Alternating Least Squares Estimation for the Single Domain DEDICOM Model

*Richard A. Harshman**

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

DEDICOM (Harshman, 1978, 1981b) is a generalization of multidimensional scaling capable of analyzing *asymmetric* relationships among a set of things (e.g., the number of telephone calls between cities, confusions among stimuli, strength of implications among propositions, or preferences among products). However, existing estimation procedures for DEDICOM have several limitations: most of the procedures are only approximately-least-squares, and none of them are well suited for ignoring the (often problematical) diagonal cells of the input data matrix. In this memo, a method is proposed for overcoming these limitations, based on "implicit equation" and true "Alternating Least Squares" (ALS) methods of estimation. The major hurdle which has previously prevented use of ALS techniques for Single Domain DEDICOM has been the requirement that the row and column space of the least squares approximation be the same. However, this hurdle can be overcome by the use of partitioned matrix equations where the symmetric and skew components of the data are represented separately, but fit jointly. The necessary ALS equations are first worked out for the basic model and then extended to include estimation of multiplicative row and/or column bias parameters for DEDICOM. Hopefully, the development of ALS methods for DEDICOM will link it to the growing literature on these often highly successful estimation techniques, allowing expertise developed for other problems to be applied to analysis of asymmetrical relationships.

August 13, 1981

*Consultant to Bell Laboratories from University of Western Ontario.

Revision of unpublished technical memorandum.
Bell Labs, Murray Hill, NJ, 1980.

Alternating Least Squares Estimation for the Single Domain DEDICOM Model

*Richard A. Harshman**

Bell Laboratories
Murray Hill, New Jersey 07974

Recently, there has been a rapid growth of interest in the problem of analyzing data describing *asymmetric* relationships between the members of a set of objects. Consider for example, data describing the amount of telephone communication or migration of population between various cities, or the pattern of confusions among a number of stimuli, or the extent to which each of a number of statements or properties is judged to imply each of the others; these are all examples of systematically asymmetric patterns of relationships among a set of things, and in all these examples, there is potentially much to be learned from a detailed study of the asymmetrical, as well as the symmetrical aspects of the relationships. Traditional factor analytic and multidimensional scaling methods have difficulty adequately representing such data, since their basic models are formulated in terms of distances, scalar products, or similarly symmetric relationships. DEDICOM (which stands for DEcomposition into DIRectional COMponents) is a generalization of factor analysis and MDS which allows for representation of observed asymmetric relationships among objects in terms of underlying asymmetric relationships among latent components or types of objects. The model and method, along with applications, are discussed elsewhere (Harshman, 1975, 1978, 1981a, 1981b). In this article, interest will be focused on developing improved methods of fitting the model to data. Such improvements (it is hoped) will provide the ability to deal efficiently with the (often problematical) diagonal elements of the data, as well as any missing data, and facilitate the estimation of additional parameters which would allow analysis of a more general set of data types.

The DEDICOM model (as described below) bears a close resemblance to PARAFAC, INDSCAL, and other models for which the Alternating Least Squares (Wold, 1966) technique works well. Our experience with fitting these models indicates that if ALS could be applied to DEDICOM it would probably provide a rapidly convergent method of fitting the two-way case which was (1) able to efficiently handle missing values and ignore the diagonal values of the data; and (2) easily incorporate estimation of additional parameters for row bias terms, three-way arrays, etc. (An efficient method of ignoring the diagonal cells of the input data is particularly important, since, for most applications of DEDICOM, the diagonal cells are either missing or not appropriately fitted by the same model as the off-diagonal cells.)

The Problem and the Basis for its Solution. Consider a square n by n data matrix **X**, which describes the (generally asymmetric) relationships among n things. For example, the n things could be cities, and the element x_{ij} might describe the number of telephone calls from city i to city j; in general, for such data, $x_{ij} \neq x_{ji}$. Now the two-way, single domain version of DEDICOM would represent **X** as follows:

$$\mathbf{X} = \mathbf{ARA}' + \mathbf{E}$$

where **A** is a vertical n by q matrix of "loadings" or weights which show the relationship of each of the n things to q underlying types or influences (q < n). The matrix **R** is a q by q matrix of (generally asymmetric) relationships among these q latent types, and **E** is an n by n matrix of error terms. Given a particular data matrix **X** we seek a way of applying the ALS methods to estimate **A** and **R**.

