Addendum to

Properties of and Algorithms for Fitting Three-way Component Models with Offset Terms

Henk A.L. Kiers

University of Groningen

In this addendum, some derivations are given that were omitted in the published paper Kiers, H.A.L. (2005) Properties of and Algorithms for Fitting Three-way Component Models with Offset Terms, *Psychometrika*. Specifically, here we give the extended version, that is the complete description including the derivation, whereas the paper only gives the description without derivations.

Algorithms for Successive Fitting Approaches (section 3.2)

3.2.1. Successive Fitting of the E1 model

The successive procedure proposed here for fitting the E1 model consists of the following (noniterative) steps.

Step 1. Center the three-way data across the A-mode, and denote the A-mode matricized form of this array as $\mathbf{X}_a^{\mathbf{c}}$

Step 2. Fit the ordinary Tucker3 or CP model to the thus centered data, considering that centering removed the offset terms, by minimizing $f(\mathbf{A}^{\mathbf{c}},\mathbf{B},\mathbf{C},\mathbf{G_a})=\| \mathbf{X}_a^{\mathbf{c}} - \mathbf{A}^{\mathbf{c}}\mathbf{G}_a(\mathbf{C}'\otimes\mathbf{B}')\|^2$

Step 3. Fit the E1 model to the original data by minimizing

$$f(\mathbf{A},\boldsymbol{\mu}_\mathbf{b})=\| \mathbf{X}_a - \mathbf{A}\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{b}')\|^2 \tag{26}$$

subject to the constraint $\mathbf{A}=\hat{\mathbf{A}}^\mathbf{c}+\mathbf{1a}'$.

The rationale behind this approach is as follows. In Step 1 we center the data over the A-mode such that we have from (9)

$$\mathbf{X}_a^\mathbf{c}= \mathbf{A}^\mathbf{c}\mathbf{G}_a(\mathbf{C}'\otimes\mathbf{B}') + \mathbf{E}_a^\mathbf{c}, \tag{27}$$

from which clearly the offset terms are removed, so that we end up with an ordinary Tucker3 or CP model plus a (centered) error term. In fact, this effect of removing offset terms was mentioned by Harshman and Lundy (1984) as the prime motivation for this type of centering prior to analysis (in their case CP). Given that we have an ordinary Tucker3 or CP model, the straightforward next step (Step 2) is to fit it to the centered data in the least squares sense. In the error free case, indeed this would lead to correct estimates of $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G_a}$, whereas $\mathbf{A^c}$ would give a centered version of the correct matrix $\mathbf{A}$. In cases with error, this will no longer hold exactly, but it can be expected that the resulting estimates $\hat{\mathbf{A}}^\mathbf{c},\hat{\mathbf{B}},\hat{\mathbf{C}}$, and $\hat{\mathbf{G}}_a$ are reasonable approximations to the correct parameter matrices $\mathbf{A^c}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{G_a}$. In Step 3, finally, we find estimates for the parameters and parameter matrix for which we do not have estimates yet: $\boldsymbol{\mu}_\mathbf{b}$ and $\mathbf{A}$. We can write $\mathbf{A}$ as $\mathbf{A} = \mathbf{A^c} +(\mathbf{11}'/I)\mathbf{A}$ and, we estimate $\mathbf{A^c}$ by the already obtained $\hat{\mathbf{A}}^\mathbf{c}$, so that the only unknown part is $\mathbf{1}'\mathbf{A}/I$, which we denote as $\mathbf{a}'$. Then it remains to solve the minimization problem in (26), hence to minimize

$$\begin{aligned} f(\mathbf{a},\boldsymbol{\mu}_\mathbf{b}) &= \| \mathbf{X}_a - (\hat{\mathbf{A}}^\mathbf{c}+\mathbf{1a}')\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{b}')\|^2 \\ &= \| \mathbf{X}_a - \hat{\mathbf{A}}^\mathbf{c}\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{b}')\|^2 \end{aligned}$$

which by permuting the matrices into B-mode matricized form, can be written as

$$f(\mathbf{a},\boldsymbol{\mu}_\mathbf{b}) = \| \mathbf{X}_b - \hat{\mathbf{B}}\hat{\mathbf{G}}_b(\hat{\mathbf{A}}^{\mathbf{c}\prime}\otimes\hat{\mathbf{C}}') - \hat{\mathbf{B}}\hat{\mathbf{G}}_b(\mathbf{a1}'\otimes\hat{\mathbf{C}}') - \boldsymbol{\mu}_\mathbf{b}(\mathbf{1}'\otimes\mathbf{1}')\|^2. \tag{28}$$

