

COMPONENT ANALYSIS WITH DIFFERENT SETS OF CONSTRAINTS ON DIFFERENT DIMENSIONS

YOSHIO TAKANE

MCGILL UNIVERSITY

HENK A. L. KIERS

UNIVERSITY OF GRONINGEN

JAN DE LEEUW

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Many of the "classical" multivariate data analysis and multidimensional scaling techniques call for approximations by lower dimensional configurations. A model is proposed, in which different sets of linear constraints are imposed on different dimensions in component analysis and "classical" multidimensional scaling frameworks. A simple, efficient, and monotonically convergent algorithm is presented for fitting the model to the data by least squares. The basic algorithm is extended to cover across-dimension constraints imposed in addition to the dimensionwise constraints, and to the case of a symmetric data matrix. Examples are given to demonstrate the use of the method.

Key words: dimensionwise constraints, across-dimension constraints, rank restrictions, constrained principal component analysis (CPCA), multidimensional scaling (MDS).

1. Introduction

One of the most important roles of multivariate data analysis is finding a representation in a reduced number of dimensions. Eigenvalue or singular value decomposition plays an important role in this process. Often, however, additional information about the data may be incorporated in the data representation. For example, Panel A of Figure 1 displays a configuration of nine colors according to the Munsell System (Torgerson, 1958). The stimuli are arranged factorially in terms of brightness (value) and saturation (chroma). In multidimensional scaling (MDS) of dissimilarities among these colors, the factorial structure may be explicitly incorporated in the stimulus representation. It is of interest to see how well the specific Munsell dimensions reflect observed dissimilarities among the colors.

One traditional approach to this problem is to obtain an unconstrained stimulus configuration, which is then transformed by a procrustes rotation to match the hypothesized structure. A goodness of agreement between the two provides an indication of how good the hypothesized structure is in accounting for the observed data. One potential problem with this approach is that it involves two distinct criteria, one for obtaining the unconstrained solution and the other for the rotation. The two criteria are

The work reported in this paper has been supported by the Natural Sciences and Engineering Research Council of Canada, grant number A6394, and by the McGill-IBM Cooperative Grant, both granted to the first author. The research of H. A. L. Kiers has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences. We thank Michael Hunter for his helpful comments on earlier drafts of this paper.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1, CANADA.

often incompatible, which makes it difficult to evaluate the goodness of the overall fit of the hypothesized structure. Another potential problem is that the initial unconstrained solution may leave out pertinent dimensions. For example, we may initially obtain an r -dimensional solution, while some of the dimensions in the hypothesized structure may not be among the first r dimensions. We may then erroneously conclude that the relevant dimensions do not exist (ten Berge, 1986). Furthermore, constructing a target matrix for the procrustes rotation may not be so straightforward. Factorial structures of stimuli, for example, require quantifications of factorial designs before they can be used as targets.

A number of attempts have been made to directly incorporate external information in dimension reduction. For example, Rao (1964) proposed the restricted eigenvalue method. He also proposed principal components of instrumental variables, also known as reduced rank regression (Anderson, 1951) or redundancy analysis (van den Wollenberg, 1977). Carroll, Pruzansky, and Kruskal (1980) proposed CANDELINC that allows impositions of constraints on both rows and columns of a data matrix (see also Golub, 1973). Ter Braak (1986) and Böckenholt and Böckenholt (1990) made analogous developments for discrete data. Takane and Shibayama (1991) proposed CPCA (Constrained Principal Component Analysis), which provided a comprehensive framework to impose linear constraints. Technically, all these methods amount to projection of a data matrix, followed by the (generalized) eigenvalue or singular value decomposition. Thus, no new numerical techniques are required.

The above methods, however, can typically accommodate only one set of constraints on each side of a data matrix. In obtaining multidimensional solutions, they form multiple sets of linear combinations of a same set of constraints. This is not very attractive, however, when the dimensions of the hypothesized structures corresponds with specific attributes of stimuli. In the Munsell color example discussed earlier, the horizontal axis represents chroma, and the vertical axis value. A separate set of constraints are necessary to characterize each dimension.

Let \mathbf{H}_1 denote a matrix of contrast vectors among levels of the chroma dimension, and similarly, \mathbf{H}_2 for the value dimension. Let \mathbf{v}_1 and \mathbf{v}_2 be coordinate vectors corresponding to the two dimensions. Then, \mathbf{v}_1 is a linear combination of \mathbf{H}_1 alone. That is, $\mathbf{v}_1 = \mathbf{H}_1 \mathbf{v}_1^*$, where \mathbf{v}_1^* is a vector of weights. Similarly, $\mathbf{v}_2 = \mathbf{H}_2 \mathbf{v}_2^*$. Let $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$. Then,

$$\mathbf{V} = [\mathbf{H}_1 \mathbf{v}_1^*, \mathbf{H}_2 \mathbf{v}_2^*]. \quad (1)$$

The methods mentioned earlier, on the other hand, join \mathbf{H}_1 and \mathbf{H}_2 into a single matrix, $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$, and obtain \mathbf{V} as linear combinations of the entire matrix. That is,

$$\mathbf{V} = \mathbf{H} \mathbf{V}^*, \quad (2)$$

where \mathbf{V}^* is a matrix of weights. This typically obtains dimensions that are mixtures of the prescribed dimensions.

In this paper we propose a matrix decomposition that directly incorporates constraints of the form in (1) in component analysis and classical MDS contexts, and that avoids the difficulties associated with the procrustes rotation approach discussed earlier. We call this decomposition DCDD (Different Constraints on Different Dimensions). In contrast, we call constraints of the form in (2) "CPCA" type. The DCDD type of constraints have been used in confirmatory factor analysis and analysis of covariance structures (e.g., Jöreskog, 1969, 1970), in "nonclassical" MDS (e.g., Heiser & Meulman, 1983; Meulman & Heiser, 1984; Ramsay, 1980; Takane, 1978, 1981), and in unfolding analysis (DeSarbo & Rao, 1984; Heiser, 1981). Yet, except for a few

special cases, their use in component analysis and classical MDS seems new and merits our consideration (see, however, a recent article by Velu, 1991; also see van der Lans, 1992).

DCDD, unlike CPCA, does not reduce to singular value decomposition. A special-purpose iterative algorithm has to be used to fit the constraints. We propose an algorithm, which is a simple adaptation of an ALS (alternating least squares) algorithm developed earlier by Meulman and Heiser (1984, see also Meulman, 1986) and by van der Burg, de Leeuw and Verdegaal (1988). In the next section, we present our method along with some empirical justifications. We define our problem more precisely (section 2.1), give a simple and efficient algorithm to solve the problem (section 2.2), and discuss a technique to speed up the convergence of the algorithm (section 2.3). We also extend our method to across-dimension constraints (see section 2.4) and to the case of a symmetric data matrix (section 2.5). In section 3, we give additional remarks on the model and the algorithm, and discuss some relationship between DCDD and other methods (e.g., Böckenholt and Böckenholt, 1990). In section 4, we illustrate practical uses of the proposed method (DCDD) through concrete examples, followed by discussions (section 5).

2. The Method

2.1. The Problem

Let \mathbf{Z} denote an n by p data matrix, and let \mathbf{G}_i (n by s_i ; $i = 1, \dots, m$) and \mathbf{H}_i (p by t_i ; $i = 1, \dots, m$) be m sets of row and column constraint matrices, respectively. Without loss of generality we assume \mathbf{G}_i and \mathbf{H}_i are of full column rank. The constraint matrices may consist of continuous variables alone, discrete variables alone, or mixtures of both, which may or may not have been columnwise centered. The discrete variables are coded into either dummy variables or contrast vectors. Examples of \mathbf{H}_i 's are given in Tables 1, 3, and 4. Not all \mathbf{G}_i 's or \mathbf{H}_i 's have to be distinct.

The row constraint matrix \mathbf{G}_i may, for example, represent subjects' demographic information such as gender, age, education levels, etcetera. The column constraint matrix \mathbf{H}_i , on the other hand, may represent any structural relationships among the columns of data matrix \mathbf{Z} , for example, a factorial structure among a set of stimuli. When there are no obvious structures to be incorporated, \mathbf{G}_i or \mathbf{H}_i may be set to an identity matrix. To motivate simultaneous use of \mathbf{G}_i and \mathbf{H}_i , consider an example from Greenacre (1993) who presents a contingency table obtained by cross-tabulating frequencies of car purchasers in terms of class sizes of cars (columns), and age-income combinations of purchasers (rows). He applied unconstrained correspondence analysis to the data, and obtained a two-dimensional solution, the first dimension of which roughly corresponded with purchasers' age and car sizes, and the second dimension with purchasers' income levels and luxuriousness of cars. We may take \mathbf{G}_1 and \mathbf{H}_1 to be matrices of dummy variables indicating purchasers' age levels and car sizes, respectively, and \mathbf{G}_2 and \mathbf{H}_2 indicating income levels and luxuriousness of cars, respectively, and examine how well these constraints account for the data.