---

*Consultant to Bell Laboratories from University of Western Ontario.

There has been one major roadblock to the application of ALS to the Single Domain DEDICOM model: the requirement that the left and right hand versions of the **A** matrix actually be the same. Since, in general, an asymmetric data matrix will not have the same row and column space, a simple unconstrained ALS fitting of different parameters for the left-hand and right-hand versions of **A** would not produce the common **A** needed by the single-domain DEDICOM model. Instead, unconstrained ALS would fit the Dual Domain model to the data. While a constrained method has seemed essential, it has not been obvious how to constrain the two **A** matrices to be the same, and still preserve the necessary property that each stage of the ALS procedure is a true conditional least-squares estimation.

The solution to this problem lies in realizing that a general asymmetric data matrix **X** is composed of two parts, each of which *does* have a row space which equals its corresponding column space. These two parts are the symmetric and skew-symmetric components of **X**. Furthermore, the symmetric and skew-symmetric parts are orthogonal to one another. Consequently, the problem can be simplified by observing that the **A** which provides an overall least squares best fit to these two orthogonal parts (when the two error sums of squares are taken together) necessarily provides a least-squares best fit to the data as a whole.

We develop the argument as follows. First, we decompose **X** into its symmetric and skew-symmetric components:

$$\mathbf{S} = \tfrac{1}{2}(\mathbf{X} + \mathbf{X}')$$

$$\mathbf{K} = \tfrac{1}{2}(\mathbf{X} - \mathbf{X}')$$

so that

$$\mathbf{X} = \mathbf{S} + \mathbf{K} \quad .$$

We note that $\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}\, k_{ij} = 0$ (orthogonality of the two components) and, as a consequence, $\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij}^2$ (additivity of their sums of squares). In fact, any skew-symmetric matrix $\overset{*}{\mathbf{K}}$ is orthogonal to any symmetric one $\overset{o}{\mathbf{S}}$ since $\overset{o}{s}_{ij}\,\overset{*}{k}_{ij} = -\overset{o}{s}_{ji}\overset{*}{k}_{ji}$, a fact which will prove useful below.

We now consider the DEDICOM least-squares approximation of **X**, namely

$$\hat{\mathbf{X}} = \mathbf{ARA}'.$$

We seek the **A** and **R** (for a given dimensionality q) such that $||\mathbf{X}-\hat{\mathbf{X}}||$ is minimized, where

$$||\mathbf{X} - \hat{\mathbf{X}}|| = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2 \quad .$$

We note that $\hat{\mathbf{X}}$ can also be decomposed into its symmetric and skew-symmetric components

$$\hat{\mathbf{X}} = \hat{\mathbf{S}} + \hat{\mathbf{K}}$$

and that the two components share the same underlying **A** matrix, differing only in terms of **R**, since

$$\hat{\mathbf{S}} = \tfrac{1}{2}(\hat{\mathbf{X}} + \hat{\mathbf{X}}') = \mathbf{A}\tfrac{1}{2}(\mathbf{R} + \mathbf{R}')\mathbf{A}' = \mathbf{A}\ \mathbf{R}_s\ \mathbf{A}'$$

$$\hat{\mathbf{K}} = \tfrac{1}{2}(\hat{\mathbf{X}} - \hat{\mathbf{X}}') = \mathbf{A}\tfrac{1}{2}(\mathbf{R} - \mathbf{R}')\mathbf{A}' = \mathbf{A}\ \mathbf{R}_k\ \mathbf{A}'$$

where $\mathbf{R}_s$ and $\mathbf{R}_k$ are the symmetric and skew components of **R**, i.e.,

$$\mathbf{R}_s = \tfrac{1}{2}(\mathbf{R} + \mathbf{R}')$$

$$\mathbf{R}_k = \tfrac{1}{2}(\mathbf{R} - \mathbf{R}') \quad .$$

By substitution,

$$||\mathbf{X} - \hat{\mathbf{X}}|| = ||(\mathbf{S} + \mathbf{K}) - (\hat{\mathbf{S}} + \hat{\mathbf{K}})||$$
$$||\mathbf{X} - \hat{\mathbf{X}}|| = ||(\mathbf{S} - \hat{\mathbf{S}}) + (\mathbf{K} - \hat{\mathbf{K}})|| \quad .$$

