol{\Sigma}) \geq \text{Tr}(\mathbf{B}^\top \mathbf{PA})$ 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 $\text{Tr}(\mathbf{B}^\top \mathbf{PA}) = \text{Tr}((\mathbf{I}_{n \times n} - \mathbf{XX}^\top)\mathbf{BA}^\top \mathbf{P}^\top) = \text{Tr}(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{P}^\top) = \text{Tr}(\boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{P}^\top \mathbf{U}) \leq \text{Tr}(\boldsymbol{\Sigma})$. The last inequality can be obtained by using a generalization version [14] of the Kristof's Theorem [15]. The equality is obtained at $\mathbf{V}^\top \mathbf{P}^\top \mathbf{U} = \mathbf{I}_{q \times q}$, i.e., $\mathbf{P} = \mathbf{UV}^\top = \hat{\mathbf{P}}$. 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., $\hat{\mathbf{P}} = \mathbf{UV}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are the orthogonal matrices obtained by performing economy (a.k.a. reduced) SVD: $\mathbf{BA}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. 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 $\mathbf{XX}^\top$ 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 $\text{rank}(\mathbf{I}_{n \times n} - \mathbf{XX}^\top) \geq n - p$.

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

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

Results 2: If $\text{rank}(\boldsymbol{\Sigma}) = q$ (we call this non-degenerative case), $\boldsymbol{\Sigma}$ may have distinct or multiple non-zero singular values. In the SVD of $(\mathbf{I}_{n \times n} - \mathbf{XX}^\top)\mathbf{BA}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, the singular vectors in $\mathbf{U}$ and $\mathbf{V}$ can be determined up to orientation. Hence, we can reformulate the SVD as

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

where $\mathbf{U}^* \in \mathbb{R}^{n \times q}$ and $\mathbf{V}^* \in \mathbb{R}^{q \times q}$ are arbitrarily oriented 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., $\mathbf{\Sigma}_{11} \geq \mathbf{\Sigma}_{22} \geq ... \geq \mathbf{\Sigma}_{qq} \geq 0$. 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 $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ have multiple singular values also has unique solution of $\hat{\mathbf{P}}$, which can be discussed in a similar way.

If $0 \leq \text{rank}(\mathbf{\Sigma}) = r < q$, there is $q - r$ (at least one) zero singular values and we call this case the degenerative case. 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 can 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

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

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

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

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 $\text{rank}(\mathbf{\Sigma}) < q$ 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

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

[6] 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.

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

[8] 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.

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

[10] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.

[11] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[12] 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.

[13] 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.

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

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