Our problem is to minimize

$$f = SS\left(\mathbf{Z} - \sum_{i=1}^m \mathbf{G}_i \mathbf{M}_i \mathbf{H}_i'\right), \quad (3)$$

(where $SS(\mathbf{X}) = \text{tr}(\mathbf{X}'\mathbf{X})$) with respect to \mathbf{M}_i (s_i by t_i) subject to the restriction that

$$\text{rank} (\mathbf{M}_i) = r_i, \tag{4}$$

where r_i is a prescribed integer. Often, $r_i = 1$ for all i . Then, \mathbf{M}_i can be reparametrized as

$$\mathbf{M}_i = d_i \mathbf{u}_i^* \mathbf{v}_i^{*'} \tag{5}$$

where d_i is a positive scalar, and \mathbf{u}_i^* and \mathbf{v}_i^* s_i -component and t_i -component vectors of weights, respectively. Minimizing (3) over \mathbf{M}_i amounts to minimizing (3) over d_i , \mathbf{u}_i^* and \mathbf{v}_i^* for $i = 1, \dots, m$. Let $\mathbf{u}_i = \mathbf{G}_i \mathbf{u}_i^*$ and $\mathbf{v}_i = \mathbf{H}_i \mathbf{v}_i^*$. We assume $\|\mathbf{u}_i\| = 1$ and $\|\mathbf{v}_i\| = 1$ for all i , where $\|\mathbf{x}\| = (\mathbf{x}'\mathbf{x})^{1/2}$. These restrictions are necessary to identify the parameters.

In some cases, the value of r_i between 1 and $\min(s_i, t_i)$ may be chosen. For example, \mathbf{H}_i may represent qualitatively distinct experimental treatments under which measurements in the data matrix were taken. We would like to capture the treatment effects, but one dimension may not be sufficient to capture all important aspects of the differences among the treatments. In this case, \mathbf{u}_i , \mathbf{v}_i , and d_i are replaced by matrices $\mathbf{U}_i(s_i \text{ by } r_i)$, $\mathbf{V}_i(t_i \text{ by } r_i)$, and $\mathbf{D}_i(r_i \text{ by } r_i)$, respectively. To identify the parameters in this case, it is convenient to require $\mathbf{U}_i' \mathbf{U}_i = \mathbf{I}$, $\mathbf{V}_i' \mathbf{V}_i = \mathbf{I}$, and \mathbf{D}_i diagonal and pd . When no rank restriction is imposed on \mathbf{M}_i (i.e., $r_i = \min(s_i, t_i)$), no special representation of \mathbf{M}_i (or $\mathbf{G}_i \mathbf{M}_i \mathbf{H}_i'$) is necessary.

Criterion (3) is defined in the usual identity metrics. However, it is straightforward to incorporate metric matrices other than identity matrices. Nonidentity metric matrices considerably widen the scope of data analysis (Meredith & Millsap, 1985). For example, it becomes possible to perform correspondence analysis or dual scaling (Greenacre, 1984; Nishisato, 1980) with the DCDD type of constraints.

Let \mathbf{K} and \mathbf{L} denote the metric matrices. Both \mathbf{K} and \mathbf{L} are assumed nnd , and to satisfy the following rank conditions; $\text{rank} (\mathbf{K} \mathbf{G}_i) = \text{rank} (\mathbf{G}_i)$ and $\text{rank} (\mathbf{L} \mathbf{H}_i) = \text{rank} (\mathbf{H}_i)$ for all i . We generalize (3) into

$$f = SS \left(\frac{\mathbf{Z} - \sum_{i=1}^m \mathbf{G}_i \mathbf{M}_i \mathbf{H}_i'}{\mathbf{K}, \mathbf{L}} \right), \tag{3'}$$

where $SS(\mathbf{X})_{\mathbf{K}, \mathbf{L}} = \text{tr}(\mathbf{K} \mathbf{X} \mathbf{L} \mathbf{X}')$. Velu (1991) recently considered a special case of (3'), where $m = 2$, $\mathbf{H}_1 = \mathbf{H}_2 = \mathbf{I}$, $\mathbf{K} = \mathbf{I}$ and \mathbf{L} taken to be the inverse of the covariance matrix among the columns of data matrix \mathbf{Z} .

Criterion (3') can be reduced to criterion (3) by simple transformations (e.g., Rao, 1980). Let $\mathbf{K} = \mathbf{R}_K \mathbf{R}_K'$ and $\mathbf{L} = \mathbf{R}_L \mathbf{R}_L'$ be any square root decompositions of \mathbf{K} and \mathbf{L} . Then, (3') can be rewritten as

$$\begin{aligned} f &= SS \left(\mathbf{R}_K' \mathbf{Z} \mathbf{R}_L - \sum_{i=1}^m \mathbf{R}_K' \mathbf{G}_i \mathbf{M}_i \mathbf{H}_i' \mathbf{R}_L \right) \\ &= SS \left(\frac{\mathbf{Z}^* - \sum_{i=1}^m \mathbf{G}_i^* \mathbf{M}_i \mathbf{H}_i^*}{\mathbf{I}} \right), \end{aligned} \tag{3''}$$

where $\mathbf{Z}^* = \mathbf{R}_K' \mathbf{Z} \mathbf{R}_L$, $\mathbf{G}_i^* = \mathbf{R}_K' \mathbf{G}_i$ and $\mathbf{H}_i^* = \mathbf{R}_L' \mathbf{H}_i$. This is essentially the same as (3), and can be minimized in the same way. Once $\mathbf{G}_i^* \mathbf{M}_i \mathbf{H}_i^*$ in (3'') is obtained, $\mathbf{G}_i \mathbf{M}_i \mathbf{H}_i'$ in (3') can be derived by $\mathbf{G}_i \mathbf{M}_i \mathbf{H}_i' = (\mathbf{R}_K^+)' \mathbf{G}_i^* \mathbf{M}_i \mathbf{H}_i^* (\mathbf{R}_L^+)$, where \mathbf{X}^+ indicates the Moore-Penrose inverse of \mathbf{X} .

An iterative algorithm to minimize (3) will be presented in the next section. However, there are three special cases where the minimization of (3) does not require an iterative algorithm. First of all, when $\mathbf{G}_i = \mathbf{G}$ and $\mathbf{H}_i = \mathbf{H}$ for all i , (3) reduces to

$$f = SS(\mathbf{Z} - \mathbf{GMH}'), \quad (6)$$

where $\mathbf{M} = \sum_{i=1}^m \mathbf{M}_i$. This is the CPCA problem discussed by Takane & Shibayama (1991). The \mathbf{GMH}' that minimizes (6) can be obtained by the singular value decomposition (SVD) of $\mathbf{P}_G \mathbf{Z} \mathbf{P}_H$, where $\mathbf{P}_G = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ and $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$.

Secondly, when $\mathbf{G}'_i \mathbf{G}_j = \mathbf{0}$ for $i \neq j$ and/or $\mathbf{H}'_i \mathbf{H}_j = \mathbf{0}$ for $i \neq j$, the minimization of (3) is equivalent to that of

$$\sum_{i=1}^m SS(\mathbf{Z} - \mathbf{G}_i \mathbf{M}_i \mathbf{H}'_i), \quad (7)$$

which can be minimized separately for each i . Each minimization problem is again a CPCA problem.

Finally, when no rank restrictions are imposed on any \mathbf{M}_i 's, there is a simple analytic solution. Let $\text{vec}(\mathbf{X})$ denote a supervector formed by column vectors of matrix \mathbf{X} . Define \mathbf{A} as

$$\mathbf{A} = [\mathbf{H}_1 \otimes \mathbf{G}_1 \mid \dots \mid \mathbf{H}_m \otimes \mathbf{G}_m], \quad (8)$$

where \otimes indicates a direct (Kronecker) product. Then, f in (3) can be rewritten as

$$f = SS(\text{vec}(\mathbf{Z}) - \mathbf{A}\mathbf{m}), \quad (9)$$

where \mathbf{m} is a supervector of $\mathbf{m}_i = \text{vec}(\mathbf{M}_i)$. A least squares estimate of \mathbf{m} is obtained by

$$\mathbf{m} = (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}' \text{vec}(\mathbf{Z}), \quad (10)$$

where \mathbf{X}^{-} denotes a g -inverse of \mathbf{X} .

2.2. Algorithm

In this section, we present an algorithm to minimize (3). It is an ALS algorithm (e.g., de Leeuw, Young, & Takane, 1976), and hence is monotonically convergent. As noted earlier, essentially the same algorithm has been used elsewhere by Meulman and Heiser (1984; Meulman, 1986) and by van der Burg et al. (1988). A similar algorithm was also used by Takane, Young, and de Leeuw (1980) for the weighted additive model, which is also a special case of the DCDD model (van der Lans, 1992).

Assume, for the moment, $r_i = 1$ for all i , and define

$$\mathbf{Z}_{(-k)} = \mathbf{Z} - \sum_{i \neq k} d_i \mathbf{u}_i \mathbf{v}'_i. \quad (11)$$

Then, for a specific k , (3) may be rewritten as

$$f_k(d_k, \mathbf{u}_k, \mathbf{v}_k) = SS(\mathbf{Z}_{(-k)} - d_k \mathbf{u}_k \mathbf{v}'_k). \quad (12)$$

One way to minimize (3) is to minimize f_k sequentially for $k = 1, \dots, m$, with respect to d_k , \mathbf{u}_k , and \mathbf{v}_k , assuming temporarily d_i , \mathbf{u}_i , and \mathbf{v}_i ($i \neq k$) are known. This sequential process is repeated until convergence is reached. The f_k , on the other hand, may be minimized by minimizing it with respect to \mathbf{u}_k (for fixed d_k and \mathbf{v}_k), \mathbf{v}_k (for fixed

d_k and \mathbf{u}_k) and d_k (for fixed \mathbf{u}_k and \mathbf{v}_k). A complete minimization of f_k is not necessary, however, and it is sufficient and usually more efficient to update \mathbf{u}_k , \mathbf{v}_k , and d_k (according to Steps 3.3 through 3.5 below) only once before moving to the next k . This decreases the value of f_k consistently. Since f is bounded below and the parameter updating equations are continuous, the iterative sequence is bound to converge to a stationary point.

