This was originally published in 1972 in *UCLA Working Papers in Phonetics*, No. *22*, pp. 111-117. Some typographical, etc. errors in the original have been corrected here.

Determination and Proof of Minimum Uniqueness Conditions for PARAFAC1

Richard A. Harshman

Introduction

The most important single fact about PARAFAC1 is that it provides a unique solution when the data fulfills certain necessary conditions. It is obviously important, then, to determine just what these conditions are. This problem has been approached in two different ways: empirically and mathematically.

Empirically, the necessary and sufficient conditions for uniqueness are explored by analyzing “synthetic” data. This synthetic data is generated by computer so that the underlying structure and mathematical properties of each data set are known exactly. The computer starts with factors and their relationships and then generates the data that those factors would produce. In testing each data set, the PARAFAC1 program is used to repeatedly analyze the data from different starting positions. When, for a given data set the PARAFAC1 program converges on the same solution from all different starting positions, it is concluded that the PARAFAC1 model provides a unique description of that set of data. From a number of such experiments on different types of data sets (i.e., different sized data sets, different types of factors underlying the data sets, etc.) it is possible to observe the common properties of those data sets which have unique solutions, and thus generalize to those conditions which any data set must fulfill in order to have a unique solution. Further, it is possible to observe any regular patterns in the types of data that yield non-unique solutions and generalize to typical ways in which data sets can be “inadequate” to uniquely determine a solution. When non-unique solutions occur, one can observe the ways in which uniqueness can partially or completely “break down”, and relate each type of non-uniqueness to a corresponding type of data “inadequacy”. The results of such empirical studies of PARAFAC1 are described in Harshman (1970).

Mathematically, the conditions for uniqueness are explored by attempting to derive them from the algebraic properties of the PARAFAC1 model. If all our questions could be answered algebraically, there would be no need for the empirical studies (except to study the quirks of a particular algorithm, or the effects of different amounts and kinds of "noise" added to the data). Unfortunately, the algebraic properties of PARAFAC1 are relatively subtle, and most of all the mathematical insights are yet to be discovered. The proof which is given below should be viewed as one more step in an ongoing effort to gain mathematical understanding of the subtler properties of PARAFAC1.

The first uniqueness proof for PARAFAC1 was discovered by Robert Jennrich of the UCLA Department of Mathematics. It was published, along with the results of a number of empirical studies on PARAFAC1, in Harshman (1970). Jennrich’s Uniqueness Theorem showed that the potential for a unique solution exists when one has as many parallel two-way data sets as there are factors underlying the data. For example, if there are three factors underlying a set of measurements, then one would need to have at least three measurements, on three objects, repeated on three different occasions, to fit the uniqueness conditions required by Jennrich's Theorem. (Certain other conditions were also required which are described in Harshman, 1970).

Empirical results, however, suggested that the conditions required by Jennrich's theorem were stronger than were really necessary in order to determine uniqueness. It was empirically possible to uniquely extract at least 6 factors from data sets having only three repetitions of the two-way tables (three different "occasions"). The *minimal* conditions of uniqueness could not be determined, however, since convergence became too slow if the number of occasions was too small compared to the number of factors. (See Harshman, 1970, chapter 4.)

The proof which follows shows that *two* occasions can be sufficient to uniquely determine any number of factors, provided that the factors change size from the first to the second occasions and that the percent change of each factor is different than that of the other factors. The proof reveals some of the necessary and sufficient conditions for uniqueness. These findings shed light, for example, on patterns of "partial breakdown" of uniqueness by clarifying how and why uniqueness fails, as well as when it fails.

