A general approach to handling missing values in Procrustes analysis

Casper J. Albers · John C. Gower

Received: 20 July 2010 / Revised: 14 October 2010 / Accepted: 15 October 2010 /

Published online: 26 October 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract General Procrustes analysis is concerned with transforming a set of given configuration matrices to closest agreement. This paper introduces an approach useful for handling missing values in the configuration matrices in the context of general linear transformations. Centring and/or standardisation are allowed. Simplifications occur in the important case where the transformations are orthogonal. In the most general case, an interesting quadratic constrained optimisation problem appears.

Keywords Procrustes analysis · Missing value estimation · Constrained optimisation

Mathematics Subject Classification (2000) 65F30

1 Introduction

For given configuration matrices $\mathbf{X}_k (k = 1, \dots, K)$, the Procrustes problem is to match the configurations by minimising over $\mathbf{T}_k (k = 1, \dots, K)$ the criterion

$$S = \sum_{h < k}^K ||\mathbf{X}_h \mathbf{T}_h - \mathbf{X}_k \mathbf{T}_k||^2 \quad (1)$$

C. J. Albers (✉)

Department of Psychometrics and Statistical Methods, University of Groningen,
Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
e-mail: c.j.albers@rug.nl

J. C. Gower

Department of Mathematics and Statistics, The Open University, Walton Hall,
Milton Keynes, MK7 6AA, UK

where $\mathbf{T}_k$ is restricted to some matrix class. Thus, every $\mathbf{X}_k$ must have the same number of rows, say n . It is important that the r th row of each configuration matrix should refer to the same item, else the concept of matching lacks meaning. The concept of matching corresponding columns in configurations is contentious. With configurations, columns refer to dimensions and there is no reason for assuming that they must match or, indeed, that all configurations must occupy the same number of dimensions. When columns refer to variables, then matching can occur but it is arguable that Procrustes methods, though often used, are not really appropriate. Throughout this paper we are concerned with configurations so the notion of missing columns is disregarded.

There are many forms of Procrustes analysis, largely depending on the chosen transformation class $\mathbf{T}_k$ and whether $K = 2$ or $K > 2$. The terms $\mathbf{X}_k \mathbf{T}_k$ may be replaced by $\rho_k \mathbf{X}_k \mathbf{T}_k$, thus allowing for possible isotropic scaling factors ρ_k and there may be a requirement that $\|\rho_k \mathbf{X}_k \mathbf{T}_k\|^2 = \|\mathbf{X}_k\|^2$. Whether or not the configurations are centred at their centroid is a further consideration. For an extensive discussion of Procrustes problems, see [Gower and Dijksterhuis \(2004\)](#), who cite many further references. In this paper, we shall be concerned with the treatment of any missing values which, fortunately, as we shall see, may be treated independently of the precise form of Procrustes analysis that may be chosen.

Any missing values in the configuration matrices need to be estimated in order to be able to use (readily available) Procrustes algorithms for non-missing data. Because (1) is a least squares criterion, it is natural also to use least squares for estimating any missing values. The problem has been addressed in the literature for various specific cases. Missing rows were studied by [Commandeur \(1991\)](#), and [ten Berge et al. \(1993\)](#) extended this to an unpatterned allocation of missing values, but both for the important special case where the transformations $\mathbf{T}_k$ are orthogonal matrices. [Gower and Dijksterhuis \(2004\)](#) addressed the problem more generally but left unanswered questions which we resolve in this paper.

To be specific, this paper is new in handling the general case that is applicable to Procrustes methods using any class of transformation matrix, rather than the more usual restriction to orthogonal matrices. Furthermore, it provides a novel methodology for simultaneously handling centring and scaling constraints. However, we do not attempt a complete algorithm. Rather, our aim is to identify the components needed in a complete general least squares algorithm and develop consequent mathematical solutions. Some initial results were indicated in [Albers \(2008\)](#) and [Albers et al. \(2010a\)](#).

After some preliminaries (Sect. 2), we deal with the general case (Sect. 3), including the modifications required when the configuration matrices are to be centred. The special case of orthogonal matrices leads to expected simplifications. Finally, we show how to handle the linked notions of isotropic scaling and standardisation of the configurations (Sect. 4), closing with a discussion in Sect. 5.

2 Preliminaries

In this section we establish notation and list some identities that are useful in the following.

The criterion (1) may be written alternatively as:

$$S = K \sum_{h=1}^K ||\mathbf{X}_h \mathbf{T}_h - \mathbf{G}||^2 \quad (2)$$

where $\mathbf{G}$ is the average, often referred to as the group-average, given by

$$\mathbf{G} = \frac{1}{K} \sum_{h=1}^K ||\mathbf{X}_h \mathbf{T}_h||^2,$$

or as

$$S = \frac{K}{K-1} \sum_{h < k} ||\mathbf{X}_h \mathbf{T}_h - \mathbf{G}_h||^2,$$

where

$$\mathbf{G}_k = \frac{1}{K-1} \sum_{h \neq k} ||\mathbf{X}_h \mathbf{T}_h||^2,$$

known as the k -excluded group average. Partitioning (1) into those terms not including $\mathbf{X}_k$ and those that do gives

