Chapter 13
Rotation Techniques

In some analysis procedures, the solution for a data set is *not uniquely determined*; multiple solutions exist. An example of such procedures is exploratory factor analysis (EFA). In this procedure, one of the solutions is first found, and then it is transformed into a useful solution that is included in multiple solutions. A family of such transformations is the *rotation* treated in this section. The rotation for EFA solutions in particular is called *factor rotation*, although the rotation can be used for solutions of procedures other than EFA. This chapter starts with illustrating why the term "rotation" is used, before explaining which solutions are useful in Sect. 13.3. This is followed by the introduction of some rotation techniques.

13.1 Geometric Illustration of Factor Rotation

As discussed with (12.16) in Sect. 12.5, when loading matrix $\hat{\mathbf{A}}$ is an EFA solution of a loading matrix, its transformed version,

$$\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}'^{-1}, \tag{13.1}$$

is *also a solution*. Here, $\mathbf{T}$ is an $m \times m$ matrix that satisfies (12.14), which is written again here:

$$\mathbf{T}'\mathbf{T} = \begin{bmatrix} 1 & & \# \\ & \ddots & \\ \# & & 1 \end{bmatrix}, \text{ or equivalently, } \quad \mathrm{diag}(\mathbf{T}'\mathbf{T}) = \mathbf{I}_m. \tag{13.2}$$

where diag() is defined in Note 12.1. In this section, we geometrically illustrate the transformation of $\hat{\mathbf{A}}$ into $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$, supposing that $\mathbf{T}$ is given.

K. Adachi, *Matrix-Based Introduction to Multivariate Data Analysis*,
https://doi.org/10.1007/978-981-15-4103-2_13

Let us use $\mathbf{a}_j'$ for the jth row of the original matrix $\hat{\mathbf{A}}$ and use $\mathbf{a}_j^{(\mathrm{T})\prime}$ for that of the transformed $\mathbf{A}_\mathrm{T}$. Then, $\mathbf{A}_\mathrm{T} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ is rewritten as

$$\mathbf{a}_j^{(\mathrm{T})\prime} = \mathbf{a}_i'\mathbf{T}'^{-1} \quad (j = 1, \ldots, p). \tag{13.3}$$

Post-multiplying both sides of (13.3) by $\mathbf{T}'$ leads to $\mathbf{a}_j^{(\mathrm{T})}{}'\mathbf{T}' = \mathbf{a}_j'$, i.e.,

$$\mathbf{a}_j' = \mathbf{a}_j^{(\mathrm{T})}{}'\mathbf{T}' \quad (j = 1, \ldots, p), \tag{13.4}$$

which shows that the original loading vector $\mathbf{a}_j'$ for variable j is expressed by the post-multiplication of the transformed $\mathbf{a}_j^{(\mathrm{T})\prime}$ by $\mathbf{T}'$. We suppose $m = 2$ and define the columns of $\mathbf{T}$ as

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2], \text{with} \quad \|\mathbf{t}_1\| = \|\mathbf{t}_2\| = 1 \tag{13.5}$$

which satisfies (13.2). Using (13.5) and $\mathbf{a}_j^{(\mathrm{T})\prime} = [a_{j1}^{(\mathrm{T})}, a_{j2}^{(\mathrm{T})}]$, (13.4) is rewritten as

$$\mathbf{a}_j' = a_{j1}^{(\mathrm{T})}\mathbf{t}_1' + a_{j2}^{(\mathrm{T})}\mathbf{t}_1'. \tag{13.6}$$

It shows that the *original loading vector* for variable j is equal to the *sum* of $\mathbf{t}_k$ (k = 1, 2) *multiplied* by the *transformed loadings*. Its geometric implications are illustrated in the next two paragraphs.

In Table 13.1(A), we again show the original loading matrix $\hat{\mathbf{A}}$ in Table 12.1(A) obtained by EFA. Its row vectors $\mathbf{a}_j'$ (j = 1, …, 8) corresponding to variables are shown in Fig. 13.1a; the vector $\mathbf{a}_7'$ for H is depicted by the line extending to [−0.63, 0.46], and the other vectors are done in parallel manners. Now, let us consider transforming $\hat{\mathbf{A}}$ into $\mathbf{A}_\mathrm{T} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ by

Table 13.1 A solution obtained with EFA (Table 12.1A) and an example of its rotated version

	(A) Before rotation			(B) After rotation		
	$\hat{\mathbf{A}}$		ψ_j	$\mathbf{A}_\mathrm{T}$		ψ_j
A	0.77	−0.38	0.26	1.03	−0.76	0.26
C	0.61	0.50	0.38	0.56	0.32	0.38
I	0.67	−0.36	0.41	0.90	−0.69	0.41
B	−0.74	−0.40	0.30	−0.75	−0.15	0.30
T	0.79	0.43	0.18	0.80	0.16	0.18
V	0.76	−0.44	0.22	1.04	−0.82	0.22
H	−0.63	0.46	0.39	−0.89	0.79	0.39
P	0.70	0.18	0.47	0.77	−0.09	0.47
ϕ_{12}	0.00			0.57		

$$\mathbf{T}'^{-1} = \begin{bmatrix} 10.18 & -0.42 \\ -0.32 & 1.14 \end{bmatrix}, \text{following from } \mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2] = \begin{bmatrix} 0.94 & 0.26 \\ 0.34 & 0.97 \end{bmatrix}. \quad (13.7)$$