Algorithm

- Step 1. Set $q = 0$ (iteration number) and initialize d_i , \mathbf{u}_i , and \mathbf{v}_i ($i = 2, \dots, m$).
- Step 2. Increment q by 1.
- Step 3. For $k = 1, \dots, m$, sequentially, repeat Step 3.1 to Step 3.5.
- Step 3.1. Calculate $\mathbf{Z}_{(-k)}$ according to (11).
- Step 3.2. (Only when $q = 1$). Initialize \mathbf{v}_k .
- Step 3.3. Update \mathbf{u}_k by $\mathbf{P}_{G(k)}\mathbf{Z}_{(-k)}\mathbf{v}_k$ and normalize, where $\mathbf{P}_{G(k)} = \mathbf{G}_k(\mathbf{G}'_k\mathbf{G}_k)^{-1}\mathbf{G}'_k$.
- Step 3.4. Update \mathbf{v}_k by $\mathbf{P}_{H(k)}\mathbf{Z}'_{(-k)}\mathbf{u}_k$ and normalize, where $\mathbf{P}_{H(k)} = \mathbf{H}_k(\mathbf{H}'_k\mathbf{H}_k)^{-1}\mathbf{H}'_k$.
- Step 3.5. Update d_k by $d_k = \mathbf{u}'_k\mathbf{Z}_{(-k)}\mathbf{v}_k$.
- Step 4. Check convergence. If not converged, go back to Step 2. The iteration is terminated when the amount of decrease in f falls below a certain threshold value; for example $(f^{(q)} - f^{(q+1)})/f^{(q)} \leq 10^{-10}$.

A few remarks are in order. First, it can be proven that each of Steps 3.3, 3.4 and 3.5 obtains the global minimum of f_k given the other parameters fixed. Secondly, it can be shown that $\mathbf{u}'_k\mathbf{Z}_{(-k)}\mathbf{v}_k = \|\mathbf{P}_{H(k)}\mathbf{Z}'_{(-k)}\mathbf{u}_k\|$, so that d_k can alternatively be obtained as the normalization factor in Step 3.4. Thirdly, Steps 3.3 and 3.4 may be more efficiently carried out by the following procedure. Let $\mathbf{G}_k = \mathbf{F}_k\mathbf{R}'_k$ be the QR decomposition of \mathbf{G}_k . Then, $\mathbf{P}_{G(k)} = \mathbf{F}_k\mathbf{F}'_k$. We calculate $\mathbf{F}'_k\mathbf{Z}_{(-k)}\mathbf{v}_k$ first, normalize it, and then pre-multiply it by \mathbf{F}_k . Often $\mathbf{F}'_k\mathbf{Z}_{(-k)}$ is much smaller in size than $\mathbf{P}_{G(k)}\mathbf{Z}_{(-k)}$. Essentially the same holds for Step 3.4. Note that the QR decompositions of \mathbf{G}_k and \mathbf{H}_k need to be obtained only once before the iteration starts.

When $1 < r_k < \min(s_k, t_k)$ for some k , Steps 3.3–3.5 should be modified accordingly. The $\mathbf{P}_{G(k)}\mathbf{Z}_{(-k)}\mathbf{u}_k$ and $\mathbf{P}_{H(k)}\mathbf{Z}'_{(-k)}\mathbf{v}_k$ are orthonormalized by the Gram-Schmidt orthonormalization procedure in Steps 3.3 and 3.4, and $\mathbf{D}_k = \text{diag}(\mathbf{U}'_k\mathbf{Z}_{(-k)}\mathbf{V}_k)$ in Step 3.5. When no rank restriction is imposed on \mathbf{M}_k for some k , $\mathbf{G}_k\mathbf{M}_k\mathbf{H}'_k$ is simply updated by $\mathbf{P}_{G(k)}\mathbf{Z}_{(-k)}\mathbf{P}_{H(k)}$. It can be shown that the iterative algorithm with this update leads to the same solution as the analytic solution mentioned earlier, when no rank restrictions are imposed on any \mathbf{M}_k 's.

2.3. Further Remarks

When $n \gg p$ and there are no row constraints to be incorporated (i.e., $\mathbf{G}_i = \mathbf{I}$ for all i), a significant cut in computation time is obtained by the following procedure. This procedure also has an important implication for avoiding the estimation of incidental parameters. Let the QR decomposition of \mathbf{Z} be $\mathbf{Z} = \mathbf{F}_Z\mathbf{R}'_Z$, where $\mathbf{F}'_Z\mathbf{F}_Z = \mathbf{I}$ and \mathbf{R}'_Z is upper trapezoidal. Then, (3) can be rewritten as follows:

$$\begin{aligned} f &= SS(\mathbf{F}_Z\mathbf{R}'_Z - \mathbf{Z}_0) \\ &= SS(\mathbf{R}'_Z - \mathbf{F}'_Z\mathbf{Z}_0) + SS(\mathbf{F}_Z\mathbf{F}'_Z\mathbf{Z}_0 - \mathbf{Z}_0) \\ &= SS(\mathbf{R}'_Z - \mathbf{F}'_Z\mathbf{Z}_0), \end{aligned} \tag{13}$$

where

$$\mathbf{Z}_0 = \sum_{i=1}^m \mathbf{M}_i \mathbf{H}'_i = \mathbf{M} \mathbf{H}', \quad (14)$$

with $\mathbf{M} = [\mathbf{M}_1, \dots, \mathbf{M}_m]$ and $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_m]$. The $SS(\mathbf{F}_Z \mathbf{F}'_Z \mathbf{Z}_0 - \mathbf{Z}_0)$ vanishes, because \mathbf{Z}_0 is in the column space of \mathbf{Z} , and therefore $\mathbf{F}_Z \mathbf{F}'_Z \mathbf{Z}_0 = \mathbf{Z}_0$. That \mathbf{Z}_0 is indeed in the column space of \mathbf{Z} can be seen from the fact that when no rank restrictions are imposed on \mathbf{M}_i 's, \mathbf{M} in (14) can be updated by

$$\mathbf{M} = \mathbf{Z} \mathbf{H} (\mathbf{H}' \mathbf{H})^{-1}.$$

Since rank restrictions only restrict \mathbf{M} to lie in a subspace of \mathbf{M} given above, \mathbf{M} should lie in the column space of \mathbf{Z} , irrespective of rank restrictions.

Let $\mathbf{N} = \mathbf{F}'_Z \mathbf{M}$. Then, (13) can be rewritten as

$$f = SS(\mathbf{R}'_Z - \mathbf{N} \mathbf{H}'). \quad (15)$$

We may apply the algorithm presented in the previous section to minimize (15). The number of rows of \mathbf{R}'_Z is at most p , which is usually much smaller than n . Once \mathbf{N} is obtained, \mathbf{M} can be recovered by

$$\mathbf{M} = \mathbf{F}_Z \mathbf{N}. \quad (16)$$

Since \mathbf{R}'_Z can also be obtained from the Cholesky factorization of $\mathbf{Z}' \mathbf{Z} = \mathbf{R}_Z \mathbf{R}'_Z$, the above procedure enables us to fit the DCDD type of constraints to covariance or correlation matrices, provided that no row restrictions are imposed on \mathbf{Z} . When we deal with covariance or correlation matrices, however, there is no way to recover \mathbf{M} from \mathbf{N} .

The above procedure has a far reaching consequence beyond mere computational convenience. Rows of data matrix \mathbf{Z} often correspond with subjects. When there are no row constraints, the number of parameters in \mathbf{M} to be estimated increases, as n increases. Such parameters are called incidental parameters. In the presence of such parameters, even nonincidental parameters in the model may not be consistently estimated (e.g., Andersen, 1980; Kiefer & Wolfowitz, 1956). The number of parameters in \mathbf{N} , on the other hand, is bounded by some function of p . That is, we are in effect bypassing the estimation of incidental parameters in the above procedure. Note that this result holds for any model \mathbf{Z}_0 that is in the column space of data matrix \mathbf{Z} , ordinary PCA being the most prominent example (see also Kiers & Krijnen, 1991; and Kiers, Kroonenberg, & ten Berge, 1992).

2.4. Additional Constraints

Additional linear constraints may be incorporated in the minimization of (3). Of particular interest are across-dimension constraints. For example, in the Munsell color example discussed in the introduction section, we may be interested in testing whether two chroma units are equivalent to one value unit, as hypothesized by the Munsell System. Since this hypothesis involves stimulus coordinates on different dimensions, across-dimension constraints are required.

Across-dimension constraints require comparability of scale across dimensions. Since the restriction that $\|\mathbf{v}_i\| = 1$ for $i = 1, \dots, m$ is completely arbitrary, it is not meaningful to impose across-dimension constraints on \mathbf{v}_i 's. However, $\tilde{\mathbf{v}}_i = d_i \mathbf{v}_i$ ($i = 1, \dots, m$) represents weights applied to \mathbf{u}_i ($i = 1, \dots, m$) to obtain a best approximation to data matrix \mathbf{Z} , so that the across-dimension constraints may be imposed on

\bar{v}_i 's. While the same holds for $\bar{u}_i = d_i u_i$, in the following discussion we assume that the across dimension constraints are imposed only on the column side and that $r_i = 1$ for all i . When $r_i > 1$ for some i , r_i copies of H_i should be made, which are treated as if they are separate constraints (de Leeuw, 1984).