Now, because any symmetric matrix is orthogonal to any skew-symmetric matrix*, ($\mathbf{S}$ - $\hat{\mathbf{S}}$) is orthogonal to ($\mathbf{K}$ - $\hat{\mathbf{K}}$) and consequently

$$||\mathbf{X} - \hat{\mathbf{X}}|| = ||\mathbf{S} - \hat{\mathbf{S}}|| + ||\mathbf{K} - \hat{\mathbf{K}}|| \quad .$$

So, if we obtain the single $\mathbf{A}$ that gives an overall least squares for the joint estimation of $\hat{\mathbf{S}}$ and $\hat{\mathbf{K}}$, i.e., such that

$$\{||\mathbf{S} - \mathbf{A}\mathbf{R}_s\mathbf{A}'|| + ||\mathbf{K} - \mathbf{A}\mathbf{R}_k\mathbf{A}'||\}$$

is minimized, we will also obtain the $\mathbf{A}$ for which $||\mathbf{X}-\hat{\mathbf{X}}||$ is minimized, i.e., our desired $\mathbf{A}$ which gives the least-squares best estimate of $\mathbf{X}$.

For the purposes of an iterative algorithm, we seek a conditional least-squares estimate of $\mathbf{A}$, that is, the estimate of $\mathbf{A}$ which minimizes the error quantity shown above, given the previous estimates of $\mathbf{R}$ and $\mathbf{A}$. There are several possible iterative methods that could be used. We will first describe the "implicit equation" (Ramsey, 1975) approach and then generalize this to the ALS approach as it is more commonly understood.

Implicit Equation Method. Consider the estimation of a common $\mathbf{A}$ for both $\hat{\mathbf{S}}$ and $\hat{\mathbf{K}}$ as a problem in regression, involving the following partitioned matrix equation:

$$\left[\frac{(\mathbf{A}\mathbf{R}_s\mathbf{A}')}{(\mathbf{A}\mathbf{R}_k\mathbf{A}')}\right] \overset{\text{l.s.}}{\approx} \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

or, equivalently,

$$\left[\frac{\mathbf{A}\mathbf{R}_s}{\mathbf{A}\mathbf{R}_k}\right]\mathbf{A}' \overset{\text{l.s.}}{\approx} \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

where "$\overset{l.s.}{\approx}$" means "provides the best q-dimensional least squares approximation of."

Here, all matrices are assumed known (from the data or from previous iterations) except $\mathbf{A}'$, which will be estimated by least-squares regression methods.

Now if we use our prior iteration values for $\mathbf{A}$ and $\mathbf{R}$ to define

$$\mathbf{G}_s = \mathbf{A}\mathbf{R}_s$$
$$\mathbf{G}_k = \mathbf{A}\mathbf{R}_k$$

so that

$$\left[\frac{\mathbf{G}_s}{\mathbf{G}_k}\right]\mathbf{A}' \overset{\text{l.s.}}{\approx} \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

then the new least-squares estimate of $\mathbf{A}'$, which we call $\mathbf{A}'_{new}$, is obtained by simply premultiplying both sides of this equation by the generalized inverse of $\left[\frac{\mathbf{G}_s}{\mathbf{G}_k}\right]$ i.e.,

$$\mathbf{A}'_{new} = \left[\frac{\mathbf{G}_s}{\mathbf{G}_k}\right]^+ \left[\frac{\mathbf{S}}{\mathbf{K}}\right].$$

The transpose of $\mathbf{A}'_{new}$ gives our new estimate of $\mathbf{A}$, while

$$\mathbf{R}_{new} = \mathbf{A}^+_{new}\mathbf{X}(\mathbf{A}'_{new})^+$$

*I would like to thank Doug Carroll both for sharpening my previously dim awareness of this fact and for pointing out the necessity of this strong orthogonality property here.

gives our new l.s. estimate of $\mathbf{R}$.

