

**ON THE SYMMETRICAL PROPERTY OF
PROCRUSTES MEASURE OF DISTANCE**

Toni Bakhtiar^{1 §}, Siswadi²

^{1,2}Department of Mathematics

Bogor Agricultural University

Dramaga Campus, Bogor, 16680, INDONESIA

Abstract: Measurement the degree of difference between two matrices by using Procrustes analysis is preceded by a series of Euclidean similarity transformations namely translation, rotation, and dilation, performed in respected order, for gaining maximal matching. It is easy, by a counter example, to show that Procrustes measure does not obey the symmetrical property, something should be owned by any distance function. In this paper we analytically proved that normalization over configuration matrices as an additional transformation results in the satisfaction of the symmetrical property by Procrustes analysis. We also proved that normalization can be undertaken prior to or after rotation to preserve symmetrical property. Moreover, we proved that Procrustes measure can be expressed in term of singular values of the matrix. We here very much exploited the characteristic of full singular value decomposition under similarity transformations.

AMS Subject Classification: 00A35, 62H20

Key Words: procrustes analysis, similarity transformation, singular value decomposition, symmetrical property

1. Procrustes Measure

Shape is all the geometrical information that remains when location, rotational,

Received: November 1, 2014

© 2015 Academic Publications, Ltd.
url: www.acadpubl.eu

[§]Correspondence author

and scaling effects are filtered out from an object. Two configuration matrices as representation of objects have the same shape if they can be translated, rotated, and dilated to each other so that they match exactly [3]. Procrustes analysis refers to a technique of comparing objects with different shapes and producing a measure of the match. It eliminates possible incommensurability of variables within the individual data sets and size differences between data sets by employing data and configuration scalings in calculating the distance, respectively. Thus, a series of Euclidean similarity transformations, i.e., translation-rotation-dilation, is carried-out to best match up to their maximal agreement. Initially, the difference d between matrices $X = (x_{ij})$ and $Y = (y_{ij})$ in $\mathbb{R}^{m \times n}$ can be quantified by

$$d(X, Y) = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - y_{ij})^2 = \text{tr}(X - Y)'(X - Y). \quad (1)$$

We mean by translation a moving process of all elements of matrix within fixed distance and into the same direction with respect to its centroid. We can minimize the distance between X and Y after translation by coinciding their centroids at the origin, without loss of generality. Thus, matrices X and Y after optimal translation are given by

$$X_T = X - C_X, \quad Y_T = Y - C_Y, \quad (2)$$

where $C_X := \frac{1}{m} \mathbf{1}_m \mathbf{1}'_m X$ and $C_Y := \frac{1}{m} \mathbf{1}_m \mathbf{1}'_m Y$ are, respectively, centroids of X and Y , with $\mathbf{1}_m$ is the $m \times 1$ vector of ones.

Rotation is a process of moving all elements of matrix under the fixed rotation angle without any changes in the point-to-centroid distances. Rotation on Y_T over X_T is carried-out by post-multiplying Y_T by a rotation matrix Q . The minimum distance after rotation is obtained by selecting $Q = VU'$, where $U\Sigma V'$ is the full singular value decomposition (FSVD) of $X'_T Y_T$ [2, 5]. It is obvious that Q is orthogonal, i.e., $QQ' = Q'Q = I$.

Dilation is a data scaling which stretches or shrinkages the point-to-centroid distance in a configuration by multiplication of a fixed scaling factor. Dilation on $Y_T Q$ over X_T is undertaken by multiplying $Y_T Q$ by a scalar $c \in \mathbb{R}$, where $c = \frac{\text{tr} X'_T Y_T Q}{\text{tr} Y'_T Y_T Q}$ minimizing the distance between X_T and $Y_T Q$ after dilation. Simultaneously, it is proved in [1] that a series of similarity transformations in the order of translation-rotation-dilation provides the lowest possible distance as stated in the following theorem, where the distance is calculated according to (1) as follows

$$d(X_T, cY_T Q) = \text{tr}(X_T - cY_T Q)'(X_T - cY_T Q). \quad (3)$$

Theorem 1. Given two matrices X and Y in $\mathbb{R}^{m \times n}$, the Procrustes measure p between X and Y after optimal translation-rotation-dilation process is provided by

$$p(X, Y) = \text{tr } X'_T X_T - \frac{\text{tr}^2 X'_T Y_T Q}{\text{tr } Y'_T Y_T}. \tag{4}$$

Proof. For complete proof, see [1].