Let $\bar{v}' = (\bar{v}'_1, \dots, \bar{v}'_m)$ and $\bar{v}^* = (\bar{v}^*_1, \dots, \bar{v}^*_m)$. Across-dimension constraints may generally be expressed as

$$C' \bar{v}^* = 0, \tag{17}$$

where C is a known constraint matrix. In some cases the constraints may be stated in terms of \bar{v} rather than \bar{v}^* , that is,

$$C^* \bar{v} = 0, \tag{18}$$

where C^* is analogous to C above. However, (18) can be reduced to (17) by

$$C^* \bar{v} = C^* D_H \bar{v}^* = C' \bar{v}^* = 0, \tag{19}$$

where $C' = C^* D_H$, and D_H is a block diagonal matrix with H_i 's as diagonal blocks. An example of C^* will be given in section 4.1. To incorporate (17), it is necessary to update all parts of \bar{v}^* simultaneously. For fixed u_i ($i = 1, \dots, m$), let

$$A = [H_1 \otimes u_1 \mid \dots \mid H_m \otimes u_m]. \tag{20}$$

Then, the LS criterion (3) can be rewritten as

$$f = SS(\text{vec}(Z) - A \bar{v}^*),$$

which is minimized with respect to \bar{v}^* subject to the constraint (17). Standard procedures are available for this constrained LS problem (see, for example, Takane, Yanai, & Mayekawa, 1991, Appendix C). Once \bar{v}^* is updated, u_i 's are sequentially updated with $\|u_i\| = 1$ for fixed \bar{v}^* as before. The simultaneous updating of \bar{v}^* and sequential updatings of u_i 's are alternated until convergence is reached.

2.5. The Symmetric Case

When a data matrix is symmetric, row and column constraints are usually identical ($G_i = H_i$, $i = 1, \dots, m$), and it is natural to assume $u_k = v_k$ (or $U_k = V_k$). Let S denote the symmetric data matrix. We assume S is (at least nearly) *nnd*. For simplicity, we also assume $r_i = 1$ for all i . An extension to $r_i > 1$ is analogous to the rectangular case.

The minimization criterion, (3), is rewritten as

$$\begin{aligned} f &= SS \left(S - \sum_{i=1}^m d_i^2 v_i v_i' \right) \\ &= SS(S_{(-k)} - d_k^2 v_k v_k'), \end{aligned} \tag{21}$$

where

$$S_{(-k)} = S - \sum_{i \neq k}^m d_i^2 v_i v_i'. \tag{22}$$

A modification necessary in the algorithm is straightforward.

Step 3 has to be replaced by

Step 3.1. Obtain $\mathbf{S}_{(-k)}$ according to (22).

Step 3.2. Update \mathbf{v}_k by $\mathbf{P}_{H(k)}\mathbf{S}_{(-k)}\mathbf{v}_k$ and normalize.

Step 3.3. Update d_k^2 by $\max(\mathbf{v}'_k\mathbf{S}_{(-k)}\mathbf{v}_k, 0)$.

Step 3.3 obtains the least squares update of d_k^2 subject to $d_k^2 \geq 0$ given $\mathbf{S}_{(-k)}$ and \mathbf{v}_k . The $\mathbf{S}_{(-k)}^* = \mathbf{P}_{H(k)}\mathbf{S}_{(-k)}\mathbf{P}_{H(k)}$ is not assured of *nnd*, and consequently $\mathbf{v}'_k\mathbf{S}_{(-k)}^*\mathbf{v}_k = \mathbf{v}'_k\mathbf{S}_{(-k)}\mathbf{v}_k$ can be negative. Step 3.3 simply sets $d_k^2 = 0$ if it is negative.

Step 3.2 is not an ALS step. However, monotonic convergence of the algorithm can be assured, if $\mathbf{S}_{(-k)}^*$ is *nnd*. In the following proof we omit subscript k ; thus, \mathbf{S} , \mathbf{v} , and d^2 denote $\mathbf{S}_{(-k)}^*$, \mathbf{v}_k , and d_k^2 , respectively. The old and new updates of \mathbf{v} are distinguished by subscripts, o and n . The problem is to minimize

$$f = \mathbf{S}\mathbf{S}(\mathbf{S} - d^2\mathbf{v}\mathbf{v}'), \quad (23)$$

subject to $\|\mathbf{v}\| = 1$. Under $d^2 = \mathbf{v}'\mathbf{S}\mathbf{v}$, this is equivalent to maximizing

$$g = \mathbf{v}'\mathbf{S}\mathbf{v}, \quad (24)$$

subject to the same restriction ($\|\mathbf{v}\| = 1$). We must prove $\mathbf{v}'_n\mathbf{S}\mathbf{v}_n \geq \mathbf{v}'_o\mathbf{S}\mathbf{v}_o$, when \mathbf{v} is updated by

$$\mathbf{v}_n = \frac{\mathbf{S}\mathbf{v}_o}{\|\mathbf{S}\mathbf{v}_o\|}. \quad (25)$$

This is straightforward by noting that using Lemma 2 by ten Berge (1986) twice, we have

$$\begin{aligned} \mathbf{v}'_n\mathbf{S}\mathbf{v}_n &= \frac{\mathbf{v}'_o\mathbf{S}^3\mathbf{v}_o}{\mathbf{v}'_o\mathbf{S}^2\mathbf{v}_o} \\ &\geq \frac{\mathbf{v}'_o\mathbf{S}^2\mathbf{v}_o}{\mathbf{v}'_o\mathbf{S}\mathbf{v}_o} \geq \mathbf{v}'_o\mathbf{S}\mathbf{v}_o \end{aligned} \quad (26)$$

Note that \mathbf{S} is assumed *nnd*, which is not assured. Empirical evidence shows, however, that the problem rarely arises in practice, if the original data, \mathbf{S} , is at least nearly *nnd*. Moreover, if necessary, we can always modify the algorithm such that it does converge monotonically. This can be done by replacing Step 3.2 by a step that takes the eigenvector associated with the largest eigenvalue of $\mathbf{P}_{H(k)}\mathbf{S}_{(-k)}\mathbf{P}_{H(k)}$.

Additional across-dimension constraints can be incorporated in a similar way to the rectangular case. In this case we may differentiate the left \mathbf{v}_i and the right \mathbf{v}_i , and treat them as separate parameters. The problem then becomes essentially the same as in the rectangular case. Although there is no theoretical guarantee that left \mathbf{v} and the right \mathbf{v} are proportional to each other at a convergence point (ten Berge & Kiers, 1991), this seems to be the case most of the time.

3. Additional Remarks

3.1. Uniqueness and Local Minima

A necessary and sufficient condition for uniqueness of parameters has been given in Kiers & Takane (1993) for a similar situation. This condition stipulates that \mathbf{G}_i ($i = 1, \dots, m$) are disjoint with all \mathbf{u}_j ($j \neq i$) and/or \mathbf{H}_i ($i = 1, \dots, m$) disjoint with all \mathbf{v}_j ($j \neq i$), assuming that all \mathbf{u}_j 's and \mathbf{v}_j 's are linearly independent. This condition is

useful, but it is stated in terms of unknown parameters. Consequently, it can only be used after the DCDD model is fitted. However, there is one clear case in which it can be determined a priori if the condition is satisfied. Recall that when no rank restrictions are imposed on any \mathbf{M}_i , there is an analytic solution. The \mathbf{M}_i can be uniquely determined if and only if matrix \mathbf{A} defined in (8) has full column rank. This condition is sufficient, but not necessary for the uniqueness of parameter estimates in the rank-restricted case. When \mathbf{A} has full column rank, Kiers and Takane's condition is automatically satisfied. However, \mathbf{A} being singular does not imply that parameters are nonunique. In the example to be presented in section 4.2, $\mathbf{G} = [\mathbf{I}, \mathbf{I}]$ and $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$, neither of which are of full column rank, and consequently $\mathbf{A} = [\mathbf{I} \otimes \mathbf{H}_1, \mathbf{I} \otimes \mathbf{H}_2]$ is not of full column rank, yet Kiers and Takane's condition is satisfied, and consequently uniqueness of parameter estimates is guaranteed.

There is a case in which uniqueness is definitely lost. This is the case where $\mathbf{G}_i = \mathbf{G}$ and $\mathbf{H}_i = \mathbf{H}$ for all i . In this instance all \mathbf{u}_j 's are contained in all $S_p(\mathbf{G}_i)$ and all \mathbf{v}_j 's are contained in all $S_p(\mathbf{H}_i)$, and Kiers and Takane's condition is clearly violated. There is indeed no unique solution in this case, because this case reduces to CPCA in which representations of \mathbf{GMH}' are not unique. In CPCA, the SVD of $\mathbf{P}_G \mathbf{ZP}_H$ gives only one of all possible solutions. If DCDD is applied in this case, it obtains one of the solutions, which is related to the SVD of $\mathbf{P}_G \mathbf{ZP}_H$ by

$$\mathbf{UDV}' = \mathbf{F}_U \mathbf{R}'_U \mathbf{DR}_V \mathbf{F}'_V = \mathbf{F}_U \mathbf{PD}^* \mathbf{Q}' \mathbf{F}'_V = \mathbf{U}^* \mathbf{D}^* \mathbf{V}^*{}', \quad (27)$$

where \mathbf{UDV}' is a DCDD solution, $\mathbf{U}^* \mathbf{D}^* \mathbf{V}^*{}'$ is the SVD of $\mathbf{P}_G \mathbf{ZP}_H$, $\mathbf{F}_U \mathbf{R}'_U$ and $\mathbf{F}_V \mathbf{R}'_V$ are QR decompositions of \mathbf{U} and \mathbf{V} , respectively, and $\mathbf{PD}^* \mathbf{Q}'$ is the SVD of $\mathbf{R}'_U \mathbf{DR}_V$.