In Ramsey's (1975) terminology, this is an "implicit equation" approach because the old value of A enters explicitly into the expression which is solved to get the new value of A on each iteration. However, as Joseph Kruskal points out (personal communication) it might be preferable to have an algorithm which keeps separate parameters for the left-hand and right-hand A matrices. This approach "two-A" works well when used by CANDECOMP or PARAFAC to solve for the structure of symmetric input data. Even though no constraints are placed on $\mathbf{A}_l$ and $\mathbf{A}_r$ they converge to the same form because of the symmetric structure of the data. Carroll and Chang (1970) have tried both methods for fitting the CANDECOMP model, and report that the "two-A" approach has better convergence rates than the "single-A" or strict "implicit equation" approach. Furthermore, the "two-A" approach is a true Alternating Least Squares (ALS) procedure. Therefore, I will now show how to slightly restructure the DEDICOM algorithm described above so that it has ALS form, and yet still takes advantage of the fact that S and K each have the same row space as column space, so as to force both left and right versions of A to have the same column space when the procedure converges.

ALS Method. First, we distinguish the left-hand A, called $\mathbf{A}_l$, from the right-hand A, called $\mathbf{A}_r$. Then, within each iteration, we do two A estimations, one based on our previously presented partitioned equation

$$\left[\frac{\mathbf{A}_l\mathbf{R}_s}{\mathbf{A}_l\mathbf{R}_k}\right]\mathbf{A}'_r \overset{\text{l.s.}}{\approx} \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

and the other based on the equivalent equation for $\mathbf{A}_l$, namely

$$\mathbf{A}_l\left[\mathbf{R}_s\mathbf{A}'_r \mid \mathbf{R}_k\mathbf{A}'_r\right] \overset{\text{l.s.}}{\approx} [\mathbf{S} \mid \mathbf{K}]$$

To simplify these equations, we define

$$\mathbf{G}_{s_l} = \mathbf{A}_l\mathbf{R}_s \quad \text{(using } \mathbf{A}_l, \mathbf{R}_s \text{ from previous iteration)}$$

$$\mathbf{G}_{k_l} = \mathbf{A}_l\mathbf{R}_k \quad \text{(using } \mathbf{A}_l, \mathbf{R}_k \text{ from previous iteration)}$$

$$\mathbf{G}_{s_r} = \mathbf{R}_s\mathbf{A}'_r \quad \text{(using } \mathbf{A}'_r, \mathbf{R}_s \text{ from previous iteration)}$$

$$\mathbf{G}_{k_r} = \mathbf{R}_k\mathbf{A}'_r \quad \text{(using } \mathbf{A}'_r, \mathbf{R}_k \text{ from previous iteration)}$$

and compute our desired estimates using the formulae

$${}_{\text{new}}\mathbf{A}'_r = \left[\frac{\mathbf{G}_{s_l}}{\mathbf{G}_{k_l}}\right]^+ \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

$${}_{\text{new}}\mathbf{A}_l = [\mathbf{S} \mid \mathbf{K}]\left[\mathbf{G}_{s_r} \mid \mathbf{G}_{k_r}\right]^+$$

with our new estimate of $\mathbf{R}$ being

$${}_{\text{new}}\mathbf{R} = {}_{\text{new}}\mathbf{A}_l^+ \, \mathbf{X} \, ({}_{\text{new}}\mathbf{A}'_r)^+.$$

We note that this generalized method reduces to the standard ALS method of PARAFAC and CANDECOMP when the skew-symmetric part is null (i.e., with symmetric data input).