2. Singular Value Decomposition

In this section we provide a short review on the singular value decomposition, which could be described as the most useful tool in matrix algebra because this factorization can be applied to any real or complex matrix, square or rectangular.

Theorem 2 ([4]). If A is an $m \times n$ matrix of rank $r \leq \min\{m, n\}$, then there exists orthogonal $m \times m$ matrix U (respectively unitary), orthogonal $n \times n$ matrix W (respectively unitary), and $m \times n$ matrix $S = (s_{ij})$ such that

$$A = USW'. \tag{5}$$

Suppose that eigenvalues of AA' can be ordered as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_m = 0$ with the corresponding eigenvectors u_i ($i = 1, \dots, r, r + 1, \dots, m$) and those of $A'A$ are $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0$ with the corresponding eigenvectors w_i ($i = 1, \dots, r, r + 1, \dots, n$). To comply with (5), it is required that either

$$u_i := \frac{Aw_i}{s_{ii}} \quad \text{or} \quad w_i := \frac{A'u_i}{s_{ii}} \tag{6}$$

must be satisfied, with $s_{ii} = \sqrt{\lambda_i}$. The s_{ii} is called the singular value of A , which is defined to be the positive square root of the ranked eigenvalue of AA' or $A'A$ (note that AA' and $A'A$ posses the same set of positive eigenvalues). Then it can be written that

$$U = [u_1 \quad \dots \quad u_r \quad u_{r+1} \quad \dots \quad u_m], \tag{7}$$

$$S = \begin{bmatrix} \Delta_r & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix}, \tag{8}$$

$$W = [w_1 \quad \dots \quad w_r \quad w_{r+1} \quad \dots \quad w_n], \tag{9}$$

where $\Delta_r = \text{diag}(s_{ii}, \dots, s_{rr})$ is an $r \times r$ diagonal matrix and 0 is a matrix of zeros with compatible dimension. Note that we have S in the form of (8) when $r < m$ and $r < n$. If $r = m = n$ then it reduces to $S = \Delta_r$. If $r = m < n$ then $S = (\Delta \ 0)$, and if $r = n < m$ then $S = (\Delta \ 0)'$. Since u_i ($i = 1, \dots, r$) and w_i ($i = 1, \dots, r$) are, respectively, eigenvectors correspond to positive eigenvalues of AA' and $A'A$, then U and W should be constructed by orthonormal columns. In fact, if the number of orthonormal columns is less than m or n then we should expand this set to orthonormal basis. If zero eigenvalues are included along with their corresponding eigenvectors, then we have an FSVD of matrix A as given by Theorem 2.

As an illustrative example of FSVD let consider a 4×3 matrix A of rank 2:

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}.$$

Matrix $A'A$ possesses eigenvalues $\lambda_1 = 12$ and $\lambda_2 = 10$ with corresponding eigenvectors construct the following W matrix:

$$W = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

While, matrix AA' has eigenvalues $\lambda_1 = 12$, $\lambda_2 = 10$, and $\lambda_3 = 0$. According to (6), we obtain

$$u_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 0 \\ 5 \end{bmatrix}.$$