The problem of multiple local minima can also be investigated using a similar numerical technique. It should be pointed out that the problem does exist, although it seems rare in practice. For all the analysis results to be presented in section 4, solutions were obtained with 10,000 random initial starts, but in no cases were multiple local minima found. It seems that the problem of multiple local minima is rather rare in practice, particularly when estimated components are nearly orthogonal to one another.

The algorithm presented in section 2.2 is generally very efficient. Most often it runs faster than PCA for extracting the same number of components. This is due perhaps to the additional structure built into the model by the DCDD type of constraints. We have observed, however, one instance in which convergence of the algorithm was very slow. It turned out that this was due to an extremely high correlation among the estimated components. High correlations among the constraint sets tend to slow down the convergence of the algorithm by creating "near" multicollinearity. Highly correlated constraint sets should be avoided for the sake of smooth convergence of the algorithm.

3.2. Relations to Other Approaches

Böckenholt and Böckenholt (1990) proposed a method, called CALC (Canonical Analysis with Linear Constraints), which is similar to, but distinct from DCDD. For ease of comparison, assume $m = 2$, and $r_i = 1$ ($i = 1, 2$). There are two ways to incorporate linear constraints (Takane et al., 1991). CALC uses the null space method in which estimates of \mathbf{u}_i and \mathbf{v}_i are obtained under $\mathbf{R}'_i \mathbf{u}_i = \mathbf{0}$ and $\mathbf{C}'_i \mathbf{v}_i = \mathbf{0}$. DCDD, on the other hand, primarily uses the reparametrization method in which \mathbf{u}_i and \mathbf{v}_i are reparametrized as $\mathbf{u}_i = \mathbf{G}_i \mathbf{u}_i^*$ and $\mathbf{v}_i = \mathbf{H}_i \mathbf{v}_i^*$. The two methods are equivalent when \mathbf{R}_i and \mathbf{G}_i , and \mathbf{C}_i and \mathbf{H}_i , are orthocomplement to each other. Such relationships will be assumed between \mathbf{R}_i and \mathbf{G}_i , and between \mathbf{C}_i and \mathbf{H}_i . Then,

$$\mathbf{P}_{G(i)} = \mathbf{Q}_{R(i)},$$

and

$$\mathbf{P}_{H(i)} = \mathbf{Q}_{C(i)},$$

for $i = 1, 2$, where $\mathbf{P}_{G(i)}$ and $\mathbf{P}_{H(i)}$ are projection operators onto column spaces of \mathbf{G}_i and \mathbf{H}_i , respectively, and $\mathbf{Q}_{R(i)} = \mathbf{I} - \mathbf{P}_{R(i)}$ and $\mathbf{Q}_{C(i)} = \mathbf{I} - \mathbf{P}_{C(i)}$.

CALC is sequential, but not iterative. It first obtains $\mathbf{u}_1, \mathbf{v}_1$, and d_1 that minimize $SS(\mathbf{Z} - d_1 \mathbf{u}_1 \mathbf{v}_1')$ subject to $\mathbf{R}'_1 \mathbf{u}_1 = \mathbf{0}$ and $\mathbf{C}'_1 \mathbf{v}_1 = \mathbf{0}$. This amounts to obtaining the largest singular value (d_1) and the associated singular vectors (\mathbf{u}_1 and \mathbf{v}_1) of $\mathbf{Q}_{R(1)} \mathbf{Z} \mathbf{Q}_{C(1)} = \mathbf{P}_{G(1)} \mathbf{Z} \mathbf{P}_{H(1)}$. CALC then obtains $\mathbf{u}_2, \mathbf{v}_2$, and d_2 that minimize $SS(\mathbf{Z}_1 - d_2 \mathbf{u}_2 \mathbf{v}_2')$, where $\mathbf{Z}_1 = \mathbf{Q}_{u(1)} \mathbf{Z} \mathbf{Q}_{v(1)}$, subject to $\mathbf{R}'_2 \mathbf{u}_2 = \mathbf{0}, \mathbf{C}'_2 \mathbf{v}_2 = \mathbf{0}, \mathbf{u}'_1 \mathbf{u}_2 = 0$, and $\mathbf{v}'_1 \mathbf{v}_2 = 0$. This amounts to obtaining the largest singular value (d_2) and the associated singular vectors (\mathbf{u}_2 and \mathbf{v}_2) of $\mathbf{P}_{QG(2)} \mathbf{Z} \mathbf{P}_{QH(2)}$. Here, $\mathbf{P}_{QG(2)}$ and $\mathbf{P}_{QH(2)}$ are projection operators defined by matrices $\mathbf{Q}_{u(1)} \mathbf{G}_2$ and $\mathbf{Q}_{v(1)} \mathbf{H}_2$, respectively. The $\mathbf{P}_{QG(2)} \mathbf{Z} \mathbf{P}_{QH(2)}$ reduces to $\mathbf{P}_{G(2)} \mathbf{Z} \mathbf{P}_{H(2)}$, if $\mathbf{u}'_1 \mathbf{G}_2 = \mathbf{0}'$ and $\mathbf{v}'_1 \mathbf{H}_2 = \mathbf{0}'$.

The first cycle of DCDD obtains the same $\mathbf{u}_1, \mathbf{v}_1$, and d_1 assuming $\mathbf{Z}_{(-1)} = \mathbf{Z}$. It then obtains the largest singular value (d_2) and the associated singular vectors (\mathbf{u}_2 and \mathbf{v}_2) of $\mathbf{P}_{G(2)} \mathbf{Z}_{(-2)} \mathbf{P}_{H(2)}$, where $\mathbf{Z}_{(-2)} = \mathbf{Z} - \mathbf{P}_{u(1)} \mathbf{Z} \mathbf{P}_{v(1)}$. Thus,

$$\mathbf{P}_{G(2)} \mathbf{Z}_{(-2)} \mathbf{P}_{H(2)} = \mathbf{P}_{G(2)} (\mathbf{Z} - \mathbf{P}_{u(1)} \mathbf{Z} \mathbf{P}_{v(1)}) \mathbf{P}_{H(2)},$$

which reduces to $\mathbf{P}_{G(2)} \mathbf{Z} \mathbf{P}_{H(2)}$ if $\mathbf{u}'_1 \mathbf{G}_2 = \mathbf{0}'$ or $\mathbf{v}'_1 \mathbf{H}_2 = \mathbf{0}'$. Matrix $\mathbf{P}_{G(2)} \mathbf{Z}_{(-2)} \mathbf{P}_{H(2)}$ is generally not equivalent to matrix $\mathbf{P}_{QG(2)} \mathbf{Z} \mathbf{P}_{QH(2)}$. The two matrices are equal only when $\mathbf{G}'_1 \mathbf{G}_2 = \mathbf{0}$ and $\mathbf{H}'_1 \mathbf{H}_2 = \mathbf{0}$. This means that even if $\mathbf{H}'_1 \mathbf{H}_2 = \mathbf{0}$, but $\mathbf{G}_1 = \mathbf{G}_2 = \mathbf{I}$ (no row constraints), CALC and DCDD will not be equivalent. CALC obtains the SVD of $\mathbf{Q}_{u(1)} \mathbf{Z} \mathbf{P}_{H(2)}$ for the second dimension, whereas DCDD obtains the SVD of $\mathbf{Z} \mathbf{P}_{H(2)}$.

Dimension extractions are sequential in CALC. There is no global fitting criterion; instead, a separate criterion is minimized for each dimension extraction. Consequently, results depend on which constraint set is accounted for first, second, and so on, unless both row and column constraint sets are mutually orthogonal. Successive solution vectors are, however, orthogonal, irrespective of the orthogonality of the constraint sets. In DCDD, on the other hand, there is a global fitting criterion which is minimized simultaneously across all dimensions, but the solution vectors need not be mutually orthogonal.

We programmed CALC, and analyzed the example data set to be reported in section 4.2. Although no details will be given, we confirmed that all our theoretical expectations about CALC given above were correct. For example, the factorial structure of a second dimension is destroyed, following extraction of a correlated first dimension. That is, coordinate values which are supposed to be equal on the second dimension are no longer equal.

Heiser (1987) considered the DCDD type of constraints in the context of correspondence analysis (CA), and developed an elegant algorithm for incorporating them. However, his method minimizes, in line with the usual (unconstrained) CA,

$$f^* = \text{tr}(\mathbf{U}'\mathbf{K}\mathbf{U}) + \text{tr}(\mathbf{V}'\mathbf{K}\mathbf{V}) - 2 \text{tr}(\mathbf{U}'\mathbf{F}\mathbf{V}),$$

where \mathbf{F} is a contingency table, and \mathbf{K} and \mathbf{L} are diagonal matrices of row and column totals of \mathbf{F} , respectively. The minimization of f^* will be equivalent to that of

$$f = SS(\mathbf{Z} - \mathbf{GMH}')_{K,L},$$

where $\mathbf{Z} = \mathbf{K}^{-1} \mathbf{F} \mathbf{L}^{-1}$, if $\mathbf{U}'\mathbf{K}\mathbf{U} = \mathbf{I}$ or $\mathbf{V}'\mathbf{L}\mathbf{V} = \mathbf{I}$. However, neither of these conditions are generally satisfied in DCDD. Thus, Heiser's method is also distinct from our method.