One Kind of Weighted Least Squares. The method presented above (in either its ALS or implicit equation form) easily generalizes to a family of weighted least squares solutions where the relative importance of the symmetric and the skew-symmetric components of the data can be adjusted as the user desires. One simply defines non-negative weighting factors for these two components, let's call these weights w_s and w_k, and then multiplies each portion of the partitioned data matrix $\left[\frac{\mathbf{S}}{\mathbf{K}}\right]$ by its corresponding weight. Thus, the weighted estimate of $\mathbf{A}_l$ for example would be

$${}_{\text{new}}\mathbf{A}_1 = [w_s\mathbf{S} \mid w_k\mathbf{K}]\,[\mathbf{G}_{s_r} \mid \mathbf{G}_{s_k}]^+.$$

In selecting w_s and w_k, it might often be useful to rescale them so that the total data variance is not changed, i.e., divide both by a constant c such that $||\frac{w_s}{c}\mathbf{S}|| + ||\frac{w_k}{c}\mathbf{K}|| = ||\mathbf{X}||$.

Without the weights (or when $w_s = w_k$), the method produces a solution with the "natural" weighting determined by the relative proportion of the total data variance contributed by the symmetric and skew-symmetric components. This is the solution which provides the (unweighted) least squares best fit to **X** as a whole. But other weightings might be useful for some purposes; for example, it might sometimes be desirable to require that the symmetric and the skew-symmetric components have *equally* strong "influence" on the form of **A**. In such a case one would pick weights which caused the trace of $w_k^2\mathbf{K}'\mathbf{K}$ to equal the trace of $w_s^2\mathbf{S}'\mathbf{S}$, i.e., equate the sums of squares of the elements of the two matrices. Still other weightings might also be useful for particular purposes.

Missing Values. To ignore the diagonal, or any other data cells which one may wish to consider "missing," one can simply add to each iteration an additional step in which the contents of these data cells are predicted by the DEDICOM model, and then replaced by their least-squares estimates, based on the current values of the model parameters. Experience with PARAFAC, which uses this approach to missing values, suggests that this method should be rapidly convergent provided that there are not "too many" missing values (e.g., more than 20%). If this previously observed robustness and rapid convergence of ALS in the presence of missing values also holds for the proposed DEDICOM ALS procedure, it will be an important advantage of this method over current non-ALS methods of fitting the DEDICOM model.

Although the method just described *does* retain ALS properties, since the estimation of missing values is just another conditional least squares step,* it is perhaps more elegant to eliminate the extra step and instead modify the main steps in such a way that they simply do not depend on the missing entries. (Indeed, Doug Carroll suggests that weighted least squares could be used in every part of the conditional least squares estimation.) However, these methods would be more complicated algorithmically than the method of replacement mentioned here.

Three-Way Generalizations. Generalizations to ALS methods for three-way models should be obvious, based on the strong analogy with PARAFAC, etc. However, three-way DEDICOM using ALS might have convergence or uniqueness difficulties similar to some previously attempted ALS versions of PARAFAC2. Empirical exploration will be necessary here.

Row or Column Bias Parameters. It is easy to generalize the ALS procedure so that each iteration includes a phase which provides least squares estimates of some of the additional terms of the "extended DEDICOM model." (The "extended DEDICOM model" (Harshman 1980a) provides for additive and multiplicative row and/or column biases.) As an example, we will here develop the machinery for estimating multiplicative row or column biases. Such bias terms might be needed, for example, to represent response biases when analyzing stimulus confusion matrices.

There are two alternative ways of extending DEDICOM to include multiplicative row (or column) biases. The first way provides a least-squares fit to the original data and involves a model of the form

$$\mathbf{X} = \mathbf{BARA}' + \mathbf{E}$$

where **B** is an n by n diagonal matrix of row bias parameters. The second approach provides a least-squares fit to the data *after* the data has been readjusted to remove the estimated row biases. This approach is represented by a model of the form

*I would like to thank Yoshio Takane for pointing out this fact.

$$\overset{*}{\mathbf{B}}\mathbf{X} = \mathbf{ARA}' + \mathbf{E} \quad .$$

The first approach is the most natural if one takes the relative importance of different rows in the matrix of observations to be given by their relative sums of squares in the original data. The second approach becomes the most natural one, however, if the biases are considered to impose distortions on the "true" structure of X, and if one would like to undo these distortions before determining the "true" relative importance of various effects underlying the data. From this perspective, the second approach provides a natural type of weighted least squares solution.