By expanding $\{u_1, u_2\}$ to an orthonormal basis in \mathbb{R}^3 and by exploiting Gram-Schmidt orthonormalization procedure [4] we get

$$u_3 = \frac{1}{\sqrt{30}} \begin{bmatrix} 1 \\ -5 \\ 2 \end{bmatrix}.$$

Thus, the corresponding eigenvectors of AA' form the following U matrix:

$$U = \begin{bmatrix} 1/\sqrt{6} & -2/\sqrt{5} & 1/\sqrt{30} \\ 1/\sqrt{6} & 0 & -5/\sqrt{30} \\ 2/\sqrt{6} & 1/\sqrt{5} & 2/\sqrt{30} \end{bmatrix}.$$

Obviously, we can avoid basis expansion and Gram-Schmidt orthonormalization by initially starting with AA' . But in this way we have to solve the third order characteristic equation.

Further, it is easy to verify that both U and W are orthonormal, i.e., $UU' = U'U = I_3$ and $WW' = W'W = I_2$. Since $m = 4$, $n = 3$, and $r = 2$, then the matrix S is provided by

$$S = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix}.$$

Alternatively, the FSVD with economic size, i.e., reduced order FSVD, can be obtained by removing columns and rows relate to zero eigenvalues, i.e.,

$$A = \begin{bmatrix} 1/\sqrt{6} & -2/\sqrt{5} \\ 1/\sqrt{6} & 0 \\ 2/\sqrt{6} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}'.$$

3. Symmetrical Property

A distance is a function which associates to any pair of vectors a real nonnegative number. The notion of distance is essential because many statistical techniques, such as principal component analysis, multidimensional scaling analysis, and correspondence analysis, are equivalent to the analysis of a specific distance table. A distance function possesses the following properties [4].

Definition 3. For any vectors x, y , and z in vector space \mathbb{S} , a function d is a distance function on \mathbb{S} if the following properties are satisfied:

1. $d(x, y) \geq 0$,
2. $d(x, y) = 0$ if and only if $x = y$,
3. $d(x, y) = d(y, x)$,
4. $d(x, z) \leq d(x, y) + d(y, z)$.

The concept of distance over vectors might be adopted for calculating the degree of difference of matrices of the same dimension. For instance, to any matrices $X = (x_{ij})$ and $Y = (y_{ij})$ in $\mathbb{R}^{m \times n}$, we may defined the difference between them as (1), as embraced by Procrustes measure within translation-rotation-dilation process. However, it is easy to show that Procrustes measure does not comply with distance properties provided in Definition 3. By taking two different matrices

$$X = \begin{bmatrix} 3 & 0 \\ 5 & 4 \\ 1 & 4 \end{bmatrix}, \quad Y = \begin{bmatrix} 3 & 5 \\ 2 & 3 \\ 4 & 3 \end{bmatrix}, \tag{10}$$

from (4) we have $p(X, Y) = 0$, which reveals a violation of the second property. This fact, however, provides another kind of idiosyncrasies of matrix algebra, in addition to the well-known fact that $AB = 0$ can happen although neither A nor B is itself a zero matrix. Another pair of matrices

$$X = \begin{bmatrix} -3 & 0 \\ 5 & 4 \\ 1 & 4 \end{bmatrix}, \quad Y = \begin{bmatrix} 3 & 5 \\ 2 & 3 \\ 4 & -5 \end{bmatrix}, \quad (11)$$

produces $p(X, Y) = 34.67$ and $p(Y, X) = 47.13$, which shows that in general Procrustes measure disobeys the third property, i.e., the symmetrical property.

This paper deals with the embedding of the symmetrical property of Procrustes measure. Its primary objective is to ascertain that an additional transformation, namely a normalization, can set the symmetrical property into Procrustes measure. In fact, normalization is already considered by MATLAB[®] built-in function `procrustes(X, Y)`. Pre-scaling to normal matrices are also employed by [3] in calculating Procrustes distance by using shape analysis approach. We here provide the proof of the symmetrical property and discuss the effect of normalization place in order of the transformation sequences. We greatly exploit the characteristic of FSVD under similarity transformations.