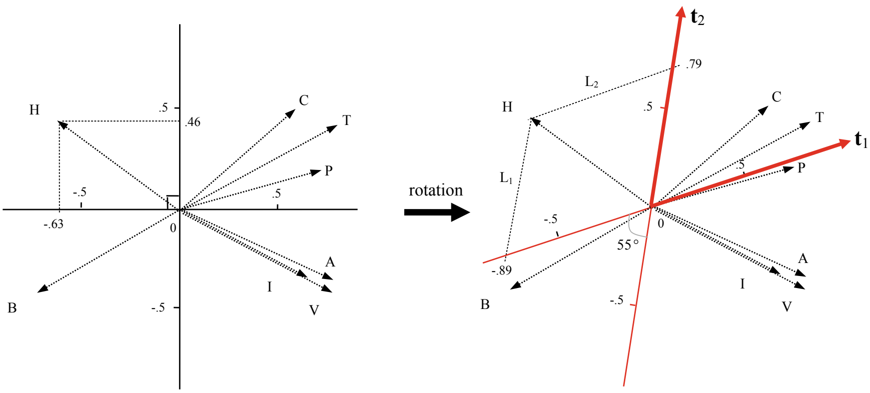
This $\mathbf{T}'^{-1}$ leads to $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ in Table 13.1(B). There, we find that the vector for H is $\mathbf{a}_7^{(T)\prime} = \mathbf{a}_7'\mathbf{T}'^{-1} = [-0.89, 0.79]$, transformed from $\mathbf{a}_7' = [-0.63, 0.46]$ in (A). Those two vectors satisfy the relationship in (13.6):

$$[-0.63, 0.46] = -0.89\mathbf{t}_1' + 0.79\mathbf{t}_2', \quad (13.8)$$

with $\mathbf{t}_1' = [0.94, 0.34]$ and $\mathbf{t}_2' = [0.26, 0.97]$.

The geometric implication of (13.8), which is an example of (13.6), is illustrated in Fig. 13.1b. There, the axes extending in the directions of $\mathbf{t}_1' = [0.94,\ 0.34]$, $\mathbf{t}_2' = [0.26,\ 0.98]$ are depicted, together with the original loading vectors $\mathbf{a}_1', \ldots, \mathbf{a}_8'$ whose locations are the same as in (A). Let us note that vector $\mathbf{a}_7'$ for H satisfies (13.8); i.e., the -0.89 times of $\mathbf{t}_1'$ plus the 0.79 times of $\mathbf{t}_2'$ is equivalent to $\mathbf{a}_7' = [-0.63,\ 0.64]$. Here, the transformed loadings -0.89 and 0.79 can be viewed as the coordinates of point H on $\mathbf{t}_1$ and $\mathbf{t}_2$ axes, as shown by the dotted lines L_1 and L_2 in Fig. 13.1b, where L_1 and L_2 extend in *parallel* to $\mathbf{t}_2$ and $\mathbf{t}_1$, respectively. This relationship holds for the other loading vectors.

In summary, transformation (13.1) implies the rotation of the original horizontal and vertical axes in Fig. 13.1a to the *new axes* extending in the direction of the column vectors of $\mathbf{T}$ as in Fig. 13.1b, where the transformed loadings are the *coordinates* on the new axes. The reason why (13.1) is called *rotation* is found above.



(a) Variable vectors for Table 13.1(A)

(b) Rotated axes, the coordinates which give the loadings in Table 13.1(B)

Fig. 13.1 Illustration of rotation as that of axes

13.2 Oblique and Orthogonal Rotation

Rotation is classified into oblique and orthogonal. The transformation illustrated in the last section is *oblique rotation*, since the new axes are intersected obliquely, as in Fig. 13.1b. On the other hand, *orthogonal rotation* refers to the rotation of *axes* by keeping their *orthogonal intersection*, whose example is described later in Fig. 13.2a. In orthogonal rotation, constraint (13.2) is strengthened so that it is the $m \times m$ identity matrix:

$$\mathbf{T}'\mathbf{T} = \mathbf{I}_m. \tag{13.9}$$

The matrix $\mathbf{T}$ satisfying (13.9) is said to be *orthonormal,* and its properties are detailed in Appendix A.1.2. Customarily, the rotation made by orthonormal $\mathbf{T}$ is not called orthonormal rotation, but rather *orthogonal rotation*. Using (13.9), (13.1) is simplified as

$$\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T} \tag{13.10}$$

in orthogonal rotation.