$$S = S_k + \sum_{u < v, u, v \neq k} ||\mathbf{X}_h \mathbf{T}_h - \mathbf{X}_k \mathbf{T}_k||^2$$

where

$$S_k = \sum_{u < v \neq k} ||\mathbf{X}_u \mathbf{T}_u - \mathbf{X}_v \mathbf{T}_v||^2,$$

the k -excluded sum-of-squares version of (1). We have

$$\begin{aligned} \sum_{h \neq k} ||\mathbf{X}_h \mathbf{T}_h - \mathbf{X}_k \mathbf{T}_k||^2 &= \sum_{h \neq k} ||\mathbf{X}_h \mathbf{T}_h - \mathbf{G}_k||^2 + (K-1) ||\mathbf{X}_k \mathbf{T}_k - \mathbf{G}_k||^2 \\ &= \frac{1}{K-1} S_k + (K-1) ||\mathbf{X}_k \mathbf{T}_k - \mathbf{G}_k||^2. \end{aligned}$$

where, in the final step, we have used the version of (2) with the k th set excluded, so defining S_k .

Hence, a fourth expression for S is

$$S = \frac{K}{K-1} S_k + (K-1) \|\mathbf{X}_k \mathbf{T}_k - \mathbf{G}_k\|^2. \quad (3)$$

These formulae are easily derived; some proofs may be found in Sect. 9.1 of [Gower and Dijksterhuis \(2004\)](#).

3 General case, with and without centring

In this section we first treat the problem for general transformation matrices and then examine the simplifications arising from orthogonal matrices.

We assume that an algorithm is available for finding the minimum when the data are complete and consider what modifications are needed when there are missing values in the data. To this end, we envisage an ALS algorithm that alternates between (i) estimating the $\mathbf{T}_k$ for current settings of the missing values and (ii) assuming the current settings of $\mathbf{T}_k$, updates the estimates of the missing values. This requires that any missing cells in each $\mathbf{X}_k$ are initialised in some way, such as setting them to zero or the mean. Assuming as we do that (i) is given, it remains only to consider (ii). Variants of the problem include (a) its basic form as just stated (see Sect. 3.1), (b) where it is required that the configurations are centred (see Sect. 3.2) and (c) where there is a size constraint on $\mathbf{X}_k$ (see Sect. 4).

Denote by Ξ_k an updating matrix with zeros everywhere except for the cells corresponding to the m_k missing values in the k th set. Throughout Sects. 3 and 4 we shall be concerned only with updating the k th setting of $\mathbf{X}_k$, so the suffix k in Ξ_k and m_k becomes redundant and these settings will be written simply as Ξ and m .

3.1 General case, no centring

If we want to update the k th set from $\mathbf{X}_k$ to $\mathbf{X}_k - \Xi$, it suffices to consider minimizing over Ξ the term $\|(\mathbf{X}_k - \Xi)\mathbf{T}_k - \mathbf{G}_k\|^2$, as this is the only term in (3) that involves $\mathbf{X}_k$. We rewrite this as

$$\|\Xi \mathbf{T}_k - \mathbf{Y}_k\|^2 \quad (4)$$

where $\mathbf{Y}_k = \mathbf{X}_k \mathbf{T}_k - \mathbf{G}_k$. Minimising (4) is equal to minimising:

$$\text{trace}(\Xi \mathbf{T} \Xi' - 2\Xi \mathbf{Y}') + \text{constant} \quad (5)$$

where $\mathbf{T} = \mathbf{T}_k \mathbf{T}_k'$ and $\mathbf{Y} = \mathbf{Y}_k \mathbf{T}_k'$.

Expression (5) is to be minimized over Ξ , which contains only m active, usually non-zero, values $\xi' = (\xi_1, \xi_2, \dots, \xi_m)$. We assume that the elements of the column-vector ξ are presented in the order given by concatenating the non-zero values in successive rows of Ξ . Thus, (5) is a quadratic form in ξ which, for some $m \times m$

symmetric matrix $\mathbf{A}$ and m -vector $\mathbf{y}$, conformable with $\boldsymbol{\xi}$, may be written

$$\boldsymbol{\xi}'\mathbf{A}\boldsymbol{\xi} - 2\boldsymbol{\xi}'\mathbf{y} + \text{constant}$$

and which is minimal when

$$\boldsymbol{\xi} = \mathbf{A}^{-1}\mathbf{y}. \quad (6)$$

(It will become clear at a later stage that $\mathbf{A}$ is indeed non-singular, hence solution (6) always exists.) It remains only to establish the precise forms of $\mathbf{A}$ and $\mathbf{y}$. To do this we need to examine the detailed structure of $\Xi\mathbf{T}\Xi'$. Normally, Ξ will have many zero rows and columns and it is only necessary to examine the non-zero part. Thus, when there are r non-zero rows, the non-zero part of $\Xi\mathbf{T}\Xi'$ may be written as the matrix