<u>Theorem</u>

If

(1) $$\mathbf{X}_1 = \mathbf{A}\mathbf{D}_1\mathbf{B}'$$

(2) $$\mathbf{X}_2 = \mathbf{A}\mathbf{D}_2\mathbf{B}'$$

where **A** and **B** are n by l matrices which are non-horizontal ($n \geq l$) and "basic" (of rank l) and $\mathbf{D}_1$ and $\mathbf{D}_2$ are nonsingular matrices such that

(2b) $$\mathbf{D}_1\mathbf{D}_2^{-1} = \mathbf{D}_\text{p}$$

where $\mathbf{D}_\text{p}$ has distinct diagonal elements.

And if there exists some alternate representation of $\mathbf{X}_1$ and $\mathbf{X}_2$, such as

(3) $$\mathbf{X}_1 = \overset{*}{\mathbf{A}}\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}'$$

(4) $$\mathbf{X}_2 = \overset{*}{\mathbf{A}}\overset{*}{\mathbf{D}}_2\overset{*}{\mathbf{B}}'$$

then

(5) $$\overset{*}{\mathbf{A}} = \mathbf{A}\boldsymbol{\Delta}_\text{x}$$

(6) $$\overset{*}{\mathbf{B}} = \mathbf{B}\boldsymbol{\Delta}_\text{y}$$

where $\boldsymbol{\Delta}_\text{x}$ is some combination of a diagonal and a permutation matrix, and the same is true of $\boldsymbol{\Delta}_\text{y}$ (i.e., $\boldsymbol{\Delta}_\text{x}$, $\boldsymbol{\Delta}_\text{y}$ are nonsingular with one nonzero element in each column).

<u>Proof</u>

We start with our expressions for $\mathbf{X}_1$ and $\mathbf{X}_2$ from (1) and (2), respectively.

Now let us seek the restrictions imposed on any alternate **A** matrix. We can describe any alternate **A** matrix as our original **A** matrix subjected to some transformation $\mathbf{T}_A$ as follows,

(7) $$\overset{*}{\mathbf{A}} = \mathbf{A}\mathbf{T}_A \quad .$$

This is possible since the columns of **A** and $\overset{*}{\mathbf{A}}$ must span the same space (e.g., $\mathbf{X}_1$). In fact, we can write an expression for **T** using equations (1), (2), (3), and (4). Since

(8) $$\overset{*}{\mathbf{A}}\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}' = \mathbf{A}\mathbf{D}_1\mathbf{B}' \;,$$

it then follows that by post-multiplying both sides by $(\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}')^{+}$, the right pseudo-inverse of $(\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}')$, i.e., $(\overset{*}{\mathbf{B}}(\overset{*}{\mathbf{B}}'\overset{*}{\mathbf{B}})^{-1})\overset{*}{\mathbf{D}}_1^{-1}$

(9) $$\overset{*}{\mathbf{A}}\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}'(\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}')^{+} = \mathbf{A}\mathbf{D}_1\mathbf{B}'(\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}')^{+}$$

(10) $$\overset{*}{\mathbf{A}} = \mathbf{A}(\mathbf{D}_1\mathbf{B}'(\overset{*}{\mathbf{B}}')^{+}\overset{*}{\mathbf{D}}_1^{-1}) \quad .$$

Let $\mathbf{T}_A = \mathbf{D}_1\mathbf{B}'(\overset{*}{\mathbf{B}}')^{+}\overset{*}{\mathbf{D}}_1^{-1}$ and we obtain, by substitution, equation (7).

In a similar fashion, we obtain

(11) $$\overset{*}{\mathbf{B}} = \mathbf{B}\mathbf{T}_B \quad .$$

Now since $\mathbf{T}_A\mathbf{T}_A^{-1}=\mathbf{I}$, $\mathbf{T}_B\mathbf{T}_B^{-1}=\mathbf{I}$, we write

(12) $$\mathbf{X}_1 = \mathbf{A}(\mathbf{T}_A\mathbf{T}_A^{-1})\,\mathbf{D}_1\,(\mathbf{T}_B'^{-1}\,\mathbf{T}_B')\mathbf{B}'$$