In summary, rotation is classified into two types:

[1] *Oblique* rotation (13.1) with $\mathbf{T}$ constrained as (13.2)
[2] *Orthogonal* rotation (13.10) with $\mathbf{T}$ constrained as (13.9)

Orthogonal rotation can be viewed as a special case of oblique rotation in which (13.2) is strengthened as (13.9).

13.3 Rotation to Simple Structure

The transformed loading matrix in Table 13.1(B) is not a useful one. That matrix is merely an example for illustrating rotation. A "good rotation procedure" is one that gives a useful matrix. Here, we have the question: "What matrix is *useful*?" A variety of answers exist; which answer is right varies from case to case.

When a matrix is a variables × factors loading matrix, usefulness can be defined as "*interpretability*", i.e., being easily interpreted. What matrix is interpretable? An ideal example is shown in Table 13.2(A), where # indicates a nonzero (positive or negative) value. This matrix has two features:

[1] *Sparse*, i.e., a number of elements are *zero*
[2] *Well classified*, i.e., different variables load different factors

Feature [1] allows us to focus on the nonzero elements to capture the relationships of variables to factors. Feature [2] clarifies the differences between factors. The matrix in Table 13.2(A) is said to have a *simple structure* (Thurstone, 1947).

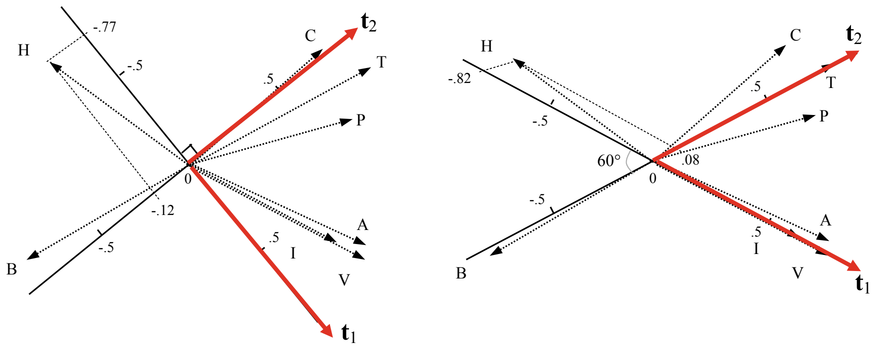
Table 13.2(A) shows an ideally simple structure, but it is almost impossible to have such a matrix; $\mathbf{T}$ cannot be chosen so that some elements of $\mathbf{A}_T = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ are exactly zero as in (A). However, it is feasible to obtain $\mathbf{A}_T = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ that approximates the ideal. It is illustrated in Table 13.2(B). There, "Small" stands for a value *close to zero*, but not exactly being zero, while "Large" expresses a value with a large absolute value. A matrix, which is not ideal but *approximates ideal structure*, is also said to have a *simple structure* in the literature for psychometrics (statistics for psychology).

Let us remember that $\mathbf{A}_T = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ can be viewed as the coordinates on rotated axes. How should the axes be rotated so as to make the loading matrix $\mathbf{A}_T$ be of a simple structure? One answer is found in Fig. 13.2, where the useful orthogonal and oblique rotation for the variable vectors in Fig. 13.1a is illustrated. First, let us note the axes of $\mathbf{t}_1$ and $\mathbf{t}_2$ in Fig. 13.2b. The former axis is approximately *parallel* to the vectors for a group of variables {A, V, I, H} (Group 1), while the latter is almost *parallel* to those for another group {C, T, B, P} (Group 2). Thus, Group 1 has the coordinates of large absolute values on the $\mathbf{t}_1$ axis, but shows those of small absolutes on the $\mathbf{t}_2$ axis. On the other hand, Group 2 shows the coordinates of large and small absolutes for $\mathbf{t}_2$ and $\mathbf{t}_1$ axes, respectively. The resulting loading matrix is presented in Table 13.3(C). There, the matrix successfully attains the simple structure as in Table 13.2(B). Orthogonal rotation is illustrated in Fig. 13.2a, where $\mathbf{t}_1$ and $\mathbf{t}_2$ are orthogonally intersected; (13.9) is satisfied. On the other hand, the axes are obliquely intersected in Fig. 13.2b. Also in (A), the $\mathbf{t}_1$ and $\mathbf{t}_2$ axes are almost parallel to Groups 1 and 2, respectively, which provides the matrix having a simple structure in Table 13.3(B).