$$\Xi\mathbf{A}\Xi' = \begin{bmatrix} \boldsymbol{\xi}'_1\mathbf{A}_{11}\boldsymbol{\xi}_1 & \boldsymbol{\xi}'_1\mathbf{A}_{12}\boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}'_1\mathbf{A}_{1r}\boldsymbol{\xi}_r \\ \boldsymbol{\xi}'_2\mathbf{A}_{21}\boldsymbol{\xi}_1 & \boldsymbol{\xi}'_2\mathbf{A}_{22}\boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}'_2\mathbf{A}_{2r}\boldsymbol{\xi}_r \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}'_r\mathbf{A}_{r1}\boldsymbol{\xi}_1 & \boldsymbol{\xi}'_r\mathbf{A}_{r2}\boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}'_r\mathbf{A}_{rr}\boldsymbol{\xi}_r \end{bmatrix} \quad (7)$$

where $\mathbf{A}_{ij}$ is obtained from $\mathbf{T}$ and Ξ , as follows. We use the columns indexed by the m_i non-zero elements $\boldsymbol{\xi}_i$ in the i th row of Ξ to index the rows of $\mathbf{T}$ and the columns indexed by the m_j non-zero elements $\boldsymbol{\xi}_j$ in the j th row of Ξ to index the columns of $\mathbf{T}$. This gives the $m_i \times m_j$ matrix $\mathbf{A}_{ij}$. Thus, $\boldsymbol{\xi}' = (\boldsymbol{\xi}'_1, \boldsymbol{\xi}'_2, \dots, \boldsymbol{\xi}'_r)$ where $\boldsymbol{\xi}_i$ derives from the non-zero values in the i th row of Ξ , but is written as a column-vector. With this notation

$$\text{trace}(\Xi\mathbf{T}\Xi') = \sum_{i=1}^r \boldsymbol{\xi}'_i\mathbf{A}_{ii}\boldsymbol{\xi}_i = \boldsymbol{\xi}'\mathbf{\Delta}\boldsymbol{\xi}$$

where $\mathbf{\Delta} = \text{diag}(\mathbf{A}_{11}, \dots, \mathbf{A}_{rr})$, an $m \times m$ block-diagonal matrix with r blocks of sizes $(m_1, m_2, \dots, m_r)$. Similarly, $\mathbf{y}$ consists of the m values of $\mathbf{Y}$ corresponding to the missing value positions in Ξ . Note that this trace operation uses only the diagonal elements of (7); the whole of (7) will be needed when we handle centring, below. Thus (6) is easily solved as $\boldsymbol{\xi} = \mathbf{\Delta}^{-1}\mathbf{y}$. We note that because of the block-diagonal structure of $\mathbf{\Delta}$, the missing values in the i th row are given by

$$\boldsymbol{\xi}_i = \mathbf{A}_{ii}^{-1}\mathbf{y}_i, \quad (8)$$

so it suffices to invert a series of $m_i \times m_i$ matrices rather than the whole of $\mathbf{\Delta}$.

3.2 General case, with centring

Normally, Procrustes analysis is conducted on column-centred matrices, because this gives the translation of each configuration that minimizes the residual sum-of-squares.

Then, each $\mathbf{X}_k$ is centred so that $\mathbf{1}'\mathbf{X}_k = 0$ and when estimating missing values, this condition has to be respected when updating the current values of $\mathbf{X}_k$. We write the centring matrix $\mathbf{I} - \mathbf{1}'\mathbf{1}/n$ as $\mathbf{I} - \mathbf{N}$. Assuming that all the $\mathbf{X}_k$ are centred, and hence also $\mathbf{Y}$ and $\mathbf{G}$, the only change to the uncentred case is to replace Ξ in criterion (3), and hence throughout, by $(\mathbf{I} - \mathbf{N})\Xi$. Thus now we have to minimize

$$\text{trace}((\mathbf{I} - \mathbf{N})\Xi\mathbf{T}\Xi'(\mathbf{I} - \mathbf{N}) - 2(\mathbf{I} - \mathbf{N})\Xi\mathbf{Y}') + \text{constant} \quad (9)$$

which, because of commutivity under the trace operator, the idempotency of $(\mathbf{I} - \mathbf{N})$ and the centring of $\mathbf{Y}$ simplifies to

$$\text{trace}((\mathbf{I} - \mathbf{N})\Xi\mathbf{T}\Xi' - 2\Xi\mathbf{Y}') + \text{constant}.$$

Now, $\text{trace}(\mathbf{I} - \mathbf{N})\Xi\mathbf{T}\Xi' = \text{trace}(\Xi\mathbf{T}\Xi') - \mathbf{1}'\Xi\mathbf{T}\Xi'\mathbf{1}/n$ the first term of which is $\xi'\Delta\xi$, as before, and the second term requires the sum of the elements of the matrix (7). We have that $\mathbf{1}'\Xi\mathbf{T}\Xi'\mathbf{1} = \xi'\mathbf{A}\xi$ so that minimising (9) gives

$$\xi = \left(\Delta - \frac{1}{n}\mathbf{A}\right)^{-1} \mathbf{y} \quad (10)$$

again requiring a matrix inverse. Now the block diagonality is lost, so there is no simplification in evaluating the inverse “row-by-row” as was possible with (8) when centring is not required.