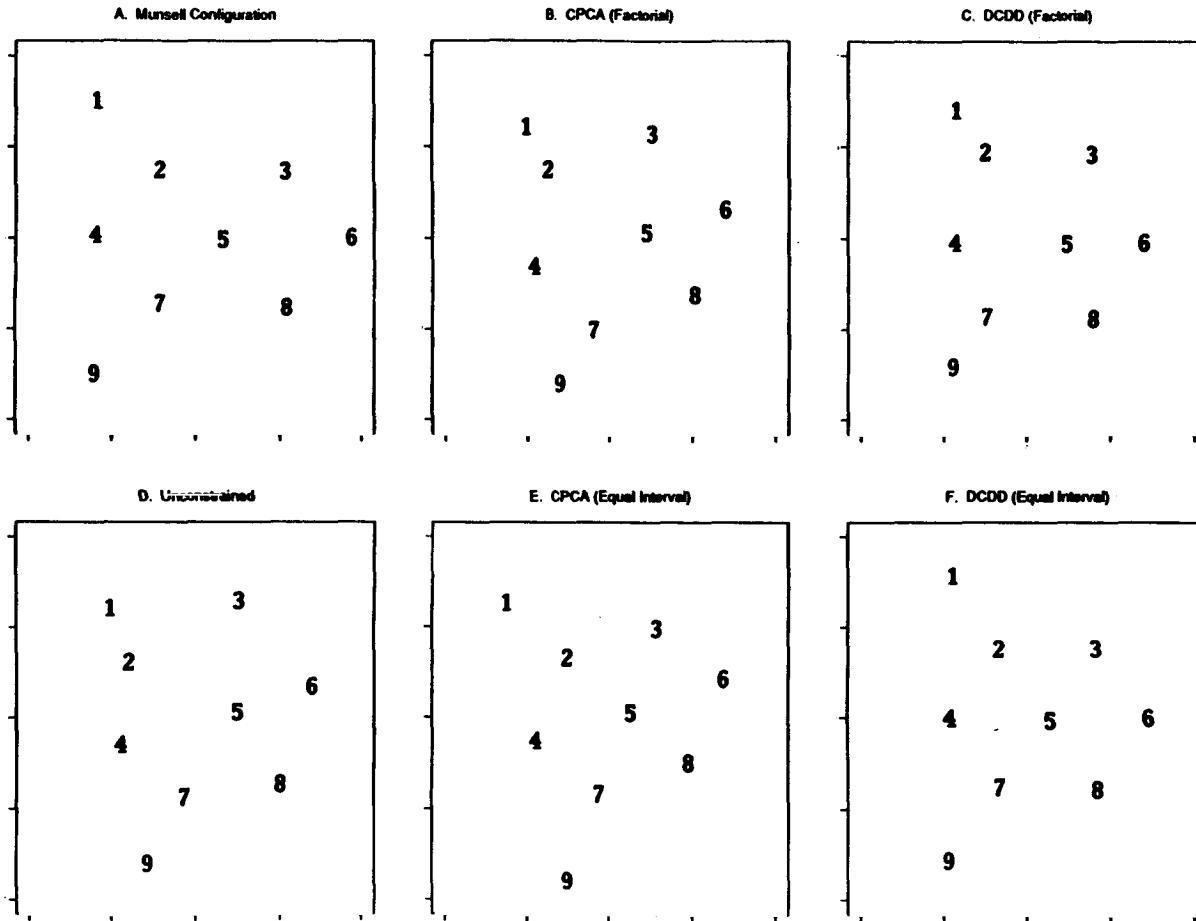


FIGURE 1.

The Munsell configuration (A) of nine colors varying in chroma and value dimensions, and stimulus configurations derived from Torgerson's (1958) data under various hypotheses.

TABLE 1
Design Matrices for Torgerson's (1958) Data

Stimulus	Factorial								Equal Interval	
	Chroma (H_1)				Value (H_2)				Chroma (H_1)	Value (H_2)
	h_{11}	h_{12}	h_{13}	h_{14}	h_{21}	h_{22}	h_{23}	h_{24}	h_1	h_2
1	2	0	0	0	1	3	0	4	-14	2
2	-1	2	0	0	0	0	1	-5	-5	1
3	-1	-1	1	0	0	0	1	-5	13	1
4	2	0	0	0	0	-2	0	4	-14	0
5	-1	-1	-1	1	0	-2	0	4	4	0
6	-1	-1	-1	-1	0	-2	0	4	22	0
7	-1	2	0	0	0	0	-1	-5	-5	-1
8	-1	-1	1	0	0	0	-1	-5	13	-1
9	2	0	0	0	-1	3	0	4	-14	-2

Correlations:					
	h_{21}	h_{22}	h_{23}	h_{24}	
h_{11}	0	0	0	0	
h_{12}	.516	.211	.365	0	Cor(h_1, h_2) = 0
h_{13}	0	0	0	0	
h_{14}	.632	-.387	-.671	0	

4. Examples of Application

Two examples in this section illustrate practical uses of DCDD. In both examples, we systematically compare unconstrained solutions, CPCA solutions, and DCDD solutions.

4.1. Torgerson's Color Data

Torgerson (1958) collected dissimilarity data on the nine colors presented in Panel A of Figure 2. He applied "classical" multidimensional scaling to his data, which amounted to the eigenvalue decomposition of the scalar product matrix derived from his original data by the Young and Householder (1938) transformation. This solution is presented in Panel D.

Recall that the nine colors have a factorial structure according to the Munsell System. This structure is captured in the stimulus design matrices, H_1 and H_2 , specified in the contrast form and given in the upper left portion of Table 1. These contrast vectors are obtained in a manner similar to that in ANOVA. There are five levels in the chroma dimension, and four contrast vectors are needed to distinguish among them. The h_{11} contrasts stimuli 1, 4 & 9 with all the rest, h_{12} stimuli 2 & 7 vs 3, 5, 6, & 8, h_{13} stimuli 3 & 8 vs 5 & 6, and h_{14} stimulus 5 vs 6. These contrast vectors are not unique. Correlations between H_1 and H_2 are given in the left side bottom of Table 1. It can be seen that they are not mutually orthogonal. The factorial structure does not assume

TABLE 2
Summary Results for Torgerson's (1958) Data

		Percent SS Explained	Number of Parameters
D.	Unconstrained	98.78	15
	Constrained (Factorial)		
B.	CPCA	97.72	13
C.	DCDD	96.77	8
	Constrained (Equal Interval)		
E.	CPCA	93.87	3
F.	DCDD	93.58	2
A.	Equal Interval Across Dimensions		
	DCDD (C:V = 1:2)	85.78	1
	DCDD (C:V = 3:8)	93.57	1

equal intervals between adjacent levels. Design matrices under this assumption are presented in the upper right corner. They happen to be mutually orthogonal.

The two design matrices, H_1 and H_2 , are joined into a single matrix, $H = [H_1, H_2]$, in CPCA solutions. CPCA in this case amounts to the eigenvalue decomposition of $P_H S P_H$. The CPCA solution with the factorial design is presented in Panel B. It is strikingly similar to the unconstrained solution (Panel D). The CPCA solutions under the equal interval hypothesis is displayed in Panel E. Both of the CPCA configurations are rotated relative to the Munsell configuration, and their dimensions do not correspond one-to-one with the Munsell dimensions. The DCDD analysis, on the other hand, extracts dimensions which coincide with the hypothesized dimensions. The DCDD solution under the dimensionwise factorial hypothesis is shown in Panel C. Stimuli having equal coordinate values on Munsell dimensions take equal coordinate values on the corresponding DCDD dimensions. Interestingly, derived v_1 and v_2 are nearly orthogonal, despite the fact that H_1 and H_2 are not mutually orthogonal. The DCDD configuration is displayed in Panel F. This configuration differs from the Munsell configuration (Panel A) only in one respect. Whereas in the latter, two chroma units are assumed equivalent to one value unit, no such assumption is made in F. Adjacent levels are assumed equally spaced within dimensions, but not across dimensions. An across-dimension equal interval hypothesis can be incorporated via across-dimension constraints, where $C^* = [1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$ combined with h_1 and h_2 given in Table 1. This C^* stipulates the chroma interval defined by Stimuli 1 and 2 is equal to the value interval defined by the same pair of stimuli. The configuration derived under this hypothesis is identical to the Munsell configuration (Panel A).

The unconstrained solution is least similar to the Munsell configuration, while it provides the best fit to the data, as indicated by the largest proportion of *SS* accounted for by the solution (see Table 2). The DCDD solution (F) derived under the within-dimension equal interval assumption still accounts for more than 93% of the total *SS*, only a 5% reduction from the unconstrained solution. The Munsell configuration (A) derived under the across-dimension equal interval hypothesis, on the other hand, accounts for less than 86% of the total *SS*, a considerable drop in fit from that of solution