In the above paragraph, we visually illustrated how $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2]$ is set to be parallel to groups of variable vectors so that $\mathbf{A}_T = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ has a simple structure. But, this task can only be attained by human vision and is impossible even by that when m exceeds three-dimensions. Indeed, the optimal $\mathbf{T}$ is obtained not visually but *computationally* with

Table 13.2 Simple structure in a matrix of variables × factors

(A) Ideally simple			(B) Simple	
Variable	F1	F2	F1	F2
1	#	0	**Large**	Small
2	0	#	Small	**Large**
3	0	#	Small	**Large**
4	#	0	**Large**	Small
5	0	#	Small	**Large**
6	#	0	**Large**	Small
7	#	0	**Large**	Small
8	0	#	Small	**Large**



(a) Orthogonal Rotation (b) Oblique Rotation

Fig. 13.2 Illustrations of rotation to a simple structure

Table 13.3 A solution obtained with EFA (Table 12.1A) and its rotated versions

	(A) Before rotation			(B) After varimax rotation			(C) After geomin rotation		
	$\mathbf{A}$		ψ_j	$\mathbf{A}_\mathrm{T}$		ψ_j	$\mathbf{A}_\mathrm{T}$		ψ_j
A	**0.77**	−0.38	0.26	**0.81**	0.28	0.26	**0.82**	0.08	0.26
C	**0.61**	0.50	0.38	0.07	**0.78**	0.38	−0.13	**0.84**	0.38
I	**0.67**	−0.36	0.41	**0.73**	0.22	0.41	**0.74**	0.04	0.41
B	**−0.74**	−0.40	0.30	−0.24	**−0.80**	0.30	−0.04	−0.82	0.30
T	**0.79**	0.43	0.18	0.25	**0.87**	0.18	0.04	0.88	0.18
V	**0.76**	−0.44	0.22	**0.85**	0.23	0.22	**0.87**	0.01	0.22
H	**−0.63**	0.46	0.39	**−0.77**	−0.12	0.39	**−0.82**	0.08	0.39
P	**0.70**	0.18	0.47	0.37	**0.63**	0.47	0.23	**0.58**	0.47
ϕ_{12}	0.00			0.00			0.48		

$$\text{maximize Simp}(\mathbf{A}_\mathrm{T}) = \text{Simp}(\hat{\mathbf{A}}\mathbf{T}'^{-1}) \text{ over } \mathbf{T} \text{ subject to (13.2) or (13.9).} \quad (13.11)$$

Here, $\text{Simp}(\hat{\mathbf{A}}\mathbf{T}'^{-1})$ is the abbreviation for the *simplicity* of $\hat{\mathbf{A}}\mathbf{T}'^{-1}$ and is a function of $\mathbf{T}$ that stands for how well $\mathbf{A}_\mathrm{T} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ approximates the ideal simple structure, that is, how simple the structure in $\mathbf{A}_\mathrm{T}$ is. The procedures formulated as (13.11) are generally called (algebraic) *rotation techniques*. In exactness, we should call them *simple structure rotation techniques* in order to distinguish them from the rotation that does not involve a simple structure. A number of simple structure rotation techniques have been proposed so far, which differ in terms of how to define $\text{Simp}(\mathbf{A}_\mathrm{T}) = \text{Simp}(\hat{\mathbf{A}}\mathbf{T}'^{-1})$. Two popular techniques are introduced in the next two sections.

13.4 Varimax Rotation

The rotation techniques with (13.9) chosen as the constraint in (13.11) are called *orthogonal* rotation techniques. Among them, the *varimax* rotation method presented by Kaiser (1958) is well known. In this method, the simplicity of $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}$ is defined as

$$\mathrm{Simp}(\mathbf{A}_{\mathrm{T}}) = \mathrm{Simp}(\hat{\mathbf{A}}\mathbf{T}) = \sum_{k=1}^{m} \mathrm{var}\left(a_{1k}^{(\mathrm{T})\,2} \cdots a_{pk}^{(\mathrm{T})\,2}\right) \tag{13.12}$$

to be maximized. Here, we have used the fact that (13.1) is simplified as (13.10) and $\mathrm{var}(a_{1k}^{(T)\,2} \cdots a_{pk}^{(\mathrm{T})\,2})$ stands for the *variance of the squared elements* in the kth column of $\mathbf{A}_{\mathrm{T}} = (a_{jk}^{(\mathrm{T})})$:

$$\mathrm{var}\left(a_{1k}^{(\mathrm{T})\,2} \cdots a_{pk}^{(\mathrm{T})\,2}\right) = \frac{1}{p}\sum_{j=1}^{p}\left(a_{jk}^{(\mathrm{T})\,2} - \frac{1}{p}\sum_{l=1}^{p} a_{lk}^{(\mathrm{T})\,2}\right)^2. \tag{13.13}$$

That is, the varimax rotation is formulated as

$$\text{maximize}\,\mathrm{simp}(\hat{\mathbf{A}}\mathbf{T}) = \frac{1}{p}\sum_{k=1}^{m}\sum_{j=1}^{p}\left(a_{jk}^{(\mathrm{T})2} - \frac{1}{p}\sum_{l=1}^{p} a_{lk}^{(\mathrm{T})2}\right)^2 \text{ over } \mathbf{T} \text{ subject } \mathbf{T}'\mathbf{T} = \mathbf{I}_m. \tag{13.14}$$

For this maximization, an iterative algorithm is needed. One of the algorithms can be included in the gradient methods introduced in Appendix A.6.3 (Jennrich, 2001). However, that is out of the scope of this book.