3.3 The orthogonal case

When the $\mathbf{T}_k$ are orthogonal matrices, denoted by $\mathbf{Q}_k$ as usual, (2) may be rewritten

$$S = K \sum_{h=1}^K \|\mathbf{X}_h - \mathbf{G}\mathbf{Q}'_h\|^2$$

which is the case considered by [ten Berge et al. \(1993\)](#). In our notation we have that $\mathbf{T} = \mathbf{I}$ and hence Δ is a unit matrix. Thus, (8) becomes simply

$$\xi = \mathbf{y}, \quad (11)$$

as found by [ten Berge et al. \(1993\)](#) who use ad hoc methods to enforce centring. More directly, (10) gives $\xi = (\mathbf{I} - \frac{1}{n}\mathbf{A})^{-1}\mathbf{y}$, where $\mathbf{A}$ will have unit values off the diagonal when there is more than one missing value in some column. Indeed, by reordering ξ by columns, rather than by the rows of Ξ , it will be found that when $\mathbf{T} = \mathbf{I}$ the reordered version of $\mathbf{A}$ becomes block-diagonal with diagonal blocks of the form $\mathbf{1}\mathbf{1}'$ of order $s \times s$, corresponding to a column with s missing values. Then, the missing values for that column may be estimated independently from those of the other columns by

solving

$$\left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}'\right) \xi_s = \mathbf{y}_s \quad (12)$$

to give

$$\xi_s = \left(\mathbf{I} + \frac{1}{n-s} \mathbf{1}\mathbf{1}'\right) \mathbf{y}_s, \quad (13)$$

where the suffix s here indicates that vectors of length s , corresponding to the s missing values and corresponding positions in $\mathbf{y}$, are involved. Eq. (13) is neat but the difference from estimating and then centring is unlikely to have a significant effect on the convergent or rate of convergence.

We note that (12) may be obtained directly by rewriting the centred form of (4) as

$$\min_{\xi} \|(\mathbf{I} - \mathbf{N})\Xi - \mathbf{Y}\|^2$$

where $\mathbf{Y} = \mathbf{X}_k - \mathbf{G}\mathbf{Q}_k'$, which may be minimized column by column. For a column with s missing values this gives

$$\min_{\xi} \left\| \left(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}'\right) \xi_s - \mathbf{y}_s \right\|^2$$

with solution (12) and thence (13).

4 Scaling and standardisation

Sometimes it is useful to include isotropic scaling or to standardise all configurations to equal size (see, e.g., Gower and Dijksterhuis 2004). In this section we shall adapt the methodology of Sect. 3 to accommodate scaling and standardisation. We show that isotropic scaling is easily handled but standardisation leads to some interesting problems.

When an isotropic scaling parameter ρ_k is included (1) becomes

$$S : \sum_{h < k}^K \|\rho_h \mathbf{X}_h \mathbf{T}_h - \rho_k \mathbf{X}_k \mathbf{T}_k\|^2 \quad (14)$$

with an identification constraint

$$C : \sum_{k=1}^K \|\rho_k \mathbf{X}_k \mathbf{T}_k\|^2 = c \quad (15)$$

where the precise non-zero value of c is immaterial. The identification constraint is essential to avoid trivial solutions $\rho_k = 0$. Isotropic parameters ρ_k are estimated, usually using eigenvalue methods, within the Procrustes step of the ALS algorithms (see [Gower and Dijksterhuis 2004](#)). They require no further consideration here.

Apart from estimating isotropic scaling, it might be convenient to standardise so that all configurations initially have the same specified size, $\text{trace}(\mathbf{X}_k' \mathbf{X}_k) = \kappa$ ($k = 1, 2, \dots, K$). Isotropic scaling and standardisation $\|\mathbf{X}_k\|^2 = \kappa$ are different and, indeed, both may be applied. Standardisation is awkward when there are missing values, as then the sizes of the configurations are not determinate and any initial settings must be modified during the iterative process to take account of updating. It follows that standardisation must be handled as part of the missing values step.

Fortunately, in the completely general case, standardisation is unnecessary because on noting that $\mathbf{X}_k \mathbf{T}_k = (\alpha \mathbf{X}_k)(\alpha^{-1} \mathbf{T}_k)$ it can be seen that α may be chosen so that $\alpha \mathbf{X}_k$ is standardised as may be required, while $\alpha^{-1} \mathbf{T}_k$ may be subsumed into the general form of $\mathbf{T}_k$. Even more generally, the identity $\mathbf{X}_k \mathbf{T}_k = (\mathbf{X}_k \mathbf{S})(\mathbf{S}^{-1} \mathbf{T}_k)$ where $\mathbf{S}$ is diagonal shows that individual columns of $\mathbf{X}_k$ may be scaled independently of one another.

These remarks do not extend to the case where $\mathbf{T}_k$ is from a specified class of matrices, and $\alpha^{-1} \mathbf{T}_k$ is not a member of the same class, as is the case with orthogonal matrices. When standardisation is required, adherence to the proper matrix class is the concern of the Procrustes step but maintaining standardisation must be handled in the missing value step; this case is considered in the following.

It is tempting to handle this by ensuring that initial settings of missing values respect the required standardisation and arranging that this is maintained by applying the constraint

$$\|\mathbf{X}_k - \Xi\|^2 = \|\mathbf{X}_k\|^2$$

or, when centring is required

$$\|(\mathbf{I} - \mathbf{N})(\mathbf{X}_k - \Xi)\|^2 = \|(\mathbf{I} - \mathbf{N})\mathbf{X}_k\|^2$$

However, this is inappropriate because only the estimated missing values will be scaled, the non-missing values retaining their initial settings. In the extreme case where $\mathbf{X}_k$ has a single missing value, only its sign could be changed.