Let given two data matrices X and Y with the same number of objects. If the number of variables of both matrices is not the same, we may place a number of columns of zeros either in the last or anywhere such that the dimension of matrices are the same [6]. Thus, without loss of generality we assume that X and Y are in the same space, i.e., $X, Y \in \mathbb{R}^{m \times n}$.

In this analysis, we mean by normalization making the norm of matrix one by dividing it by its Frobenius norm. We introduce a normalization procedure following a translation. Hence, the normalization over X_T and Y_T produces

$$\bar{X}_T = aX_T, \quad \bar{Y}_T = bY_T, \quad (12)$$

where

$$a := \frac{1}{\|X_T\|_F} = \frac{1}{\sqrt{\text{tr } X_T' X_T}}, \quad b := \frac{1}{\|Y_T\|_F} = \frac{1}{\sqrt{\text{tr } Y_T' Y_T}}, \quad (13)$$

with X_T and Y_T are given by (2). In (13) we denote by $\|\cdot\|_F$ the Frobenius norm of a matrix, hence $\|\bar{X}_T\|_F = \|\bar{Y}_T\|_F = 1$.

Theorem 4. *Given two matrices X and Y in $\mathbb{R}^{m \times n}$, the Procrustes measure p between X and Y after optimal translation-normalization-rotation-dilation process obeys the symmetrical property and is provided by*

$$p(X, Y) = p(Y, X) = 1 - \left(\sum_{i=1}^r \sigma_{ii} \right)^2, \quad (14)$$

where r and σ_{ii} ($i = 1, \dots, r$) respectively are the rank and singular values of $\bar{X}'_T \bar{Y}_T$ or $\bar{Y}'_T \bar{X}_T$ with \bar{X}_T and \bar{Y}_T are given in (12).

Proof. By performing optimal rotation over \bar{X}_T and \bar{Y}_T and subsequently optimal dilation, then by considering (1) we obtain the Procrustes measure of X and Y as follows

$$p(X, Y) = \inf_c d(\bar{X}_T, c\bar{Y}_T Q_1) = \inf_c \text{tr}(\bar{X}_T - c\bar{Y}_T Q_1)'(\bar{X}_T - c\bar{Y}_T Q_1),$$

where $Q_1 = V_1 U_1'$ and $U_1 \Sigma_1 V_1'$ is the FSVD of $\bar{X}'_T \bar{Y}_T$. We can expand the above equation to get a quadratic form in c :

$$p(X, Y) = \inf_c (\text{tr} \bar{X}'_T \bar{X}_T - 2c \text{tr}(\bar{X}'_T \bar{Y}_T Q_1) + c^2 \text{tr} \bar{Y}'_T \bar{Y}_T). \tag{15}$$

By optimally selecting

$$c = \frac{\text{tr}(\bar{X}'_T \bar{Y}_T Q_1)}{\text{tr} \bar{Y}'_T \bar{Y}_T} \tag{16}$$

and substituting it back into quadratic form in (15) alongside with (12)–(13) we have

$$p(X, Y) = 1 - \text{tr}^2(\bar{X}'_T \bar{Y}_T Q_1) = 1 - \text{tr}^2 \Sigma_1. \tag{17}$$

Note that the last expression is obtained by remembering that $Q_1 = V_1 U_1'$ and $\bar{X}'_T \bar{Y}_T = U_1 \Sigma_1 V_1'$. We now consider the opposite direction, where the matrix X should be fitted by similarity transformations for obtaining the best matching to the observed matrix Y . In other words, we aim to calculate the Procrustes measure $p(Y, X)$. By following the similar technique, we may shortly arrive at

$$p(Y, X) = 1 - \text{tr}^2(\bar{Y}'_T \bar{X}_T Q_2) = 1 - \text{tr}^2 \Sigma_2, \tag{18}$$