(13) $$\mathbf{X}_2 = \mathbf{A}(\mathbf{T}_A\mathbf{T}_A^{-1})\,\mathbf{D}_2\,(\mathbf{T}_B'^{-1}\,\mathbf{T}_B')\mathbf{B}'$$

or, simply regrouping parentheses,

(14) $$\mathbf{X}_1 = (\mathbf{A}\mathbf{T}_A)(\mathbf{T}_A^{-1}\,\mathbf{D}_1\,\mathbf{T}_B'^{-1})(\mathbf{T}_B'\,\mathbf{B}')$$

(15) $$\mathbf{X}_2 = (\mathbf{A}\mathbf{T}_A)(\mathbf{T}_A^{-1}\,\mathbf{D}_2\,\mathbf{T}_B'^{-1})(\mathbf{T}_B'\,\mathbf{B}')$$

which can be rewritten, substituting $\overset{*}{\mathbf{A}} = \mathbf{A}\mathbf{T}_A$, $\overset{*}{\mathbf{B}}' = \mathbf{T}_B'\,\mathbf{B}'$ as

(16) $$\mathbf{X}_1 = \overset{*}{\mathbf{A}}(\mathbf{T}_A^{-1}\,\mathbf{D}_1\,\mathbf{T}_B'^{-1})\overset{*}{\mathbf{B}}'$$

(17) $$\mathbf{X}_2 = \overset{*}{\mathbf{A}}(\mathbf{T}_A^{-1}\,\mathbf{D}_2\,\mathbf{T}_B'^{-1})\overset{*}{\mathbf{B}}' \quad .$$

Now since equations (3) and (16) both give expressions for $\mathbf{X}_1$

(18) $$\overset{*}{\mathbf{A}}\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}' = \overset{*}{\mathbf{A}}(\mathbf{T}_A^{-1}\,\mathbf{D}_1\,\mathbf{T}_B'^{-1})\overset{*}{\mathbf{B}}' \quad .$$

Using the fact that $\overset{*}{\mathbf{A}}$ and $\overset{*}{\mathbf{B}}$ are basic and thus have left pseudo-inverses, we can pre- and post-multiply by appropriate inverses, obtaining

(19) $$(\overset{*}{\mathbf{A}}{}^{+}\overset{*}{\mathbf{A}})\overset{*}{\mathbf{D}}_1(\overset{*}{\mathbf{B}}{}'\overset{*}{\mathbf{B}}{}'^{+}) = (\overset{*}{\mathbf{A}}{}^{+}\overset{*}{\mathbf{A}})(\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_1\mathbf{T}_{\mathrm{B}}'^{-1})(\overset{*}{\mathbf{B}}{}'\overset{*}{\mathbf{B}}{}'^{+})$$

or

(20) $$\overset{*}{\mathbf{D}}_1 = \mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_1\mathbf{T}_{\mathrm{B}}'^{-1} \quad .$$

In precisely the same fashion, by operating on equations for $\mathbf{X}_2$ (i.e., equations (4) and (17)) we can obtain

(21) $$\overset{*}{\mathbf{D}}_2 = \mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_2\mathbf{T}_{\mathrm{B}}'^{-1} \quad .$$

Now equations (20) and (21) can be premultiplied on both sides as follows

(22) $$\overset{*}{\mathbf{D}}_1^{-1}\,\overset{*}{\mathbf{D}}_1 = \overset{*}{\mathbf{D}}_1^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\mathbf{D}_1\mathbf{T}_{\mathrm{B}}'^{-1})$$

(23) $$\overset{*}{\mathbf{D}}_2^{-1}\,\overset{*}{\mathbf{D}}_2 = \overset{*}{\mathbf{D}}_2^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\mathbf{D}_2\mathbf{T}_{\mathrm{B}}'^{-1})$$

or

(24) $$\mathbf{I} = \overset{*}{\mathbf{D}}_1^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_1\mathbf{T}_{\mathrm{B}}'^{-1})$$