Since the second approach to row biases is both easier to develop mathematically and also seems more appropriate for many situations, it will be explained first.

To estimate the row bias parameters for the model

$$\overset{*}{\mathbf{B}}\mathbf{X} = \mathbf{ARA}' + \mathbf{E}$$

one adds an extra phase at the end of each ALS iteration. First, $\hat{\mathbf{X}}$ is computed using

$$\hat{\mathbf{X}} = \mathbf{ARA}'$$

Next, the model is redefined in terms of $\hat{\mathbf{X}}$, as

$$\overset{*}{\mathbf{B}}\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} \quad (\overset{*}{\mathbf{B}} \text{ diagonal}) \quad .$$

To solve for $\overset{*}{\mathbf{B}}$, we decompose this matrix equation into a set of n parallel bivariate regression problems (involving regression "through the origin" unless we are also estimating additive bias parameters). If $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ are row vectors corresponding to the ith row of $\mathbf{X}$ and $\hat{\mathbf{X}}$ respectively, then the ith bivariate regression problem is simply

$$\overset{*}{b}_i \mathbf{x}_i = \hat{\mathbf{x}}_i + \mathbf{e}_i \quad .$$

The least-squares solution to this is simply the regression weight

$$\overset{*}{b}_i = \frac{\sum_j x_{ij}\hat{x}_{ij}}{\sum_j x_{ij}^2}$$

On the first ALS iteration, $\overset{*}{\mathbf{B}}$ is taken to be the identity matrix for the purposes of estimating A and R (then it is itself estimated at the end of the iteration). Generally, however, the $\overset{*}{\mathbf{B}}$ estimated at the end of the previous iteration must be used to adjust the rows of the S and K matrices before they are themselves used to estimate new values for A and R. For the ith iteration we define

$$^{i}\mathbf{S} = \overset{*}{\mathbf{B}}\mathbf{S} \quad (\text{using } \overset{*}{\mathbf{B}} \text{ from the previous iteration})$$

$$^{i}\mathbf{K} = \overset{*}{\mathbf{B}}\mathbf{K} \quad (\text{using } \overset{*}{\mathbf{B}} \text{ from the previous iteration})$$

and these matrices are used in place of S and K in the formulae for A, i.e.,

$$\mathbf{A}_l = [{}^{i}\mathbf{S} \mid {}^{i}\mathbf{K}] \left[\mathbf{G}_{s_r} \mid \mathbf{G}_{k_r}\right]^{+} \quad .$$

$$\mathbf{A}_r' = \left[\frac{\mathbf{G}_{s_l}}{\mathbf{G}_{k_l}}\right]^{+} \left[\frac{{}^{i}\mathbf{S}}{{}^{i}\mathbf{K}}\right]$$

and R is estimated as follows:

$$\mathbf{R} = \mathbf{A}_l^{+}\overset{*}{\mathbf{B}}\mathbf{X}(\mathbf{A}_r')^{+} \quad .$$

Yoshio Takane, who has investigated the question of row bias parameters for the DEDICOM model in connection with (non-ALS) derivative-based estimation procedures (discussed elsewhere), has noted that even when the $\overset{*}{b}_i$ are constrained so that $\sum_i \overset{*}{b}_i^2 = 1$, the row biases in his method tended toward a degenerate solution where all the bias terms were zero except one (Takane, personal communication, 1980). It is not clear that a such a problem would emerge with this ALS method; however, to guard against such a problem, a suitable normalization scheme could be adopted at the end of each iteration, which (without loss of generality) normalized the bias parameters so that their *product* is equal to 1.