Minimizing f($\mathbf{a}$,$\boldsymbol{\mu}_\mathbf{b}$) over $\boldsymbol{\mu}_\mathbf{b}$ gives, via linear regression, and using that $\hat{\mathbf{A}}^{\mathrm{c}}{}'\mathbf{1}=\mathbf{0}$,

$$\boldsymbol{\mu}_\mathbf{b} = \bar{\mathbf{x}}_b - \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\mathbf{a}\otimes\hat{\mathbf{C}}'\mathbf{1}/K), \quad (29)$$

and upon substituting (29) for $\boldsymbol{\mu}_\mathbf{b}$ in (28), we end up with

$$\mathrm{f}(\mathbf{a}) = \| \mathbf{X}_\mathrm{b} - \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\hat{\mathbf{A}}^{\mathrm{c}}{}'\otimes\hat{\mathbf{C}}') - \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\mathbf{a}\mathbf{1}'\otimes\hat{\mathbf{C}}')$$

$$- \bar{\mathbf{x}}_b(\mathbf{1}'\otimes\mathbf{1}') + \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\mathbf{a}\otimes\hat{\mathbf{C}}'\mathbf{1}/K)(\mathbf{1}'\otimes\mathbf{1}')\|^2$$

$$= \| \mathbf{X}_\mathrm{b} - \bar{\mathbf{x}}_b(\mathbf{1}'\otimes\mathbf{1}') - \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\hat{\mathbf{A}}^{\mathrm{c}}{}'\otimes\hat{\mathbf{C}}') - \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\mathbf{a}\mathbf{1}'\otimes\hat{\mathbf{C}}') + \hat{\mathbf{B}}\,\hat{\mathbf{G}}_b(\mathbf{a}\mathbf{1}'\otimes\hat{\mathbf{C}}'\mathbf{1}\mathbf{1}'/K)\|^2.$$

Backpermuting this expression into A-mode matricized form we find

$$\mathrm{f}(\mathbf{a}) = \| \mathbf{X}_\mathrm{a} - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}') - \hat{\mathbf{A}}^{\mathrm{c}}\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') +$$

$$\mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\mathbf{1}\mathbf{1}'/K\otimes\hat{\mathbf{B}}'))\|^2$$

$$= \| \mathbf{X}_\mathrm{a} - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}') - \hat{\mathbf{A}}^{\mathrm{c}}\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{\mathrm{c}}{}'\otimes\hat{\mathbf{B}}')\|^2,$$

from which now $\mathbf{a}$ can be solved by Penrose (1956) regression as

$$\mathbf{a}' = (\mathbf{1}/I)'(\mathbf{X}_\mathrm{a} - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}') - \hat{\mathbf{A}}^{\mathrm{c}}\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}'))(\hat{\mathbf{C}}^{\mathrm{c}}\otimes\hat{\mathbf{B}})\hat{\mathbf{G}}_a{}'(\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{\mathrm{c}}{}'\hat{\mathbf{C}}^{\mathrm{c}}\otimes\hat{\mathbf{B}}'\hat{\mathbf{B}})\hat{\mathbf{G}}_a{}')^{-1}$$

$$= (\mathbf{1}'\mathbf{X}_\mathrm{a}/I)(\hat{\mathbf{C}}^{\mathrm{c}}\otimes\hat{\mathbf{B}})\hat{\mathbf{G}}_a{}'(\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{\mathrm{c}}{}'\hat{\mathbf{C}}^{\mathrm{c}}\otimes\hat{\mathbf{B}}'\hat{\mathbf{B}})\hat{\mathbf{G}}_a{}')^{-1}, \quad (28)$$

where it was used that $(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}')(\hat{\mathbf{C}}^{\mathrm{c}}\otimes\hat{\mathbf{B}}) = (\mathbf{1}'\hat{\mathbf{C}}^{\mathrm{c}}\otimes\bar{\mathbf{x}}_b{}'\hat{\mathbf{B}}) = \mathbf{0}$, because $\mathbf{1}'\hat{\mathbf{C}}^{\mathrm{c}} = \mathbf{0}$, and $(\mathbf{1}/I)'\,\hat{\mathbf{A}}^{\mathrm{c}} = \mathbf{0}$, because $\mathbf{1}'\hat{\mathbf{A}}^{\mathrm{c}} = \mathbf{0}$. Note that, in case of singularity, the inverse should be replaced by the Moore-Penrose inverse.