We should note that variance (13.13) is not defined for loadings $a_{jk}^{(\mathrm{T})}$ but for its squares $a_{jk}^{(\mathrm{T})}$; they are irrelevant to whether $a_{jk}^{(\mathrm{T})}$ are positive or negative, but are relevant to the absolute values of $a_{jk}^{(\mathrm{T})}$. If variance (13.13) is larger, the *absolute values* of the loadings in each column of $\mathbf{A}_{\mathrm{T}}$ would take a *variety* of values so that

$$\text{some absolute values are larger, but others are small,} \tag{13.15}$$

as illustrated in Table 13.2(B).

The sum of the above variances over m columns defines the simplicity as in (13.12). By maximizing the sum, all m columns can have loadings with (13.15). This allows us to consider the two different $\mathbf{A}_{\mathrm{T}}$ results illustrated in Table 13.4(A) and (B). There, we find that (A) is equivalent to Table 13.2(B); i.e., it shows a simple structure, while Table 13.4(B) is not simple, in that the same variables heavily load two factors. However, (13.14) hardly provides a loading matrix $\mathbf{A}_{\mathrm{T}}$, as

Table 13.4 Variables × factors matrices with and without a simple structure

Variable	(A) Simple		(B) Not simple	
	F1	F2	F1	F2
1	**Large**	Small	**Large**	**Large**
2	Small	**Large**	Small	Small
3	Small	**Large**	Small	Small
4	**Large**	Small	**Large**	**Large**
5	Small	**Large**	Small	Small
6	**Large**	Small	**Large**	**Large**
7	**Large**	Small	**Large**	**Large**
8	Small	**Large**	Small	Small

in Table 13.4(B), since it necessitates $\mathbf{t}_1$ and $\mathbf{t}_2$ extending almost in parallel, which contradicts constraint (13.9).

The varimax rotation for loading matrix $\hat{\mathbf{A}}$ in Table 13.3(A) provides the rotation matrix

$$\mathbf{T} = \begin{bmatrix} 0.705 & 0.710 \\ -0.711 & 0.704 \end{bmatrix}, \tag{13.16}$$

which is the solution for (13.14). Post-multiplication of $\hat{\mathbf{A}}$ in Table 13.3(A) by (13.16) yields the matrix $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}$ in Table 13.3(B) that shows a simple structure. Indeed, Fig. 13.2a has been depicted according to (13.16).

Let us compare $\hat{\mathbf{A}}$ in Table 13.3(A) and $\mathbf{A}_{\mathrm{T}}$ in (B). It is difficult to reasonably interpret the former loadings in (A), as all variables show the loadings of large absolute values for Factor 1 and those of rather small absolutes for Factor 2. It obliges one to consider that Factor 1 explains all variables, while Factor 2 is irrelevant to all variables, which implies that Factor 2 is trivial. On the other hand, $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}$ can be reasonably interpreted in the same manner as described in Sect. 12.7.

13.5 Geomin Rotation

The phrase "*maximize* Simp($\mathbf{A}_{\mathrm{T}}$)" in (13.11) is equivalent to "*minimize* $-1 \times$ Simp ($\mathbf{A}_{\mathrm{T}}$)". Here, $-1 \times$ Simp($\mathbf{A}_{\mathrm{T}}$) can be rewritten as Comp($\mathbf{A}_{\mathrm{T}}$) which abbreviates the *complexity* of $\mathbf{A}_{\mathrm{T}}$ and represents to what extent $\mathbf{A}_{\mathrm{T}}$ deviates from a simple structure. Some rotation techniques are formulated as substituting "minimize Comp($\mathbf{A}_{\mathrm{T}}$)" for "maximize Simp($\mathbf{A}_{\mathrm{T}}$)" in (13.11). One of them is Yates's (1987) *geomin* rotation method, in which complexity is defined as

$$\mathrm{Comp}(\mathbf{A}_{\mathrm{T}}) = \mathrm{Comp}(\hat{\mathbf{A}}\mathbf{T}'^{-1}) = \sum_{j=1}^{p}\left\{\prod_{k=1}^{m}\left(a_{jk}^{(\mathrm{T})\,2} + \varepsilon\right)\right\}^{1/m}, \tag{13.17}$$

with ε a specified small positive value such as 0.01. The geomin rotation method has orthogonal and oblique versions. In this section, we treat the latter, i.e., the *oblique geomin* rotation, which is formulated as

$$\text{minimize } \mathrm{Comp}(\hat{\mathbf{A}}\mathbf{T}'^{-1}) = \sum_{j=1}^{p}\left\{\prod_{k=1}^{m}(a_{jk}^{(\mathrm{T})\,2} + \varepsilon)\right\}^{1/m} \text{ over } \mathbf{T} \text{ subject to (13.2)}. \tag{13.18}$$

For this minimization, an iterative algorithm is needed. One of the algorithms can be included in the gradient methods introduced in Appendix A.6.3 (Jennrich, 2002). However, that is beyond the scope of this book.