(25) $$\mathbf{I} = \overset{*}{\mathbf{D}}_2^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_2\mathbf{T}_{\mathrm{B}}'^{-1}) \quad .$$

Setting these expressions equal to one another, we get

(26) $$\overset{*}{\mathbf{D}}_1^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_1\mathbf{T}_{\mathrm{B}}'^{-1}) = \overset{*}{\mathbf{D}}_2^{-1}(\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_2\mathbf{T}_{\mathrm{B}}'^{-1})$$

and post-multiplying both sides by $\mathbf{T}_{\mathrm{B}}'$ we get

(27) $$\overset{*}{\mathbf{D}}_1^{-1}\,\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_1 = \overset{*}{\mathbf{D}}_2^{-1}\,\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_2 \quad .$$

Pre-multiplying both sides by $\overset{*}{\mathbf{D}}_1$ and post-multiplying by $(\mathbf{D}_1)^{-1}$ we obtain

(28) $$\mathbf{T}_{\mathrm{A}}^{-1} = \overset{*}{\mathbf{D}}_1\,\overset{*}{\mathbf{D}}_2^{-1}\,\mathbf{T}_{\mathrm{A}}^{-1}\,\mathbf{D}_2\mathbf{D}_1^{-1} \quad .$$

Now $\mathbf{D}_2(\mathbf{D}_1)^{-1}=\mathbf{D}_{\mathrm{p}}$ from (2b) and we can let

(29) $$\mathbf{D}_{\mathrm{q}} = \overset{*}{\mathbf{D}}_1\,\overset{*}{\mathbf{D}}_2^{-1} \quad .$$

We know that $\mathbf{D}_{\mathrm{q}}$ is nonsingular since $\overset{*}{\mathbf{D}}_1,\ \overset{*}{\mathbf{D}}_2$ are nonsingular (if they were singular, the rank of $\overset{*}{\mathbf{A}}\,\overset{*}{\mathbf{D}}_1\overset{*}{\mathbf{B}}{}'$ would not be the rank of $\mathbf{X}_1$ and similarly for $\mathbf{X}_2$). So from (29) and (2b), we can rewrite (28) as

(30) $$\mathbf{T}_{\mathrm{A}}^{-1} = \mathbf{D}_{\mathrm{q}}\mathbf{T}_{\mathrm{A}}^{-1}\mathbf{D}_{\mathrm{p}} \quad .$$

Now this implies that $\mathbf{T}_{\mathrm{A}}^{-1}$ (and thus $\mathbf{T}_{\mathrm{A}}$) is diagonal and/or permutation. To see this, consider two elements in a given row of $\mathbf{T}_{\mathrm{A}}^{-1}$. Let us take t_{ab}^{-1} and t_{ac}^{-1} elements from row a and in column b and c. From (30) we can write the scalar equations

(31) $$t_{ab}^{-1} = (d_{aa}^{q})(t_{ab}^{-1})(d_{bb}^{p})$$

(32) $$t_{ac}^{-1} = (d_{aa}^{q})(t_{ac}^{-1})(d_{cc}^{p}) \quad .$$

Now if *both* t_{ab}^{-1} and t_{ac}^{-1} are nonzero we could simplify by dividing both sides of equation (31) by t_{ab}^{-1}, and both sides of (32) by t_{ac}^{-1}, such that

(33) $$1 = (d_{aa}^{q})(d_{bb}^{p})$$

(34) $$1 = (d_{aa}^{q})(d_{cc}^{p})$$

from which we would obtain

(35) $$(d_{aa}^{q})^{-1} = d_{bb}^{p}$$

(36) $$(d_{aa}^{q})^{-1} = d_{cc}^{p}$$

or

(37) $$d_{bb}^{p} = d_{cc}^{p} \quad .$$

But this contradicts our hypothesis that all diagonal elements of $\mathbf{D}_p$ are distinct. Evidently, we cannot take two nonzero elements from any row of $\mathbf{T}_A^{-1}$. But since $\mathbf{T}_A^{-1}$ is nonsingular, each row has at least one nonzero element. In the same fashion (but by taking two elements from the same *column*) we can show that each column has at most one nonzero element.