In the above procedure, it was chosen to remove the offset term by centering across mode A. An alternative possibility would be to center across mode C. Then, switching roles of the A- and C- modes throughout, a different procedure for successively fitting the E1 model can be derived fully analogously, which, however, will lead to different parameter estimates. Here we only considered the procedure based on A-mode centering, in particular because centering data across subjects seems to be the most common choice in practice.

3.2.2. Successive Fitting of the E2 model

The procedure for successive fitting of the E2 model differs from that for fitting the E1 model only with respect to Step 3. To fit the E2 model, in Step 3, we minimize

$$f(\mathbf{A},\boldsymbol{\mu}_\mathbf{b},\boldsymbol{\mu}_\mathbf{c})=\| \mathbf{X}_\mathrm{a} – \mathbf{A}\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') – \mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{b}') - \mathbf{1}(\boldsymbol{\mu}_\mathbf{c}'\otimes\mathbf{1}')\|^2 \tag{30}$$

subject to the constraints $\mathbf{A}=\hat{\mathbf{A}}^\mathrm{c}+\mathbf{1a}'$ and $\mathbf{1}'\boldsymbol{\mu}_\mathbf{c}=0$.

This boils down to minimizing

$$f(\mathbf{a},\boldsymbol{\mu}_\mathbf{b},\boldsymbol{\mu}_\mathbf{c}) = \| \mathbf{X}_\mathrm{a} - \hat{\mathbf{A}}^\mathrm{c}\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') – \mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{b}') -\mathbf{1}(\boldsymbol{\mu}_\mathbf{c}'\otimes\mathbf{1}')\|^2. \tag{31}$$

The optimal $\boldsymbol{\mu}_\mathbf{b}$ is, as in the previous section, obtained via cycling permutation and regression, and given by

$$\boldsymbol{\mu}_\mathbf{b} = \bar{\mathbf{x}}_b - \hat{\mathbf{B}}\hat{\mathbf{G}}_b(\mathbf{a}\otimes\hat{\mathbf{C}}'\mathbf{1}/K). \tag{32}$$

The reason that the expressions for $\boldsymbol{\mu}_\mathbf{b}$ are equal for the E1 and E2 models, whereas the loss functions do differ, is that the difference in loss functions only pertains to the term $\mathbf{1}(\boldsymbol{\mu}_\mathbf{c}'\otimes\mathbf{1}')$, which upon cyclic permutation is written as $\mathbf{1}(\mathbf{1}'\otimes\boldsymbol{\mu}_\mathbf{c}')$, and hence would add a term $\mathbf{1}(1\otimes\boldsymbol{\mu}_\mathbf{c}'\mathbf{1}/K)$, which, however, is zero due to the identification constraint on $\boldsymbol{\mu}_\mathbf{c}$. Substituting (32) for $\boldsymbol{\mu}_\mathbf{b}$ in (31), analogously to the previous section, we end up with

$$f(\mathbf{a},\boldsymbol{\mu}_\mathbf{c}) = \| \mathbf{X}_\mathrm{a}- \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b') - \hat{\mathbf{A}}^\mathrm{c}\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}')- \mathbf{1a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{\mathrm{c}\prime}\otimes\hat{\mathbf{B}}') -\mathbf{1}(\boldsymbol{\mu}_\mathbf{c}'\otimes\mathbf{1}')\|^2,$$

which we permute into C-mode matricized form as

$$f(\mathbf{a},\boldsymbol{\mu}_\mathbf{c}) = \| \mathbf{X}_c - \mathbf{1}(\bar{\mathbf{x}}_b{}'\otimes\mathbf{1}') - \hat{\mathbf{C}}\,\hat{\mathbf{G}}_c(\hat{\mathbf{B}}'\otimes\hat{\mathbf{A}}^{c\prime}) - \hat{\mathbf{C}}^c\hat{\mathbf{G}}_c(\hat{\mathbf{B}}'\otimes\mathbf{a}\mathbf{1}') - \boldsymbol{\mu}_\mathbf{c}(\mathbf{1}'\otimes\mathbf{1}')\|^2.$$