Let us note the parenthesized part in the right-hand side of (13.17):

$$\prod_{k=1}^{m}\left(a_{jk}^{(\mathrm{T})2} + \varepsilon\right) = \left(a_{j1}^{(\mathrm{T})2} + \varepsilon\right) \times \cdots \times \left(a_{jm}^{(\mathrm{T})2} + \varepsilon\right). \tag{13.19}$$

It is close to zero, if some of $a_{jk}^{(\mathrm{T})}$ are close to zero, which would give a matrix approximating that in Table 13.2(A). The sum of (13.19) over p variables is minimized as in (13.18). This minimization for $\hat{\mathbf{A}}$ in Table 13.3(A) provides the rotation matrix

$$\mathbf{T}'^{-1} = \begin{bmatrix} 0.581 & 0.582 \\ -0.979 & 0.979 \end{bmatrix}. \tag{13.20}$$

Post-multiplication of $\hat{\mathbf{A}}$ in Table 13.3(A) by (13.20) yields $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}'^{-1}$ in Table 13.3(C). This has also been presented in Table 12.1(B), as described in Sect. 12.7.

The reason for adding a small positive constant ε to loadings, as in (13.19), is as follows: (13.19) would be $\prod_{k=1}^{m} a_{jk}^{(\mathrm{T})\,2} = a_{j1}^{(\mathrm{T})\,2} \times \cdots \times a_{jm}^{(\mathrm{T})\,2}$ without ε. Then, the solution which allows $\prod_{k=1}^{m} a_{jk}^{(\mathrm{T})\,2}$ to attain the lower bound 0 is not uniquely determined; multiple solutions could exist. For example, let m be 2. If $a_{j1}^{(\mathrm{T})} = 0$, then $a_{j1}^{(\mathrm{T})\,2} \times a_{j2}^{(\mathrm{T})\,2} = 0$ whatever value $a_{j2}^{(\mathrm{T})}$ takes. This existence of multiple solutions is avoided by adding ε as in (13.19).

13.6 Orthogonal Procrustes Rotation

In this section, we introduce *Procrustes* rotation, whose purpose is *different* from the procedures treated so far. Procrustes rotation generally refers to a class of rotation techniques to rotate $\hat{\mathbf{A}}$ so that the resulting $\mathbf{A}_{\mathrm{T}}$ is *matched* with a *target* matrix $\mathbf{B}$. The rotation was originally conceived by Mosier (1939) and named by Hurley and Cattell (1962) after a figure appearing in Greek mythology.

Let us consider *orthogonal Procrustes rotation* with (13.9), i.e., $\mathbf{T}$ ($m \times m$) constrained to be orthonormal. This is formulated as

$$\text{minimize}\, f(\mathbf{T}) = \left\|\mathbf{B} - \hat{\mathbf{A}}\mathbf{T}\right\|^2 \text{ over } \mathbf{T} \text{ subject to } \mathbf{T}'\mathbf{T} = \mathbf{I}_m. \tag{13.21}$$

This is useful for every case, in which one wishes to match $\hat{\mathbf{A}}\mathbf{T}$ to target $\mathbf{B}$ and examine how *similar* the resulting matrix $\mathbf{A}_{\mathrm{T}} = \hat{\mathbf{A}}\mathbf{T}$ is to the target, under constraint (13.9).

The function $f(\mathbf{T})$ in (13.21) can be expanded as

$$f(\mathbf{T}) = \|\mathbf{B}\|^2 - 2\mathrm{tr}\mathbf{B}'\hat{\mathbf{A}}\mathbf{T} + \mathrm{tr}\mathbf{T}'\hat{\mathbf{A}}'\hat{\mathbf{A}}\mathbf{T} = \|\mathbf{B}\|^2 - 2\mathrm{tr}\mathbf{B}'\hat{\mathbf{A}}\mathbf{T} + \|\mathbf{A}\|^2, \tag{13.22}$$

where we have used $\mathbf{T}\mathbf{T}' = \mathbf{I}_m$ following from (13.9). In the right-hand side of (13.22), only $-2\mathrm{tr}\mathbf{T}'\hat{\mathbf{A}}'\mathbf{B}$ is relevant to $\mathbf{T}$. Thus, the minimization of (13.22) amounts to

$$\text{maximize}\, g(\mathbf{T}) = \mathrm{tr}\mathbf{B}'\hat{\mathbf{A}}\mathbf{T} \text{ over } \mathbf{T} \text{ subject to } \mathbf{T}'\mathbf{T} = \mathbf{I}_m. \tag{13.23}$$

This problem is equivalent to the one in Theorem A.4.2 (Appendix A.4.2). As found there, the solution of $\mathbf{T}$ is given through the singular value decomposition of $\hat{\mathbf{A}}'\mathbf{B}$.