where $Q_2 = V_2 U_2'$ and $U_2 \Sigma_2 V_2'$ is the FSVD of $\bar{Y}'_T \bar{X}_T$. Since for any matrix A , the spectrum of $A'A$ is the same as that of AA' , then

$$\text{tr} \Sigma_1 = \text{tr} \Sigma_2 = \sum_{i=1}^r \sigma_{ii},$$

where r and σ_{ii} ($i = 1, \dots, r$) respectively are the rank and singular values of $\bar{X}'_T \bar{Y}_T$ or $\bar{Y}'_T \bar{X}_T$. Thus, we proved the symmetrical property of Procrustes measure, i.e., $p(X, Y) = p(Y, X) = 1 - (\sum_{i=1}^r \sigma_{ii})^2$.

Remark 5. Procrustes measure p in (14) is basically calculated by using (1). Since $\text{tr } A'A \geq 0$ for any matrix A , then it is guaranteed that the nonnegativity property of a distance function is satisfied, i.e., $p(X, Y) \geq 0$. Furthermore, suppose that there exists matrices X and Y such that $p(X, Y) > 1$. Then we have $1 - (\sum_{i=1}^r \sigma_{ii})^2 > 1$, or equivalently $(\sum_{i=1}^r \sigma_{ii})^2 < 0$, which is a contradiction. It must be the case that $p(X, Y) \leq 1$. As a consequence of these facts, we conclude $0 \leq p(X, Y) \leq 1$ for any matrices X and Y of the same dimension. Result presented in Theorem 4 is in agreement with that of [3].

Remark 6. Note that the optimal scaling parameter c in (16) can further be written as

$$c = \sum_{i=1}^r \sigma_{ii}.$$

Remark 7. Suppose that we have the following FSVD: $X'_T Y_T = U_3 \Sigma_3 V'_3$. Together with (17) we obtain

$$\begin{aligned} \bar{X}'_T \bar{Y}_T = U_1 \Sigma_1 V'_1 &\Leftrightarrow ab X'_T Y_T = U_1 \Sigma_1 V'_1 \\ &\Leftrightarrow X'_T Y_T = U_1 \frac{\Sigma_1}{ab} V'_1, \end{aligned} \tag{19}$$

from which we conclude $U_1 = U_3$, $V_1 = V_3$, $\Sigma_1 = ab \Sigma_3$, and furthermore we have the relation between singular values $\sigma_{ii} = ab \delta_{ii}$, where δ_{ii} is the singular value of $X'_T Y_T$. Thus we can alternatively rewrite (14) as

$$p(X, Y) = p(Y, X) = 1 - a^2 b^2 \left(\sum_{i=1}^r \delta_{ii} \right)^2.$$

Remark 8. By (14), if we have $p(X, Y) = 0$ then it means that X and Y have exactly the same shape. Conversely, X and Y with $p(X, Y) = 1$ have entirely different shape. In this regard we call p a dissimilarity measure. If we would like to have Procrustes measure ρ with opposite interpretation, then we define $\rho(X, Y) := 1 - p(X, Y) = (\sum_{i=1}^r \sigma_{ii})^2$. It means that the bigger the coefficient ρ , the closer the similarity of matrices X and Y . Thus we call ρ a similarity measure which can conduct as a goodness-of-fit of the best matching.

4. When Should We Normalize?

In the previous section we have analyzed the incorporating normalization as an additional similarity transformation into the procedure. It is shown that

by normalizing the matrices after translation, the symmetrical property is asserted by Procrustes measure. In this section we examine the situation where normalization is performed sequentially different. As an illustrative example, let consider again matrices X and Y in (11). Table below provides Procrustes measure under different sequence of transformations, e.g., we mean by T-N-R-D the degree of difference between matrices is calculated after translation (T), normalization (N), rotation (R), and dilation (D), respectively.