4.1 Standardisation without centring

To overcome the problem just discussed, a standardising parameter σ must be applied to the whole of $\mathbf{X}_k$, and chosen to minimise

$$S : \|\sigma(\mathbf{X}_k - \Xi)\mathbf{T}_k - \mathbf{G}_k\|^2, \quad (16)$$

over σ and Ξ subject to the constraint

$$C : \|\sigma(\mathbf{X}_k - \Xi)\|^2 = \|\mathbf{X}_k\|^2. \quad (17)$$

Note that S and C in (16) and (17) have been redefined from their settings in (14) and (15). Expressions (16) and (17) look remarkably like the isotropic scaling expressions (14) and (15) but there is the important difference that σ applies only to the k th configuration and does not have to satisfy a global constraint like ρ_k in (15). With this formulation, the whole of $\mathbf{X}_k$ gets updated and rescaled to $\sigma(\mathbf{X}_k - \Xi)$ in the missing value step of the algorithm. This is entirely reasonable because it would be unrealistic to assume that the correct scaling could be applied at the outset. When one wished also to combine isotropic scaling with standardisation then, during the missing value step, ρ_k may be considered absorbed into $\mathbf{T}_k$.

We note that if σ is considered fixed, then both S and C are quadratic forms in the non-zero elements of Ξ so that the minimisation of S subject to C is included as a special case of the class of problems discussed by Albers (2008) and Albers et al. (2010a,b). Similarly, if Ξ is considered fixed, minimising (16) subject to (17) is a trivial quadratic problem. Thus we might achieve the required minimisation using an alternating algorithm. In the following we consider the possibilities for direct minimisation.

The general case

Recalling that we are estimating the missing values in $\mathbf{X}_k$ solely, we can continue to omit subscripts k for convenience. Thus, expanding the constraint (17) gives

$$C : \sigma^2(\|\mathbf{X}\|^2 - 2\xi'\mathbf{x} + \xi'\xi) = \|\mathbf{X}\|^2 \quad (18)$$

where $\mathbf{x}$ consists of the m values of $\mathbf{X}$ corresponding to the missing value positions in Ξ . On defining $\varphi = \sigma\xi$, (18) becomes

$$C : \sigma^2\|\mathbf{X}\|^2 - 2\sigma\varphi'\mathbf{x} + \varphi'\varphi = \|\mathbf{X}\|^2 \quad (19)$$

which is manifestly a quadratic form in φ and σ .

Similarly, expanding S gives

$$S = \sigma^2\text{trace}(\mathbf{X} - \Xi)'(\mathbf{X} - \Xi)\mathbf{T} - 2\sigma\text{trace}(\mathbf{X} - \Xi)'\mathbf{G} + \|\mathbf{G}\|^2 \quad (20)$$

(the subscript for $\mathbf{T}_k$ is retained, to avoid confusion with $\mathbf{T} = \mathbf{T}_k\mathbf{T}_k'$). On substituting for φ , (20) takes the form

$$S = \varphi'\Delta\varphi - 2\sigma\varphi'\text{mvec}(\mathbf{X}\mathbf{T}) + \sigma^2\|\mathbf{X}\|^2 + 2\varphi'\text{mvec}(\mathbf{G}\mathbf{T}_k') - 2\sigma\text{trace}(\mathbf{X}'\mathbf{G}\mathbf{T}_k') + \|\mathbf{G}\|^2 \quad (21)$$

where mvec is the operator that concatenates conformably with $\boldsymbol{\varphi}$, the ‘active elements’ of subsequent columns of a matrix into a vector. Next we write (21) as

$$S : \begin{pmatrix} \boldsymbol{\varphi} \\ -\sigma \end{pmatrix}' \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{w} \\ \mathbf{w}' & ||\mathbf{X}||^2 \end{bmatrix} \begin{pmatrix} \boldsymbol{\varphi} \\ -\sigma \end{pmatrix} + 2 \left(\text{trace}(\mathbf{X}' \mathbf{G} \mathbf{T}'_k) \right)' \begin{pmatrix} \boldsymbol{\varphi} \\ -\sigma \end{pmatrix} + ||\mathbf{G}||^2$$

where $\mathbf{v} = \text{mvec}(\mathbf{G} \mathbf{T}'_k)$ and $\mathbf{w} = \text{mvec}(\mathbf{X} \mathbf{T})$.