A numerical example is given in Table 13.5. The matrices $\mathbf{B}$ and $\hat{\mathbf{A}}$ presented there seem to be very different. The orthogonal Procrustes rotation for them provide $\mathbf{T} = \begin{bmatrix} 0.53 & 0.85 \\ -0.85 & 0.53 \end{bmatrix}$. The resulting $\hat{\mathbf{A}}\mathbf{T}$ is shown in the right-hand side of Table 13.5, where $\hat{\mathbf{A}}\mathbf{T}$ is found to be very similar to $\mathbf{B}$.

Table 13.5 Example of orthogonal Procrustes rotation

$\mathbf{B}$		$\hat{\mathbf{A}}$		$\hat{\mathbf{A}}\mathbf{T}$	
0.0	0.8	0.6	0.4	−0.02	0.72
0.3	0.7	0.8	0.1	0.34	0.73
0.6	0.6	0.8	−0.2	0.59	0.57
0.8	0.1	0.5	−0.6	0.77	0.11
0.9	0.0	0.5	−0.8	0.94	0.00

13.7 Bibliographical Notes

Simple structure rotation techniques are exhaustively described in Browne (2001) and Mulaik (2011). Procrustes rotation techniques are detailed in Gower and Dijksterhuis (2004), with its special extended version presented by Adachi (2009). The simple structure rotation can be related to the sparse estimation, as discussed in Sect. 22.9 and other literature (e.g., Trendafilov, 2014).

Exercises

13.1. Show that $\mathbf{T} = \mathbf{S}\text{diag}(\mathbf{S}'\mathbf{S})^{-1/2}$ satisfies (13.2), where diag($\mathbf{S}'\mathbf{S}$) denotes the $m \times m$ diagonal matrix whose diagonal elements $d_1, \ldots, d_m$ are those of $\mathbf{S}'\mathbf{S}$ (Note 12.1) and $\text{diag}(\mathbf{S}'\mathbf{S})^{-1/2}$ is the $m \times m$ diagonal matrix whose diagonal elements are $1/d_1^{1/2}, \ldots, 1/d_m^{1/2}$.

13.2. Show that a 2×2 orthonormal matrix $\mathbf{T}$ is expressed as $\mathbf{T} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

13.3. Thurstone (1947) defined simple structure with provisions, which have been rewritten more clearly by Browne (2001, p. 115) as follows:

[1] Each row should contain at least one zero.
[2] Each column should contain at least m zeros, with m the number of factors.
[3] Every pair of columns should have several rows with a zero in one column but not the other.
[4] If $m \geq 4$, every pair of columns should have several rows with zeros in both columns.
[5] Every pair of columns should have a few rows with nonzero loadings in both columns.

Present an example of a 20×4 matrix meeting provisions [1]–[5].

13.4. Minimizing $\frac{1}{m}\sum_{k=1}^{m-1}\sum_{l=k+1}^{m}\sum_{j=1}^{p}(a_{jk}^{(\mathrm{T})2} - \bar{a}_{.k}^{(\mathrm{T})2})(a_{jl}^{(\mathrm{T})2} - \bar{a}_{.l}^{(\mathrm{T})2})$ over $\mathbf{T}$ subject to diag($\mathbf{T}'\mathbf{T}$) = $\mathbf{I}_m$ is included in a family of oblique rotation called *oblimin rotation* (Jennrich & Sampson, 1966), where $a_{jk}^{(\mathrm{T})}$ is the (j, k) element of the rotated loading matrix $\hat{\mathbf{A}}\mathbf{T}'^{-1}$. Discuss the purpose of the above minimization.

13.5. Oblique rotation tends to give a matrix of a *simpler* structure than orthogonal rotation. Explain its reason.

13.6. Show that orthogonal rotation is feasible for the $p \times m$ matrix $\mathbf{A}$ that minimizes $\|\mathbf{V} - \mathbf{A}\mathbf{A}'\|^2$ subject to $\mathbf{A}'\mathbf{A} = \mathbf{I}_m$ for given $\mathbf{V}$.

13.7. Show that oblique rotation is feasible for the solution of principal component analysis, if constraint (5.25) is relaxed as n^{-1}diag($\mathbf{F}'\mathbf{F}$) = $\mathbf{I}_m$ without (5.26). Here, diag() defined in Note 12.1.