For standardization of the form of the solution, **A** could also be normalized (or even, in the two-way case, orthonormalized) at the end of each iteration, with the scale of the data being reflected in the size of the entries in the **R** matrix. When convergence was attained, **A** could then be "rotated" to a preferred form for interpretation, with the inverse transformation applied to either side of **R**.

To estimate the row bias parameters using the alternative approach, where one wants a least-squares fit to the "unadjusted" data, one can proceed as follows. As in the "adjusted X" approach we compute $\hat{\mathbf{X}} = \mathbf{ARA}'$ at the end of each iteration, so that we have our model redefined as

$$\mathbf{X} = \mathbf{B}\hat{\mathbf{X}} + \mathbf{E} \quad (\mathbf{B} \text{ diagonal})$$

To solve this, we decompose it into a set of parallel bivariate regression problems, as before, only now the equations have the form

$$\mathbf{x}_i = b_i \hat{\mathbf{x}}_i + \mathbf{e}_i$$

the solution to each of which is simply the regression weight

$$b_i = \frac{\sum_i \hat{x}_{ij} x_{ij}}{\sum_j \hat{x}_{ij}^2}$$

These weights are then incorporated into the ALS algorithm by revising the estimation of $\mathbf{A}_r'$ as follows

$$\mathbf{A}_r' = \left[\frac{\mathbf{BG}_{s_l}}{\mathbf{BG}_{k_l}}\right]^+ \left[\frac{\mathbf{S}}{\mathbf{K}}\right]$$

and revising the estimation of $\mathbf{A}_l$ as follows

$$\mathbf{A}_l = \mathbf{B}^{-1}\,[\mathbf{S} \mid \mathbf{K}]\,[\mathbf{G}_{s_r} \mid \mathbf{G}_{k_r}]^+ \;.$$

(Note that since the estimation of a given row of $\mathbf{A}_l$ includes only products involving the corresponding row of $[\mathbf{S} \mid \mathbf{K}]$, the reweighting of that row by the constant term b_i does not alter the relative pattern of sizes of elements within the row of $\mathbf{A}_l$, and thus, does not change the least-square properties of this estimation.) **R** is then estimated* by

$$\mathbf{R} = (\mathbf{BA}_l)^+ \mathbf{X} (\mathbf{A}_r')^+$$

Joint row and column bias estimates. It is obvious that the procedure for estimating row biases will instead estimate column biases, if one simply applies it to the transpose of **X**. What may be less obvious, however, is that the "unweighted least squares" approach to estimating bias terms (where the bias terms are on the right hand side, with the rest of the model) provides a single solution which can be transformed without loss of fit into either a row bias, column bias, or joint row and column bias form. All these alternative forms still retain the

*This is still a least-squares estimation; see, for example, Rao & Mitra (1971), pp. 60-61.

least squares property of the original solution. As Hollman (1979) has shown, row bias parameters imposed upon an otherwise symmetric metric representation can be transformed into column bias parameters by redefining the symmetric part of the model. We can apply this idea to our DEDICOM asymmetric model as follows, simply let

$$\mathbf{\overset{*}{A}} = \mathbf{\overset{*}{B}A}$$

then

$$\mathbf{X} = \mathbf{\overset{*}{A}R\overset{*}{A}'B^{-1}} + \mathbf{E}$$

which gives a column bias model, with $\mathbf{B}^{-1}$ as the diagonal matrix of column bias terms. We can generalize this approach even further, to obtain a n-parameter family of solutions. If **D** is an arbitrary non-singular n by n diagonal matrix (with the n diagonal cells constituting our n parameters), we can define

$$\mathbf{\overset{*}{A}} = \mathbf{D^{-1}A}$$

and obtain

$$\mathbf{X} = \mathbf{BD\overset{*}{A}R\overset{*}{A}'D} + \mathbf{E} \quad ,$$

which lets us take the diagonal elements of (**BD**) as our row bias terms and the diagonals of **D** as our column bias terms. (Related indeterminacies are also discussed by Constantine and Gower (1980)).