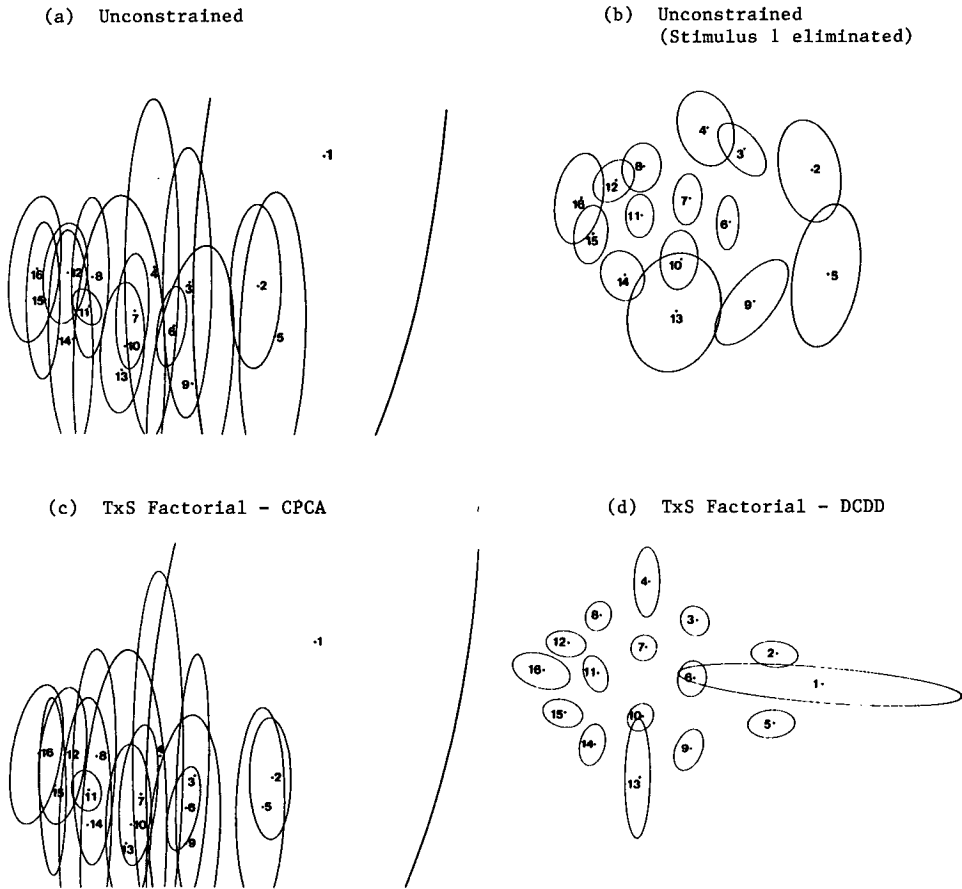


FIGURE 2.

Stimulus configurations derived from Delbeke's (1978) data under various hypotheses. (See Figure 2b.)

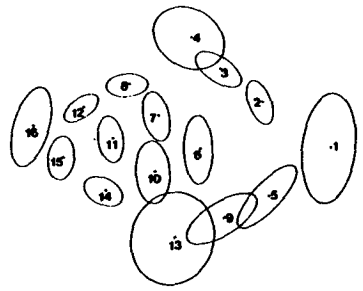
F. Solution F indicates that eight chroma units are roughly equivalent to three value units. When this hypothesis was incorporated, the fit turned out to be 93.57%, which was almost as good as that of solution F.

Table 2 also provides the numbers of parameters estimated in various solutions. Let m denote the dimensionality of the representation space, p the number of stimuli, t_i the number columns in H_i ($i = 1, 2$), and $t = \text{rank}(\mathbf{H})$ where $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$. Then, the effective number of parameters is given by $(p - 1)m - m(m - 1)/2$ for the unconstrained solution. The second term is subtracted because of rotational indeterminacy. This also applies to CPCA solutions in which the number of parameters is calculated by $tm - m(m - 1)/2$. There is no rotational indeterminacy in DCDD in which the number of parameters is given by $t_1 + t_2$.

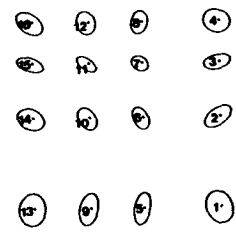
4.2. Delbeke's Data

Delbeke (1978) constructed a set of stimuli by systematically varying the number of boys and the number of girls in a family. By factorially combining four levels (0 to 3) each of the two variables, 16 stimuli were constructed. The stimuli are numbered from 1 to 16 by varying the number of boys first from 0 to 3, and then the number of girls in a similar way. Stimulus 1 thus represents the combination of 0 boys and 0 girls,

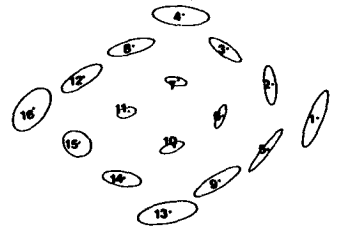
(e) BxG Factorial - CPCA



(f) BxG Factorial - DCDD



(g) TxS, BxG Equal Interval - CPCA
(The two solutions are identical)



(h) TxS Equal Interval - DCDD



(i) BxG Equal Interval - DCDD



FIGURE 2.
(continued)

TABLE 3

Design Matrices (TxS) for Delbeke's (1978) Data

Stimulus	Factorial						Equal Interval							
	Total (H_1)						Bias (H_2)						Total Bias	
	h_{11}	h_{12}	h_{13}	h_{14}	h_{15}	h_{16}	h_{21}	h_{22}	h_{23}	h_{24}	h_{25}	h_{26}	h_1	h_2
1	1	0	0	2	1	-1	0	0	0	0	0	3	-3	0
2	0	1	0	-1	1	-1	0	0	1	0	-1	-1	-2	-1
3	0	0	1	0	-1	-1	0	1	0	-1	1	-1	-1	-2
4	0	0	0	0	0	3	1	0	0	2	1	-1	0	-3
5	0	1	0	-1	1	-1	0	0	-1	0	-1	-1	-2	1
6	0	0	1	0	-1	-1	0	0	0	0	0	3	-1	0
7	0	0	0	0	0	3	0	0	1	0	-1	-1	0	-1
8	0	0	-1	0	-1	-1	0	1	0	-1	1	-1	1	-2
9	0	0	1	0	-1	-1	0	-1	0	-1	1	-1	-1	2
10	0	0	0	0	0	3	0	0	-1	0	-1	-1	0	1
11	0	0	-1	0	-1	-1	0	0	0	0	0	3	1	0
12	0	-1	0	-1	1	-1	0	0	1	0	-1	-1	2	-1
13	0	0	0	0	0	3	-1	0	0	2	1	-1	0	3
14	0	0	-1	0	-1	-1	0	-1	0	-1	1	-1	1	2
15	0	-1	0	-1	1	-1	0	0	-1	0	-1	-1	2	1
16	-1	0	0	2	1	-1	0	0	0	0	0	3	3	0

Correlations (in metric D_c):

	h_{21}	h_{22}	h_{23}	h_{24}	h_{25}	h_{26}	
h_{11}	.027	.030	.023	-.026	-.036	-.392	
h_{12}	.029	.031	-.016	-.028	.310	.176	
h_{13}	.015	-.007	.012	.116	-.127	-.047	
h_{14}	.011	.013	-.069	-.011	.349	.509	Cor(h_1, h_2) = -.031
h_{15}	.019	-.133	.081	.355	-.643	-.275	
h_{16}	.182	-.079	.042	.514	-.272	-.393	

stimulus 2 that of 1 boy and 0 girls, and so on. Delbeke asked 82 university students in Belgium to rank order the 16 stimuli according to their preference.

Heiser (1981) previously analyzed the data by unfolding analysis (Coombs, 1964). He obtained both unconstrained and constrained unfolding solutions. The latter was obtained under similar constraints to those used in the present study. He also applied correspondence analysis, which is a special form of unfolding analysis. We extend this line of analysis by incorporating constraints.

By construction, the stimuli have a factorial structure in terms of the number of boys and the number of girls (the $B \times G$ hypothesis). Alternatively, the stimuli can be characterized by the total number of children ($T = B + G$) and the sex bias ($S = B - G$). These two structural designs were coded into matrices of contrasts in a manner similar to that in the previous example, and are shown in the upper portions of Table 3 (the $T \times S$ hypothesis) and Table 4 (the $B \times G$ hypothesis). Correlations among them are given in the bottom portions of the tables. These correlations were calculated with

TABLE 4

Design Matrices (BxG) for Delbeke's (1978) Data

Stimulus	Factorial						Equal Interval	
	Boys (H_1)			Girls (H_2)			Boys (H_1)	Girls (H_2)
	h_{11}	h_{12}	h_{13}	h_{21}	h_{22}	h_{23}	(h_1)	(h_2)
1	1	0	1	1	0	1	-3	-3
2	-1	0	1	1	0	1	-1	-3
3	0	1	-1	1	0	1	1	-3
4	0	-1	-1	1	0	1	3	-3
5	1	0	1	-1	0	1	-3	-1
6	-1	0	1	-1	0	1	-1	-1
7	0	1	-1	-1	0	1	1	-1
8	0	-1	-1	-1	0	1	3	-1
9	1	0	1	0	1	-1	-3	1
10	-1	0	1	0	1	-1	-1	1
11	0	1	-1	0	1	-1	1	1
12	0	-1	-1	0	1	-1	3	1
13	1	0	1	0	-1	-1	-3	3
14	-1	0	1	0	-1	-1	-1	3
15	0	1	-1	0	-1	-1	1	3
16	0	-1	-1	0	-1	-1	3	3

Correlations (in metric D_c):

	h_{21}	h_{22}	h_{23}	
h_{11}	.057	-.021	-.042	
h_{12}	-.036	.033	.004	Cor(h_1, h_2) = .062
h_{13}	-.037	-.012	-.033	

metric D_c , so that perfectly balanced designs can produce nonzero correlations. The data were analyzed by correspondence analysis, CPCA, and DCDD. In CPCA and DCDD, the two factorial hypotheses were fitted with and without the within-dimension equal interval assumption. In no cases were constraints imposed on the subjects.

Derived stimulus configurations are presented in Figure 2. These configurations exhibit stimulus points only. Biplots of stimulus points and subject vectors could have been informative (Gabriel, 1971; ter Braak, 1990), but there were too many subjects to be plotted for the size of each panel. Ellipses surrounding estimated points are 95% confidence regions obtained by the Bootstrap method (Efron, 1979). They were obtained by generating 100 Bootstrap samples, analyzing them separately, and calculating variances and covariances of estimated point coordinates. Under the asymptotic normality assumption, these variance and covariance estimates can be turned into ellipses that indicate the degrees of reliability of the estimated points.

