

SOLUTIONS OF ORTHOGONAL PROCRUSTES PROBLEMS UNDER PARTIALLY KNOWN PRIOR

JUN XU

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 priors

1. Introduction

The classical orthogonal Procrustes problem has been applied in psychometrics, multidimensional scaling, factor analysis, machine learning, computer vision, optical imaging, and robotics.

2. Definition of the Problem and Solution

Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times m}$ be two given data matrices. Define $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{P} \in \mathbb{R}^{n \times q}$ where $p + q = n$. \mathbf{X} is the partially known prior which could be used to guide the solutions of \mathbf{P} . We formulate the orthogonal Procrustes problem with partially known priors as:

$$\hat{\mathbf{P}} = \arg \min_{\mathbf{D}} \|\mathbf{B} - [\mathbf{X} \mathbf{P}] \mathbf{A}\|_F^2 \quad s.t. \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}, \mathbf{X}^\top \mathbf{P} = \mathbf{0}_{p \times q}. \quad (1)$$

In fact, as have been proofed, if the matrix $\mathbf{B} \mathbf{A}^\top$ has no zero singular value, then the solution of $\hat{\mathbf{P}} = \mathbf{U} \mathbf{V}^\top$ is unique and we do not need any preceding results.

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

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

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

$$\hat{\mathbf{P}} = \arg \min_{\mathbf{P}} \|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 \quad s.t. \quad \mathbf{P}^\top \mathbf{P} = \mathbf{I}_{q \times q}, \mathbf{X}^\top \mathbf{P} = \mathbf{0}_{p \times q} \quad (3)$$

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 (aka. reduced) SVD:

$$(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \quad (4)$$

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

$$\hat{\mathbf{P}} = \arg \min_{\mathbf{P}} \|\mathbf{B} - \mathbf{P}\mathbf{A}\|_F^2 = \arg \min_{\mathbf{P}} \|\mathbf{B}\|_F^2 + \|\mathbf{P}\mathbf{A}\|_F^2 - 2\text{Tr}(\mathbf{B}^\top \mathbf{P}\mathbf{A}) = \arg \max_{\mathbf{P}} \text{Tr}(\mathbf{A}\mathbf{B}^\top \mathbf{P}). \quad (5)$$

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

$$\mathcal{L} = \text{Tr}(\mathbf{A}\mathbf{B}^\top \mathbf{P}) - \text{Tr}(\Gamma_1(\mathbf{P}^\top \mathbf{P} - \mathbf{I}_{q \times q})) - \text{Tr}(\Gamma_2(\mathbf{P}^\top \mathbf{X})), \quad (6)$$

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

$$\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = \mathbf{B}\mathbf{A}^\top - \mathbf{P}(\Gamma_1 + \Gamma_1^\top) - \mathbf{X}\Gamma_2^\top \quad (7)$$

After setting the equation (7) to zero, we get that

$$\mathbf{B}\mathbf{A}^\top - \mathbf{P}(\Gamma_1 + \Gamma_1^\top) - \mathbf{X}\Gamma_2^\top = 0. \quad (8)$$

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 Equ. (8) by \mathbf{X}^\top , we have

$$\mathbf{X}^\top \mathbf{B}\mathbf{A}^\top = \Gamma_2^\top. \quad (9)$$

Put the results back to Equ. (8), we have

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

Or equivalently,

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

Right multiplying Equ. (11) by \mathbf{P}^\top , we have

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

This shows that $(\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top)\mathbf{B}\mathbf{A}^\top \mathbf{P}^\top$ is a symmertric matrix of order $n \times n$. Then we perfrom 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 symmertric, we have

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

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 symmertric matrix that every symmertric 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 Equ. (4), since $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_{p \times p}$, we have

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

It means that $\mathbf{X}^\top \mathbf{U}\Sigma\mathbf{V}^\top = \mathbf{0}_{p \times p}$ 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

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

Until now, if we want to proof that $\hat{\mathbf{P}} = \mathbf{U}\mathbf{V}^\top$ 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, over all \mathbf{P} satisfying the subject condition in Equ. (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

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

Now we need to proof that $\text{Tr}(\Sigma) \geq \text{Tr}(\mathbf{B}^\top \mathbf{P}\mathbf{A})$ 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 $\text{Tr}(\mathbf{B}^\top \mathbf{P}\mathbf{A}) = \text{Tr}((\mathbf{I}_{n \times n} - \mathbf{X}\mathbf{X}^\top) \mathbf{B}\mathbf{A}^\top \mathbf{P}^\top) = \text{Tr}(\mathbf{U}\Sigma\mathbf{V}^\top \mathbf{P}^\top) = \text{Tr}(\Sigma\mathbf{V}^\top \mathbf{P}^\top \mathbf{U}) \leq \text{Tr}(\Sigma)$. The last inequality can be easily obtained by using the generalization [1] of the Kristof's Theorem [2]. 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.

(Since \mathbf{P} is independent of $\text{Tr}(\Sigma)$, for which to be maximum, we need all the diagonal elements in Σ to be nonnegative. Hence, once the \mathbf{V} is given, the orientation of \mathbf{U} must be chosen to retain the nonnegative diagonal elements in $\text{Tr}(\Sigma)$.)

Now we discuss the uniqueness of the solution \mathbf{P} .

For $\text{Tr}(\mathbf{A}\mathbf{B}^\top \mathbf{P})$

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 perfroming economy (aka. reduced) SVD: $\mathbf{B}\mathbf{A}^\top = \mathbf{U}\Sigma\mathbf{V}^\top$. 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 orgothonal to the external prior $\mathbf{P}^\top \mathbf{X}$.

In this section, we examine the sensitivity of the solution to perturbation in the data. To measure this sensitivity, we give the relative residuals and the Fro-norm condition numbers of the solutions. The condition number of the matrix \mathbf{A} is defined as $k_F(\mathbf{A}) = \frac{\sigma_1}{\sigma_r}$, where $r = \text{rank}(\mathbf{A})$.

FIGURE 1.
North Carolina End-of-Grade Math Skills Test Subscores.

3. Concluding Remarks

References

- [1] Jos M. F. Ten Berge. A generalization of kristof's theorem on the trace of certain matrix products. *Psychometrika*, 48(4):519–523, 1983.
- [2] Walter Kristof. A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, 7(3):515 – 530, 1970.