Minimizing this function of $\boldsymbol{\mu}_\mathbf{c}$ by regression, temporarily ignoring the identification constraint, we get

$$\begin{aligned}\boldsymbol{\mu}_\mathbf{c} &= \bar{\mathbf{x}}_c - \bar{x}\,\mathbf{1} - \hat{\mathbf{C}}^c\hat{\mathbf{G}}_c(\hat{\mathbf{B}}'\mathbf{1}/J\otimes\mathbf{a})\\ &= \bar{\mathbf{x}}_c{}^{\mathbf{c}} - \hat{\mathbf{C}}^c\hat{\mathbf{G}}_c(\hat{\mathbf{B}}'\mathbf{1}/J\otimes\mathbf{a}),\end{aligned} \qquad (33)$$

where it has been used that $\hat{\mathbf{C}}\,\hat{\mathbf{G}}_c(\hat{\mathbf{B}}'\otimes\hat{\mathbf{A}}^{c\prime})(\mathbf{1}'\otimes\mathbf{1}) = \mathbf{0}$; it should be noted that the identification constraint $\mathbf{1}'\boldsymbol{\mu}_\mathbf{c}=0$ is satisfied by (33) even though it was not explicitly imposed.

Finally, substituting (32) for $\boldsymbol{\mu}_\mathbf{b}$ and (33) for $\boldsymbol{\mu}_\mathbf{c}$ in (31), and repeatedly using cyclic permutations as in the previous section, we obtain

$$\begin{aligned}f(\mathbf{a}) &= \| \mathbf{X}_a - \hat{\mathbf{A}}^c\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}')\\ &\quad + \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\mathbf{1}\mathbf{1}'/K\otimes\hat{\mathbf{B}}')) - \mathbf{1}(\bar{\mathbf{x}}_c{}^{c\prime}\otimes\mathbf{1}') + \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}'\mathbf{1}\mathbf{1}'/J))\|^2\\ &= \|\mathbf{X}_a - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}') - \mathbf{1}(\bar{\mathbf{x}}_c{}^{c\prime}\otimes\mathbf{1}') - \hat{\mathbf{A}}^c\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}') - \mathbf{1}\mathbf{a}'\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{c\prime}\otimes\hat{\mathbf{B}}^{c\prime})\|^2,\end{aligned}$$

which can be minimized over $\mathbf{a}$ by Penrose (1956) regression:

$$\begin{aligned}\mathbf{a}' &= (\mathbf{1}/I)'(\mathbf{X}_a - \mathbf{1}(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}') - \mathbf{1}(\bar{\mathbf{x}}_c{}^{c\prime}\otimes\mathbf{1}') - \hat{\mathbf{A}}^c\,\hat{\mathbf{G}}_a(\hat{\mathbf{C}}'\otimes\hat{\mathbf{B}}'))\\ &\qquad (\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^c)\hat{\mathbf{G}}_a{}'(\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{c\prime}\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^{c\prime}\hat{\mathbf{B}}^c)\hat{\mathbf{G}}_a{}')^{-1}\\ &= (\mathbf{1}'\mathbf{X}_a/I)(\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^c)\hat{\mathbf{G}}_a{}'(\hat{\mathbf{G}}_a(\hat{\mathbf{C}}^{c\prime}\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^{c\prime}\hat{\mathbf{B}}^c)\hat{\mathbf{G}}_a{}')^{-1},\end{aligned} \qquad (31)$$

where it was used that $(\mathbf{1}'\otimes\bar{\mathbf{x}}_b{}')(\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^c) = \mathbf{0}$, $(\bar{\mathbf{x}}_c{}^{c\prime}\otimes\mathbf{1}')(\hat{\mathbf{C}}^c\otimes\hat{\mathbf{B}}^c) = \mathbf{0}$, and $\mathbf{1}'\hat{\mathbf{A}}^c=0$. Again, in case of singularity, the inverse should be replaced by the Moore-Penrose inverse.