Poster presented at TRICAP '97, a conference on three-way methods in chemistry and psychology, Lake Chelan, Washington, May, 1997.

An "Extended PARAFAC Model" Incorporating Singly-Subscripted Constants: Theory and Application

Richard A. Harshman and Margaret E. Lundy
Dept. of Psychology, Univ. of Western Ontario
London, Ontario, Canada

Theory

(For a more detailed discussion of these issues, See Harshman and Lundy, 1984, Chapter 6 of Law et al.)

The problem of additive offsets

Factor analysis models require ratio-scale measurements in order to recognize proportional multiplicative effects of factor loading changes across levels of each mode.

Chemical data sometimes contain unwanted offsets, additive measurement biases, background constants, etc. that make true zero different from measured zero. Psychological data are often much worse; they are rarely ratio scale in their raw form. Variables like subjective ratings, IQ, strength of reinforcement, etc. do not have a clear (or at least known) zero point. Such data would seem incompatible with a factor analytic model.

The usual solution: centering (removing variable means)

In the two-way case, this baseline problem is not usually noticed because of the use of correlations, covariances, or deviation scores. All such data preprocessing subtracts out the mean level of each variable, and along with it, the mean of the constant offset—which of course removes the offset itself.

For three-way arrays, several centering operations might be needed—such as centering data within each row, and /or within each column, and/or within each 'tube'. Surprisingly, the issue of centering turns out to be more complex for three-way arrays. For example, centering should not be across entire 'slabs' or for the array as a whole, since it disturbs the zero points of the latent contributions of individual factors. We have analyzed these effects more formally elsewhere (Harshman & Lundy, 1984, Chapter 6 of Law et al.).

Disadvantages and limitations of centering

Centering that is mathematically appropriate (i.e., removal of 'fiber' means) will eliminate unwanted biases that are constant across the elements of the centered fibers. But it also subtracts out the means of the factor loadings for the factors (in the loading table for the centered mode). This removes the true zero point from factor loadings and replaces it with the factor mean; loadings become redefined in terms of deviations from the factor mean.

An extended Parafac model

As an alternative, we have defined an 'extended Parafac' model which incorporates added sets of constants to account for the offsets in the data. The model as originally proposed included both singly-subscripted and double-subscripted constants. In this poster, we concentrate on an extended model that only includes additional terms that are singly-subscripted:

(1)
$$x_{ijk} = \sum_r a_{ir} b_{jr} c_{jr} + {}_a h_i + {}_b h_j + {}_c h_k$$

We discuss here a simple method of fitting the extended model (1), and an example application.

Estimation

The new constant terms h_i, h_j and h_k can be considered as three additional factors being fit to the data. These added factors are constrained to have constant loadings in two of the three modes. In a two-factor case, the extended factor loading tables would look like this:

Mode A

a_{11}	a_{12}	${}_a h_1$	1	1
a_{21}	a_{22}	${}_a h_2$	1	1
a_{31}	a_{32}	${}_a h_3$	1	1

Mode B

b_{11}	b_{12}	1	${}_b h_1$	1
b_{21}	b_{22}	1	${}_b h_2$	1
b_{31}	b_{32}	1	${}_b h_3$	1

Mode C

c_{11}	c_{12}	1	1	${}_c h_1$
c_{21}	c_{22}	1	1	${}_c h_2$
c_{31}	c_{32}	1	1	${}_c h_3$

Uniqueness

The extra factors that represent the singly-subscripted constants h_i, h_j and h_k do not fulfill all the conditions normally required for uniqueness. Specifically, the two factors that are constant in any given mode would normally cause an indeterminacy of rotation in the subspace that they span. However, these factors are constrained to be constant down the column. Consequently, any linear transformation that would combine any of these three extra factors with either other constant factors or other 'standard' factors would involve mixing some non-constant loadings in with constant ones, which would violate the constancy requirements for the three extra factors (in at least one mode, if not more). The uniqueness is maintained by the constraint conditions.