Finally, we “complete the square” using $\mathbf{z}' \mathbf{A} \mathbf{z} + 2 \mathbf{a}' \mathbf{z} = (\mathbf{z} + \mathbf{A}^{-1} \mathbf{a})' \mathbf{A} (\mathbf{z} + \mathbf{A}^{-1} \mathbf{a}) + \text{constant}$ to give

$$S : \boldsymbol{\zeta}' \begin{bmatrix} \boldsymbol{\Delta} & \mathbf{w} \\ \mathbf{w}' & ||\mathbf{X}||^2 \end{bmatrix} \boldsymbol{\zeta} + \text{constant} \quad (22)$$

where

$$\boldsymbol{\zeta} = \begin{pmatrix} \boldsymbol{\varphi} + \boldsymbol{\Delta}^{-1}(\mathbf{v} - t \mathbf{w}) \\ -\sigma + t \end{pmatrix}$$

and

$$t = \frac{\mathbf{w} \boldsymbol{\Delta}^{-1} \mathbf{v} - \text{trace}(\mathbf{X}' \mathbf{G} \mathbf{T}'_k)}{\mathbf{w}' \boldsymbol{\Delta}^{-1} \mathbf{w} - ||\mathbf{X}||^2}.$$

Thus, the quadratic form (20) in $\boldsymbol{\varphi}$ and σ with linear terms has been transformed into a simple quadratic form in $\boldsymbol{\zeta}$. Similarly, the constraint C may be transformed into a quadratic form in $\boldsymbol{\zeta}$, incidentally generating linear terms. To minimise S subject to a quadratic constraint C is precisely the problem discussed by Albers et al. (2010a,b).

All that is required for a unique solution is that the symmetric matrix in (22) be positive semi-definite. We know that $\boldsymbol{\Delta}$ is psd because it is formed from diagonal blocks of the inner-product matrix $\mathbf{T}$. Furthermore, S , being a sum-of-squares is never negative so the determinant of the full matrix is not negative. It follows that the psd requirement is satisfied.

In conclusion, the solution to (22) will deliver a unique $\boldsymbol{\zeta}$ and hence $\boldsymbol{\varphi}$ and σ and thence $\boldsymbol{\xi}$. This provides all that is required to update $\mathbf{X}_k$ for the general case.

Standardisation when $\mathbf{T}_k$ is orthogonal

When $\mathbf{T}_k$ is orthogonal, $\mathbf{T} = \mathbf{I}$ and the quadratic terms in (16) and (17) coincide, which greatly simplifies the optimisation problem of Albers et al. (2010a,b) allowing (20) to be solved explicitly, as follows. We first observe that the first term in (16) automatically satisfies the constraint so we now have to minimise

$$S : ||\mathbf{X}||^2 + 2 (\sigma \text{trace}(\mathbf{G}' \mathbf{X} \mathbf{T}_k) - \boldsymbol{\varphi}' \mathbf{v}) \quad (23)$$

where $\mathbf{v} = \text{mvec}(\mathbf{G}\mathbf{T}'_k)$, as above. Thus, we have to maximize the linear expression in square brackets, subject to the previous quadratic constraint

$$C : \sigma^2 \|\mathbf{X}\|^2 - 2\sigma \boldsymbol{\varphi}' \mathbf{x} + \boldsymbol{\varphi}' \boldsymbol{\varphi} = \|\mathbf{X}\|^2.$$

When $\mathbf{A}$ is pd, the maximum of $\mathbf{a}'\mathbf{z}$ subject to $\mathbf{z}'\mathbf{A}\mathbf{z} = k$ is given by

$$\mathbf{z} = \mathbf{A}^{-1} \mathbf{a} \sqrt{\frac{k}{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}}. \quad (24)$$

In the present case we have

$$\mathbf{z} = \begin{pmatrix} \varphi \\ \sigma \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} -\mathbf{v} \\ \text{trace}(\mathbf{G}'\mathbf{X}\mathbf{T}_k) \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{I} & -\mathbf{x} \\ -\mathbf{x}' & \|\mathbf{X}\|^2 \end{pmatrix}.$$

Now,

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} + \alpha \mathbf{x}\mathbf{x}' & \alpha \mathbf{x} \\ \alpha \mathbf{x}' & \alpha \end{pmatrix}, \quad \text{where } \alpha = (\|\mathbf{X}\|^2 - \mathbf{x}\mathbf{x}')^{-1}.$$

Inserting into (24) gives

$$\begin{pmatrix} \varphi \\ \sigma \end{pmatrix} = \begin{pmatrix} \beta \mathbf{x} - \mathbf{v} \\ \beta \end{pmatrix} \sqrt{\frac{\|\mathbf{X}\|^2}{\beta^2/\alpha + \mathbf{v}\mathbf{v}'}}, \quad \text{where } \beta = \alpha (\text{trace}(\mathbf{G}'\mathbf{X}\mathbf{T}'_k) - \mathbf{x}'\mathbf{v}) \quad (25)$$

which, with $\boldsymbol{\varphi} = \sigma \boldsymbol{\xi}$, determines the required scaling and the update vector.