But any nonsingular matrix which has only one nonzero element in each row and column is either a diagonal or a permutation matrix, or some combination thereof. Hence, $\mathbf{T}_A^{-1}$ (and thus $\mathbf{T}_A$) is diagonal and/or permutation.

But $\mathbf{\Delta}_x$ is defined as some matrix which has just this form, so we can let

(38) $$\mathbf{\Delta}_x = \mathbf{T}_A$$

(39) $$\overset{*}{\mathbf{A}} = \mathbf{A}\mathbf{T}_A = \mathbf{A}\mathbf{\Delta}_x \quad .$$

In the same fashion using symmetry we can obtain

(40) $$\overset{*}{\mathbf{B}} = \mathbf{B}\mathbf{T}_B = \mathbf{B}\mathbf{\Delta}_y \quad .$$

We can see that when two elements of $\mathbf{D}_p$ are not distinct, the corresponding factors are not uniquely determined, while the remaining factors (with elements of $\mathbf{D}_p$ distinct from all other elements of $\mathbf{D}_p$) will *still* be uniquely determined. This explains how the uniqueness can break down “in stages” as discovered empirically in Harshman (1970, pp. 39-44).

Application to INDSCAL and other models

By taking the case where matrix **B**=**A**, the minimal conditions theorem applies to analysis of scalar-product and cross-product matrices, providing that one assumes orthogonal factors. Thus it applies to Carroll and Chang's INDSCAL model (Carroll & Chang, 1970), and provides minimal conditions of uniqueness for INDSCAL solutions. By the same type of interpretation, it provides a uniqueness proof for the orthogonal factor case of PARAFAC2 (for a description of PARAFAC2, see Harshman, 1972). The general oblique proof for PARAFAC2 has not yet been discovered (although progress has been made using reasoning which is along the same lines as the proof reported here).

Further work to be done

Of course, not all questions about the uniqueness of PARAFAC1 have been answered by this minimal conditions theorem. For example, it does not deal with the circumstance where some or all of the $\mathbf{D}_I$ matrices are of a lower rank than the full number of factors—i.e., when some factors have a zero influence on some occasions. Empirical results show that this circumstance need not interfere with uniqueness when there are a sufficient number of occasions. Just how many are necessary and sufficient has not been mathematically determined. Of course, if any two of the $\mathbf{D}_I$ matrices satisfy the conditions of the preceding theorem, then they will suffice to provide the unique solution. But when no two such $\mathbf{D}_I$ exist, uniqueness must be determined by a more complicated set of interdependencies. Jennrich's Uniqueness Theorem applies to this more complicated circumstance whenever there are as many "occasions" as factors. It thus establishes some possible sufficient conditions for uniqueness in such a case. In this respect Jennrich's theorem in "stronger" than the minimal conditions theorem presented here, although it does not handle the situations where there are more factors than occasions, or provide *minimal* conditions for uniqueness when two $\mathbf{D}_I$ are of full rank. In these respects the theorem presented here is "stronger".

Another interesting empirical finding which is not covered by either proof is the discovery that 10 factors can apparently be uniquely determined by an 8 by 8 data set. This would correspond to the case where the **A** and **B** matrices in the minimal conditions proof were horizontal rather than the current stipulation that they are square or vertical. Mathematical treatment of this number of factors in such a sized data set poses the problem of carrying through the proof without the pseudo-inverses of **A** and **B**. This, too, is a problem for further work in the future.

References

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika, 35*, 283-319.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an ‘explanatory’ multi-modal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1-84.

Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 30-44.

Acknowledgements

Without the constructive criticism of Robert Jennrich and David Cantor (UCLA Department of Mathematics), the proof would have been far less clear and rigorous. Their interest and help is greatly appreciated.