Example Application

To demonstrate the procedure, we apply the extended model (1) to data that have been previously analyzed by the traditional method (using centering of Mode A and Mode B) to remove unwanted offsets. This is the 'Cars and Stars' dataset—consisting of ratings of 12 automobiles and 12 celebrities (and 'yourself') on 39 bipolar rating scales (see table) by 34 people (see Harshman & Lundy, 1984b, Appendix C of Law et al.).

With the new analysis, no centering is done and instead the three additive offsets are estimated. The tables of factor loadings for Modes A, B, and C are shown below for the 2D solution (Mode C loadings correspond to subject weights or sensitivities for the two dimensions).

Note that we here report a 2D solution rather than the 3D solution that was obtained with the centered data. In our reanalysis, the 3D solution was somewhat degenerate. We interpret this to mean that there are additional confounding components in these data besides the one-way constants. As demonstrated algebraically in Harshman and Lundy (1984a) each centering also removes some of the doubly subscripted constant terms. Thus, several sets of doubly-subscripted constants were removed by the centering done before the original analysis, but were not removed by our extended analysis.

Advantage of not centering

Although this example has demonstrated some limitations of the extended model (if there are additional confounding components in the data besides the singly-subscripted constants) it also shows the advantages of this type of analysis. The baseline offsets for each level of each mode are recovered in the last three columns of the factor loading tables. Note that the Mode B loadings of the bipolar rating scales (with options labeled 1 at the bottom of the scale and 7 at the top) generally seem to have a constant offset of between 3.5 and 4.5. this makes sense, since it shows that the subjects were not treating the scale as representing strengths of 1 to 7, but instead considered the zero point of the scale to be around the middle.

The nicest effect of estimating the additive constants instead of centering is that the factor loadings are not centered, and so can be interpreted as ratio-scale values for which a loading of zero represents a zero amount of effect of the given factor at that level of the data.

In the original analysis of these data, the stimulus loadings appeared as deviation scores around the centroid of the stimulus space. This made the stimuli appear to be located at the ends of bipolar stimulus dimensions. The new analysis shows that Factor 1 is completely unipolar, and Factor 2 is unipolar except for two stimuli. This changes the interpretation of the factors, since stimuli at the low end are now seen to not involve the stimulus dimension rather than be interpreted as being at the opposite extreme of some stimulus dimension.

A=

0.8487	0.9386	3.5908	1.0	1.0
1.4593	1.5629	3.6190	1.0	1.0
0.1521	0.5545	3.9051	1.0	1.0
0.0195	-0.5361	3.8241	1.0	1.0
1.5406	1.6095	3.9395	1.0	1.0
1.5303	1.9834	4.0484	1.0	1.0
0.2661	0.1411	3.8013	1.0	1.0
1.2821	1.1741	3.6177	1.0	1.0
0.7825	0.3963	3.8917	1.0	1.0
0.1882	-0.4906	3.7929	1.0	1.0
1.2359	1.1402	3.8671	1.0	1.0
0.1882	0.1791	3.7157	1.0	1.0
1.7817	0.9216	3.7241	1.0	1.0
1.4433	1.3942	3.9354	1.0	1.0
1.2838	0.6597	3.6567	1.0	1.0
1.2945	2.0799	4.1445	1.0	1.0
1.3323	1.2546	3.9256	1.0	1.0
1.5668	0.0289	3.8486	1.0	1.0
1.9664	1.6581	4.1020	1.0	1.0
1.5817	0.4994	3.7906	1.0	1.0
2.0894	0.9192	3.9056	1.0	1.0
1.3102	2.1742	4.1557	1.0	1.0
1.0054	0.6278	3.8646	1.0	1.0
1.1294	0.8469	3.9577	1.0	1.0
0.9276	1.0131	3.8458	1.0	1.0
		^^^		
		Mode A offsets		