4.2 Standardisation with centring

Now σ and Ξ , must be chosen to minimise

$$S : \|\sigma(\mathbf{I} - \mathbf{N})(\mathbf{X}_k - \Xi)\mathbf{T}_k - \mathbf{G}_k\|^2, \quad (26)$$

subject to the standardising constraint

$$C : \|\sigma(\mathbf{I} - \mathbf{N})(\mathbf{X}_k - \Xi)\|^2 = \|\mathbf{X}_k\|^2. \quad (27)$$

Recall that $\mathbf{X}_k$ and $\mathbf{G}_k$ are initially centred and remain so throughout any iterations. Thus the constraint may be written

$$C : \|\sigma(\mathbf{X}_k - \Xi) + \sigma \mathbf{N}\Xi\|^2 = \|\mathbf{X}_k\|^2$$

the first part of which is the uncentred constraint and $\sigma \mathbf{N}\Xi$ is its mean. Now $\|\sigma \mathbf{N}\Xi\|^2 = \sigma^2 \mathbf{1}' \Xi \Xi' \mathbf{1} / n$ which is a quadratic form in $\boldsymbol{\xi}$ which we shall write as $\boldsymbol{\xi}' \mathbf{C} \boldsymbol{\xi}$. Because $\mathbf{1}' \Xi$ is a row-vector of the sum of the m_i missing-value corrections in the i th column

of Ξ , $\mathbf{C}$ will have $m_i \times m_i$ blocks of units when the elements of ξ are numbered by columns (with a suitable permutation if they are numbered by rows). It follows that

$$C : \sigma^2(||\mathbf{X}||^2 - 2\xi'\mathbf{x} + \xi'\xi) - \sigma^2\xi'\mathbf{C}\xi/n = ||\mathbf{X}||^2$$

where the first part is precisely the same as in the uncentred case. Setting $\varphi = \sigma\xi$ and $\mathbf{J} = (\mathbf{I} - \mathbf{C}/n)$, C becomes

$$C : \sigma^2||\mathbf{X}||^2 - 2\sigma\varphi'\mathbf{x} + \varphi'\mathbf{J}\varphi = ||\mathbf{X}||^2$$

which has a very similar form to (19).

Similarly the criterion may be written

$$S : ||\sigma(\mathbf{X}_k - \Xi)\mathbf{T}_k - \mathbf{G}_k + \sigma\mathbf{N}\Xi\mathbf{T}_k||^2,$$

which may be expanded as in (20) but supplemented by the additional term $\sigma^2\xi'\mathbf{C}_T\xi/n$ where $\mathbf{C}_T$ is defined by $\{\mathbf{C}_T(i, j)\} = t_{ij}\mathbf{1}_i'\mathbf{1}_j$ and $\mathbf{1}_i$ is a vector of m_i units. Note that when $\mathbf{T} = \mathbf{I}$, $\mathbf{C}_T$ is the same as $\mathbf{C}$. On substituting for φ , the only difference from (21) and (22) is that Δ is replaced by $(\Delta - \mathbf{C}_T)/n$ which retains the psd property. With this small change, the subsequent development and solution remains as in the uncentred case of Sect. 4.1.

Standardisation with centring when $\mathbf{T}_k$ is orthogonal

When $\mathbf{T}_k$ is orthogonal, centring does not introduce major complications, thus we may derive an explicit solution similar to that of the previous section. As before the first term in (26) automatically satisfies the constraint

$$S : ||\mathbf{X}||^2 + 2(\sigma\text{trace}(\mathbf{G}'\mathbf{X}\mathbf{T}_k) - \varphi'\mathbf{v}),$$

which is the same as (23) because the centring operator has no effect on the already centred matrix $\mathbf{G}$. Again we must maximize the linear expression in square brackets. The centred form of the constraint has to be applied

$$C : \sigma^2||\mathbf{X}||^2 - 2\sigma\varphi'\mathbf{x} + \varphi'\mathbf{J}\varphi = ||\mathbf{X}||^2$$

so that

$$\mathbf{z} = \begin{pmatrix} \varphi \\ \sigma \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} -\mathbf{v} \\ \text{trace}(\mathbf{G}'\mathbf{X}\mathbf{T}_k) \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{J} & -\mathbf{x} \\ -\mathbf{x}' & ||\mathbf{X}||^2 \end{pmatrix}.$$

We have that

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{I} + \alpha\mathbf{J}^{-1}\mathbf{x}\mathbf{x}'\mathbf{J}^{-1} & \alpha\mathbf{J}^{-1}\mathbf{x} \\ \alpha\mathbf{x}'\mathbf{J}^{-1} & \alpha \end{pmatrix}, \quad \text{where } \alpha = (||\mathbf{X}||^2 - \mathbf{x}'\mathbf{J}^{-1}\mathbf{x})^{-1}.$$

Table 1 Overview of methods for situations without standardisation

	Centring not required	Centring required
General $\mathbf{T}_k$	Formula (8)	Formula (10)
Orthogonal $\mathbf{T}_k$	Formula (11)	Formula (13)

Table 2 Overview of methods for situations where standardisation is required

	Centring not required	Centring required
Unrestricted $\mathbf{T}_k$	Formula (8)	Formula (10)
Orthogonal $\mathbf{T}_k$	Formula (25)	Formula (28)
Other types of $\mathbf{T}_k$	min (22) s.t. (19)	min (26) s.t. (27)

The constrained optimisation problems for ‘other types of $\mathbf{T}_k$ ’ can be solved using [Albers et al. \(2010a,b\)](#)

Inserting into (24) gives

$$\begin{pmatrix} \boldsymbol{\varphi} \\ \sigma \end{pmatrix} = \begin{pmatrix} \beta \mathbf{J}^{-1} \mathbf{x} - \mathbf{v} \\ \beta \end{pmatrix} \sqrt{\frac{\|\mathbf{X}\|^2}{\beta^2/\alpha + \mathbf{v}'\mathbf{v}}} \quad \text{where } \beta = \alpha \left(\text{trace}(\mathbf{G}'\mathbf{X}\mathbf{T}_k) - \mathbf{x}'\mathbf{J}^{-1}\mathbf{v} \right) \quad (28)$$

which, with $\boldsymbol{\varphi} = \sigma \boldsymbol{\xi}$, determines the required scaling and the update vector $\boldsymbol{\xi}$.