No	Transformations	$p(X, Y)$	$p(Y, X)$
1	N-T-R-D	0.5174	0.5355
2	T-N-R-D	0.8125	0.8125
3	T-R-N-D	0.8125	0.8125
4	T-R-D-N	1.1340	1.1340

It is shown that normalizing prior to rotation (T-N-R-D) or after rotation (T-R-N-D) provides a consistent result and guarantees the symmetrical property. It can generally be explained as follows. After translation and rotation we have X_T and $Y_T Q_3$, where $Q_3 = V_3 U_3'$ and $U_3 \Sigma_3 V_3'$ is the FSVD of $X_T' Y_T$. Normalizing both matrices provides $\bar{X}_T = a X_T$ as in (12) and $\bar{Z}_T := \frac{Y_T Q_3}{\|Y_T Q_3\|_F} = b Y_T Q_3$. The procrustes measure after optimal dilation is achieved by selecting $c = \frac{\text{tr} \bar{X}_T' \bar{Z}_T}{\text{tr} Z_T' Z_T}$ with

$$p(X, Y) = 1 - \text{tr}^2 \bar{X}_T' \bar{Z}_T = 1 - \text{tr}^2 (ab \Sigma_3) = 1 - \text{tr}^2 \Sigma_1, \tag{20}$$

where the last expression is came from (19). Equation (20), however, proves that Procrustes measures under T-N-R-D and T-R-N-D are the same. Obviously, we can also prove the symmetrical property similarly.

From the table we can also see that when normalization is applied in the first stage, i.e., we consider N-T-R-D, then the symmetrical property is generally not satisfied, even though this sequence may provide a smaller measure. While, when we put normalization in the last stage, i.e., we consider T-R-D-N, then we get a different magnitude of measure with the symmetrical property convinced. However, the measure is not always staying in $[0, 1]$ interval, something which is not expected for a goodness of fit coefficient. In fact, Procrustes measure under T-R-D-N scheme is given by

$$p(X, Y) = \text{tr}(a X_T - b Y_T Q)'(a X_T - b Y_T Q) = 2(1 - ab \text{tr} \Sigma). \tag{21}$$

5. Conclusion

In this paper we have shortly reviewed the optimal similarity transformations of Procrustes analysis. It has been shown that by performing translation, rotation, and dilation consecutively over configuration matrices, then their difference is minimized. However, by simple example it has also been revealed that Procrustes measure violates the symmetrical property. To overcome the problem, we propose the inclusion of normalization, that is a division by its Frobenius norm, as an additional similarity transformation. We analytically proved that Procrustes measure can assert the symmetrical property as long as normalization procedure is accomplished prior to or after rotation. By this we also guarantee that the measure belongs to the interval of $[0, 1]$. In particular, we have shown that the goodness-of-fit between X and Y can be expressed as the squared sum of singular values of $\bar{X}'_T \bar{Y}_T$.

References

- [1] T. Bakhtiar, Siswadi, Orthogonal procrustes analysis: its transformation arrangement and minimal distance, *Int. J. App. Math. Stat.*, **20**, No. M11 (2011), 16-24, **doi:** 10.12988/ams.
- [2] J.M.F. ten Berge, Orthogonal procrustes rotation for two or more matrices, *Psychometrika*, **42**, No. 2 (1977), 267–276, **doi:** 10.1007/BF02294053.
- [3] I.L. Dryden, K.V. Mardia, *Statistical Shape Analysis*, John Wiley & Sons, New York (1998).
- [4] J.R. Schott, *Matrix Analysis for Statistics*, John Wiley & Sons, New Jersey (2005).
- [5] R. Sibson, Studies in the robustness of multidimensional scaling: procrustes statistics, *J. R. Statist. Soc. B*, **40**, No. 2 (1978), 234-238.
- [6] Siswadi, T. Bakhtiar, R. Maharsi, Procrustes analysis and the goodness-of-fit of biplots: some thoughts and findings, *App. Math. Sci.*, **6**, No. 72 (2012), 3579-3590, **doi:** 10.12988/ams.