B=

1.0087	-0.8008	1.0	-0.9564	1.0
0.9976	-1.5011	1.0	-0.4289	1.0
0.7151	-1.6173	1.0	0.6508	1.0
1.3393	-0.7791	1.0	-1.3434	1.0
-1.2113	0.6192	1.0	1.2703	1.0
-1.7932	0.8846	1.0	1.9327	1.0
1.2754	-1.4444	1.0	-0.5606	1.0
-1.4570	0.6981	1.0	1.2707	1.0
1.0363	-1.4835	1.0	-0.1949	1.0
1.0758	-0.8791	1.0	-1.1923	1.0
-0.9217	1.0357	1.0	1.4392	1.0
-0.4468	1.0080	1.0	-0.4139	1.0
0.2469	-1.4930	1.0	0.5794	1.0
-0.0672	-1.0997	1.0	0.8334	1.0
0.8012	-0.7281	1.0	-1.4123	1.0
1.2411	-1.0243	1.0	-1.2073	1.0
-0.6552	1.0702	1.0	0.5938	1.0
1.0742	-0.9717	1.0	-0.9572	1.0
0.0936	-0.5524	1.0	0.1749	1.0
1.2340	-0.6824	1.0	-1.5518	1.0
1.0095	-1.5209	1.0	-0.4357	1.0
0.3271	-1.0704	1.0	0.1240	1.0
-1.1566	0.1172	1.0	1.7194	1.0
1.2655	-1.3736	1.0	-0.7458	1.0
-1.3109	0.5300	1.0	1.5101	1.0
-0.8072	1.1426	1.0	0.1183	1.0
0.5205	-0.7872	1.0	-0.4416	1.0
-1.1989	0.5654	1.0	1.1714	1.0
1.0668	-0.8781	1.0	-0.9236	1.0
0.3595	0.1718	1.0	-1.0484	1.0
-1.4529	0.6690	1.0	2.5915	1.0
-0.4214	-0.8387	1.0	0.9396	1.0
-0.5795	-0.2275	1.0	0.5152	1.0
1.4009	-0.8937	1.0	-1.1210	1.0
0.7544	-0.5547	1.0	-0.7824	1.0
-0.0433	1.1199	1.0	-0.7375	1.0
1.1840	-1.0263	1.0	-0.6415	1.0
0.3279	-1.4678	1.0	0.6270	1.0
1.2365	-0.8638	1.0	-0.9652	1.0
			^^^	
			Mode B offsets	

C=

1.0096	0.9712	1.0	1.0	-0.0804
1.1757	1.1507	1.0	1.0	0.0747
0.8631	1.2106	1.0	1.0	0.1775
0.9483	0.6451	1.0	1.0	-0.0974
1.0101	1.2028	1.0	1.0	-0.0117
1.0442	1.0279	1.0	1.0	0.0939
1.0937	1.0355	1.0	1.0	0.1694
0.9114	0.7700	1.0	1.0	0.0282
1.1121	0.8398	1.0	1.0	0.0715
0.7304	0.9249	1.0	1.0	-0.1361
1.3384	1.1606	1.0	1.0	0.1473
1.3893	1.1982	1.0	1.0	0.1604
1.1038	1.1890	1.0	1.0	0.1046
1.2986	1.0598	1.0	1.0	0.0850
1.3826	1.1580	1.0	1.0	0.1913
0.9420	1.0425	1.0	1.0	-0.0037
0.8656	0.6380	1.0	1.0	-0.0224
0.8141	0.6520	1.0	1.0	0.0731
1.4752	1.3615	1.0	1.0	0.3034
1.2933	1.3703	1.0	1.0	-0.1711
1.0273	1.1169	1.0	1.0	-0.0660
0.5839	1.0612	1.0	1.0	0.0452
0.9118	1.0298	1.0	1.0	0.1512
0.6212	1.2410	1.0	1.0	-0.1930
0.5291	0.6059	1.0	1.0	-0.2899
0.7494	0.9032	1.0	1.0	-0.1604
0.8067	0.8430	1.0	1.0	0.0078
0.9293	0.8065	1.0	1.0	0.0415
1.0953	0.8673	1.0	1.0	-0.0624
1.2455	1.0085	1.0	1.0	0.0259
0.3380	0.7709	1.0	1.0	-0.5063
0.6017	0.7996	1.0	1.0	-0.2632
0.5804	0.9381	1.0	1.0	-0.0164
0.9082	0.6401	1.0	1.0	0.0382
				^^^
				Mode C offsets