Throughout, we require the inverse of $\mathbf{J} = \mathbf{I} - \mathbf{C}/n$. Recall that $\mathbf{C}$ is defined as a block-diagonal matrix with elements $\{\mathbf{C}(i, i)\} = \mathbf{1}'_i \mathbf{1}_i$ where the length of $\mathbf{1}_i$ is the number m_i of missing values in the i th active column of Ξ . It follows that $\mathbf{J}^{-1}$ is block-diagonal with elements $\left(\mathbf{I} + \frac{n}{n-m_i} \mathbf{1}_i \mathbf{1}'_i \right)$ which helps with computation.

5 Discussion

In this paper we develop a method for estimating missing values in Procrustes analysis. Our method goes beyond what was already available in the literature, in the sense that the transformations need not be restricted to the class of orthogonal matrices. Furthermore, standardisation and centring are embedded in our approach. In Sect. 4 we show that isotropic scaling does not have impact on the estimation of missing values and thus can be disregarded in this paper.

An overview of which formulae to use in which situation is given in Table 1 for situations where standardisation is not required, and in Table 2 for situations where standardisation is required.

Recall that the updating process discussed in Sects. 3 and 4 apply only to the k th set. It is essential that, during the missing values step of the ALS-algorithm, only one of the matrices is updated. This update in $\mathbf{X}_k$ will estimate the missing values such that $\mathbf{X}_k \mathbf{T}_k$ is closest to $\mathbf{G}_k$; after this change in $\mathbf{X}_k$, a new Procrustes step is needed to recompute the matrices $\mathbf{G}_h$ ($h \neq k$) and to find the optimal transformations. Then, by simply alternating between the matrices with missing values in the missing values

Table 3 General outline for an algorithm

Loop
Compute S from (1)
For set $k = 1$ to K
Compute $\tilde{S}$ from (1)
Set missing values in $\mathbf{X}_k$ to 0 (or some other value)
1. For current estimates, compute transformations $\mathbf{T}_i$ using Procrustes algorithms
2. For current $\mathbf{T}_i$, compute missing value estimates using Tables 1 and 2
Compute $\tilde{S}^*$ from (1)
3. Repeat 1. and 2. until $\tilde{S} - \tilde{S}^* < \varepsilon$
End For
Compute S^* from (1)
Repeat until $S - S^* < \varepsilon$

updating step, convergence can be guaranteed. However, the usual caveats about ALS algorithms possibly finding local minima apply and convergence can be slow, as our studies on various data sets confirmed: after a relatively small number of initial steps where substantive reductions in S are achieved, the convergence becomes quite slow. In practice, the transformations $\mathbf{X}_k \mathbf{T}_k$ as well as the estimates for the missing values only change slightly in these steps, so, depending on the nature of the research performed, one might be satisfied with a threshold ε (see Table 3) larger than what one initially would specify.

The approach discussed in this paper can be modified as follows. Instead of performing only 1 step in the Procrustes analysis in each loop of the ALS-algorithm, one can also perform a full Procrustes analysis, i.e. finding the optimal transformations, given the current estimates of the missing values, before updating missing values. This approach has the advantage that no alterations have to be made to the (already available) Procrustes code, but that a ‘shell’ involving the missing values step that repeatedly calls this code suffices. The drawback, however, is that this is at the cost of convergence speed. Computations indicate that the processor time required roughly doubles. This will, however, only be a serious drawback in very large data sets.

We have identified the components needed but have not assembled them into a general algorithm. Table 3 outlines the form that a general algorithm could take but further study is needed to identify the best updating strategy, influence of the setting of the threshold ε and initial settings of the missing values. Only when these have been done, can we go on to study efficiency. Not that we think that speed is especially important: when compared to the human time spent on research, algorithmic time is unlikely to be a serious consideration, except possibly for important special cases, such as orthogonal Procrustes methods, when bespoke algorithms may be worth considering.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Albers CJ (2008) Some quadratic optimisation problems in psychometrics. In: Shigemasu K (ed) *New trends in psychometrics*. Universal Academic Press, Tokyo, 1–6
- Albers CJ, Critchley F, Gower JC (2010) Quadratic minimisation problems in statistics. *J Multivariate Anal.* doi:[10.1016/j.jmva.2009.12.018](https://doi.org/10.1016/j.jmva.2009.12.018)
- Albers CJ, Critchley F, Gower JC (2010) Explicit minimisation of a convex quadratic under a quadratic constraint (In preparation)
- Commandeur JJF (1991) *Matching configurations*. DSWO Press, Leiden
- Gower JC, Dijksterhuis GB (2004) *Procrustes problems*. Oxford Statistical Science Series 30. University Press, Oxford
- ten Berge JMF, Kiers HAL, Commandeur JJF (1993) Orthogonal Procrustes rotation for matrices with missing values. *Br J Math Stat Psychol* 46(1):119–134