Now, as long as we stay with two-way versions of the DEDICOM model, any member of this family of solutions will fit the data as well as any other. In fact, they all give rise to the same residual matrix, **E**. One obviously needs some auxiliary principle to provide a basis for choice among the alternative versions; presumably the choice would be based on the characteristics of the particular data, and one's evaluation of which model describes the data in the most natural or appropriate way. With three-way DEDICOM (Harshman, 1980a), this indeterminacy should disappear.

In contrast to the "unweighted" approach just discussed, a solution obtained by using the second or "weighted least squares" method of estimating row biases (described earlier) cannot in general be transformed into a family of equivalent solutions with different bias terms. Unless the fit to the data is perfect, nonequivalent solutions (i.e. solutions with different residual matrices and fit values) will generally be obtained when shifting from row-bias to column-bias or to intermediate models. This is because different forms of the model give rise to different weightings of the elements of **X**, and thus different least-squares solutions.

Of course, by simply applying the "weighted" estimation procedure to the transpose of **X**, one obtains column bias parameters for **X**. However, joint estimation of row and column bias parameters would require a minor generalization of the "weighted least squares" row bias method described earlier. Since the form of such a generalization is fairly obvious, it will not be detailed here.

When using the "weighted" approach to estimation of bias parameters, it might be interesting to consider the significance of differences in fit between weightings using row bias terms, vs. those using column bias terms, vs. those using intermediate row-and-column bias estimates. Perhaps if there were substantial differences in fit, such differences might provide a principled basis for preferring a given weighting scheme, at least in the absence of other guidelines.

Final Note. Other issues, such as constrained solutions, three-way solutions, overrelaxation for acceleration of convergence, etc. remain to be discussed. Presumably, many of these could be implemented as in PARAFAC. Also, this ALS approach may facilitate generalization of DEDICOM estimation procedures to take advantage of more general estimation methods such as those used in ALSCAL (Takane, Young and deLeeuw, 1977) to allow relaxation of metrical assumptions about the data. In any case, the method described here, or variants of it, will at least serve to provide a foundation linking the DEDICOM Single Domain model to the ever widening literature on ALS estimation. Hopefully, such a link would allow present (and future) developments in the ALS technique to be applied more easily to the analysis of asymmetric data.

Acknowledgements

The author would like to thank J. D. Carroll, Joseph Kruskal and Yoshio Takane for stimulating discussions and/or valuable and insightful comments on the points discussed in this article.

REFERENCES

Carroll, J. D. and Chang, J-J. Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of Eckert-Young Decomposition. *Psychometrika,* 35, 283-319, 1970.

Constantine, A. G. and Gower, J. Models for the Analysis of Inter-Regional Migration. Unpublished manuscript, Rothamsted Experimental Station, England, 1980.

Harshman, R. A. Models for the Analysis of Asymmetrical Relationships Among n Objects or Stimuli. Talk presented to the U.S.-Japan Seminar on Multidimensional Scaling, Santa Barbara, 1975.

Harshman, R. A. Models for the Analysis of Asymmetrical Relationships Among n Objects or Stimuli. Paper presented to the First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, Hamilton, Ontario, 1978.

Harshman, R. A. A Generalization of Factor Analysis and Multidimensional Scaling for the Analysis of Asymmetric Relationships. Part I: Theory. Manuscript in preparation, 1981a.

Harshman, R. A. DEDICOM Analysis of Skew-Symmetric Data. Part I: Theory. Bell Labs Technical Memorandum 80-11229-6, 1981b.

Hollman, E. Monotonic Models for Asymmetric Proximities. *Journal of Mathematical Psychology,* 20: 1-15, 1979.

Ramsey, J. O. Solving Implicit Equations in Psychometric Data Analysis. *Psychometrika,* 40, 337-360, 1975.

Rao, C. R. and Mitra, S. K. *Generalized Inverse of Matrices and its Applications,* New York: Wiley & Sons, 1971.

Takane, Y., Young, F. W., deLeeuw, J. Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika,* 42: 7-67, 1977.

Wold, H. Estimation of Principal Components and Related Models by Iterative Least Squares. In *Multivariate Analysis*, ed. P. R. Krishnaiah. New York: Academic Press, 1966.