13.8. Show the objective function (13.12) in the varimax rotation can be rewritten as

$$f = \frac{1}{n}\mathrm{tr}\mathbf{T}'\hat{\mathbf{A}}'\{(\hat{\mathbf{A}}\mathbf{T}) \odot (\hat{\mathbf{A}}\mathbf{T}) \odot (\hat{\mathbf{A}}\mathbf{T})\} - \frac{1}{n^2}\mathrm{tr}\mathbf{T}'\hat{\mathbf{A}}'\hat{\mathbf{A}}\mathbf{T}\{\mathrm{diag}(\mathbf{T}'\hat{\mathbf{A}}'\hat{\mathbf{A}}\mathbf{T})\}.$$

(ten Berge, Knol, & Kiers, 1988). Here, diag() is defined in Note 12.1, and $\odot$ denotes the element-wise product called the *Hadamard product* and defined as (17.69):

$$\mathbf{X} \odot \mathbf{Y} = \begin{bmatrix} x_{11}y_{11} & \cdots & x_{1p}y_{1p} \\ & \vdots & \\ x_{n1}y_{n1} & \cdots & x_{np}y_{np} \end{bmatrix} = \left(x_{ij}y_{ij}\right)(n \times p) \text{ for } n \times p \text{ matrices}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ & \vdots & \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1p} \\ & \vdots & \\ y_{n1} & \cdots & y_{np} \end{bmatrix}.$$

13.9. *Generalized orthogonal rotation* is formulated as minimizing $\sum_{k=1}^{K} ||\mathbf{H} - \mathbf{A}_k\mathbf{T}_k||^2$ over $\mathbf{H}, \mathbf{T}_1, \ldots, \mathbf{T}_{\mathrm{K}}$ subject to $\mathbf{T}_k'\mathbf{T}_k = \mathbf{T}_k\mathbf{T}_k' = \mathbf{I}_m$, $k = 1, \ldots, K$, for given $p \times m$ matrices $\mathbf{A}_1$, …, $\mathbf{A}_K$. Show that the minimization can be attained by the following algorithm:

Step 1. Initialize $\mathbf{T}_1, \ldots, \mathbf{T}_K$.
Step 2. Set $\mathbf{H} = K^{-1}\sum_{k=1}^{K} \mathbf{A}_k\mathbf{T}_k$.
Step 3. Compute the SVD $\mathbf{A}_k'\mathbf{H} = \mathbf{K}_k\mathbf{\Lambda}_k\mathbf{L}_k'$ to set $\mathbf{T}_k = \mathbf{K}_k\mathbf{L}_k'$ for k = 1, …, K.
Step 4. Finish if convergence is reached; otherwise, go back to Step 2.

13.10. Show

$$K = \sum_{k=1}^{K} ||\mathbf{H} - \mathbf{A}_k\mathbf{T}_k||^2 = \sum_{k=1}^{K-1}\sum_{l=k+1}^{K} ||\mathbf{A}_k\mathbf{T}_k - \mathbf{A}_l\mathbf{T}_l||^2$$

for $\mathbf{H}$ in Step 2 described in Exercise 13.9.

13.11. Let us consider the minimization of $||[\mathbf{M}, \mathbf{c}] - \mathbf{AT}||^2$ over $\mathbf{T}(m \times m)$ and $\mathbf{c}$ $(p \times 1)$ subject to $\mathbf{T}'\mathbf{T} = \mathbf{TT}' = \mathbf{I}_m$ for given $\mathbf{M}(p \times (m-1))$ and $\mathbf{A}(p \times m)$. Here, [$\mathbf{M}$, $\mathbf{c}$] is the $p \times m$ matrix whose final column $\mathbf{c}$ is unknown. Show that the minimization can be attained by the following algorithm:

Step 1. Initialize $\mathbf{T}$.
Step 2. Set $\mathbf{c}$ to the final column of $\mathbf{AT}$.
Step 3. Compute the SVD $\mathbf{A}'[\mathbf{M}, \mathbf{c}] = \mathbf{K\Lambda L}'$ to set $\mathbf{T} = \mathbf{KL}'$.
Step 4. Finish if convergence is reached; otherwise, go back to Step 2.

13.12. Kier's (1994) *simplimax rotation*, which is used for having a matrix of simple structure, is a generalization of the Procrustes rotation introduced in Sect. 13.6. In the simplimax rotation, target matrix $\mathbf{B}$ is unknown except for that $\mathbf{B}$ is constrained to have a specified number of zero elements: $\left\|\mathbf{B} - \hat{\mathbf{A}}\mathbf{T}'^{-1}\right\|^2$ is minimized over $\mathbf{B}$ and $\mathbf{T}$ subject to (13.2) or (13.9) and s elements being zero in $\mathbf{B}$, though the locations of the s zero elements are unknown. Show that, for fixed $\mathbf{T}$, the optimal $\mathbf{B} = (b_{jk})$ is given by $b_{jk} = \begin{cases} 0 & if\ a_{jk}^{[\mathrm{T}]2} \le a_{<s>}^{[\mathrm{T}]2} \\ a_{jk}^{[\mathrm{T}]} & otherwise \end{cases}$, where $a_{jk}^{[\mathrm{T}]}$ is the (j,k) element of $\hat{\mathbf{A}}\mathbf{T}'^{-1}$ and $a_{<s>}^{[\mathrm{T}]2}$ is the sth smallest value among the squares of the elements in $\hat{\mathbf{A}}\mathbf{T}'^{-1}$.