The unconstrained solution is displayed in Panel a. It appears that the horizontal direction roughly corresponds with the total number of children, and the vertical direction with the sex bias, but this is obscured by large variabilities in the estimated point locations. The instability is largely due to point 1 (the combination of 0 boys and 0 girls) which is so unpopular. When the analysis was repeated with Point 1 removed, the remaining points were much more reliably estimated (see Panel b). These findings are consistent with Heiser (1981). Panel c presents the CPCA solution with the $T \times S$

factorial structure. This configuration is remarkably similar to the unconstrained solution. Obviously, the constraints are not very restrictive. The $T \times S$ design, being incomplete factorial, allows quite a bit of freedom. In particular, stimulus 1 is the only stimulus with the total number of children equal to zero. The DCDD analysis with the same $T \times S$ design yields a much more stable configuration (see Panel d), although Point 1 still exhibits considerable instability along the horizontal direction. The CPCA solution with the $B \times G$ design is displayed in Panel e. This configuration rotates the prescribed $B \times G$ dimensions approximately 45° , so that the horizontal direction roughly corresponds with the T dimension and the vertical direction the S dimension. Although confidence ellipses are quite different, the configuration of points itself looks very much like the DCDD configuration obtained under the $T \times S$ hypothesis. The DCDD solution with the $B \times G$ design is displayed in Panel f.

The three configurations shown in the bottom row of Figure 2 were obtained under the within-dimension equal interval assumption. Two CPCA solutions, one obtained under the $T \times S$ hypothesis and the other under the $B \times G$ hypothesis, turned out to be identical, and are shown in Panel g. This is due to the fact that the $T \times S$ design matrix and the $B \times G$ design matrix happen to span an identical space.

The DCDD solutions under equal interval hypotheses are shown in Panel h and in Panel i. It can be observed that the more stringent the constraints are, the more reliable the point estimates become. This is indicated by tighter confidence regions in the configurations derived under the more stringent constraints.

Table 5 presents proportions of SS accounted for by various solutions along with Bootstrap results; means, standard deviations, maxima and minima of the percent SS . The unconstrained solution accounts for approximately two thirds of the total SS . This proportion reduces only by 4% in solutions in Panels g, h, and i. Considering the variabilities in the percent SS from the Bootstrap study, ranging roughly from 2.50 to 3.00, this is well within the range of sampling variability. The four solutions given in the bottom row of Figure 3 have an identical fit. We have already seen why the two CPCA solutions are identical. In addition, the two DCDD solutions (h and i) have an identical fit with each other and with the two CPCA solutions. In these solutions, no rank restrictions are imposed on \mathbf{M}_i , and \mathbf{G}_i 's are common across i . Under these circumstances, DCDD and CPCA yield an equivalent fit, because if we let $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_m]$ and $\mathbf{M} = [\mathbf{M}_1, \dots, \mathbf{M}_m]$, the DCDD model can be written as \mathbf{GMH}' , the full rank SVD of which gives the CPCA solution.

An across-dimension equal interval hypothesis can be incorporated via across-dimension constraints. The hypothesis that a one-unit difference in B is equivalent to a one-unit difference in G only negligibly worsened the goodness of fit. However, a similar hypothesis on the $T \times S$ dimensions resulted in a considerable drop in fit. Solution h indicates that a two-unit difference in S is roughly equivalent to a one-unit difference in T . This post hoc hypothesis yielded a fit of 63.41%, which was almost identical to the fit of Solution h. It is not clear which of the two hypothesized structures, $T \times S$ or $B \times G$, better accounts for the data.

It might be pointed out that similar observations to the above could perhaps be made by constrained unfolding analysis. However, DCDD is computationally much simpler and faster than unfolding analysis, which is also prone to degenerate solutions.

5. Concluding Remarks

Principal component analysis (PCA) is a useful technique for a spatial representation of a data matrix (e.g., Jackson, 1991; Jolliffe, 1986; Lebart, Morineau, & Warwick,

TABLE 5
Summary Results for Delbeke's (1978) Data

		Percent SS Explained	Number of Parameters	Bootstrap on Percent SS Explained			
				mean	sd	max.	min.
(a)	Unconstrained	67.45	29	69.24	2.65	75.03	62.81
	TxS; Factorial						
(c)	CPCA	66.69	21	68.40	2.68	74.49	60.67
(d)	DCDD	65.60	12	66.26	2.67	72.85	58.74
	TxS; Equal Interval						
(g)	CPCA	63.41*	2	64.05	2.82	70.61	52.78
(h)	DCDD	63.41*	2	64.05	2.82	70.61	52.78
	TxS; Equal Interval Across Dimensions						
	DCDD (T:S = 1:1)	56.99	1	58.40	2.15	63.88	52.73
	DCDD (T:S = 2:1)	63.41	1	63.08	3.18	69.88	54.90
	BxG; Factorial						
(e)	CPCA	64.91	11	65.48	2.65	72.71	58.43
(f)	DCDD	64.76	6	65.26	2.92	73.06	58.32
	BxG; Equal Interval						
(g)	CPCA	63.41*	2	64.05	2.82	70.61	52.78
(i)	DCDD	63.41*	2	64.05	2.82	70.61	52.78
	BxG; Equal Interval Across Dimensions						
	DCDD	63.34	1	63.51	3.00	69.81	55.70

*These solutions are identical in fit.

1984; Velicer & Jackson, 1990). The derived spatial representation captures the most prevailing tendency in the data in as few dimensions as possible.

The interpretation of the spatial representation can be greatly facilitated by external information incorporated through linear constraints (Böckenholt & Böckenholt, 1990; Carroll, et al., 1980; Takane & Shibayama, 1991; Takane, et al., 1991; ter Braak, 1986). Typically the resultant spatial representation is simpler, and exhibits a greater degree of stability (see section 4.2). External constraints may represent a theoretical hypothesis. By comparing constrained and unconstrained solutions, empirical validity of the hypothesis can be investigated. External constraints may allow certain useful predictions. For example, stimuli, which are not included in the original representation, can be *a posteriori* mapped into the representation via the constraints (Carroll, 1972). In some cases, external constraints are useful in eliminating incidental parameters from a model (van der Leeden, 1990).

In this paper a method (DCDD) to impose different sets of constraints on different dimensions was proposed. The DCDD type of constraints often represents original empirical hypotheses more accurately than the CPCA type, and consequently more meaningful analyses may be possible. A simple, yet versatile, algorithm to fit the dimensionwise constraints was given, and extended in various directions, including across-dimension constraints, symmetric data, etc. Although only two examples were given in this paper, a host of other examples can readily be generated. Simple structures

in factor analysis and analysis of MTMM (multitrait-multimethod) matrices are but two of the possibilities that immediately come to mind.

It may be useful to incorporate orthogonality or partial orthogonality constraints on U in DCDD. More generally, it will be interesting to incorporate various correlational patterns in $U'U$ which are implied, for example, by hypothesized "causal" relationships among the extracted dimensions (= latent variables). This makes DCDD more similar in spirit to the PLS approach (Wold, 1982; Lohmöller, 1989) to structural equation models. Both analyze original data matrices rather than covariance matrices. However, the two approaches are radically different in the principle of algorithm construction. Whereas in DCDD a global minimization criterion (such as (3)) is consistently minimized, no such criterion exists in PLS. In the latter, convergence points are defined to be wherever its algorithm achieves a state of "equilibrium".

Throughout this paper we have primarily focused on the analysis of structural parts of data. However, equally important is the analysis of initially unstructured portions of data, that is, the analysis of residuals from prescribed structures. The current DCDD program is also capable of analyzing the residuals by PCA.

References

- Andersen, E. B. (1980). *Discrete statistical models with social science applications*. Amsterdam: North-Holland.
- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distribution. *Annals of Mathematical Statistics*, 22, 327-351.
- Böckenholt, U., & Böckenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55, 633-639.
- Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling, Vol. 1*, (pp. 105-155). New York: Seminar Press.
- Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45, 3-24.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Delbeke, L. (1978). Enkele analyses op voorkeuroordelen voor gezinssamenstellingen [Some analyses of preference judgments for family compositions]. Centrum voor Mathematische Psychologie en Psychologische Methodologie, Katholieke Universiteit Leuven.
- de Leeuw, J. (1984). The Gifi-system of nonlinear multivariate analysis. In E. Diday, M. Jambu, L. Lebart, J. Pagès, & R. Thomassone (Eds.), *Data analysis and informatics III* (pp. 415-424). Amsterdam: North Holland.
- de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, 471-503.
- DeSarbo, W. S., & Rao, V. R. (1984). GENFOLDS: A set of models and algorithms for the GENERAL UNFOLDING analysis of preference/dominance data. *Journal of Classification*, 1, 147-186.
- Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 7, 1-26.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467.
- Golub, G. H. (1973). Some modified eigenvalue problems. *SIAM Review*, 15, 318-335.
- Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.
- Greenacre, M. J. (1993). *Correspondence analysis in practice*. London: Academic Press.
- Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished doctoral dissertation, University of Leiden.
- Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189-221). Berlin: Springer Verlag.
- Heiser, W. J., & Meulman, J. (1983). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement*, 7, 381-404.
- Jackson, J. E. (1991). *A user's guide to principal components*. New York: Wiley.
- Jolliffe, I. T. (1986). *Principal component analysis*. Berlin: Springer Verlag.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, 57, 239-251.

- Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, 27, 887–906.
- Kiers, H. A. L., & Krijnen, W. P. (1991). An efficient algorithm for PARAFAC of three-way data with large numbers of observation units. *Psychometrika*, 56, 147–152.
- Kiers, H. A. L., Kroonenberg, P. M., & ten Berge, J. M. F. (1992). An efficient algorithm for TUCKALS 3 on data with large numbers of observation units. *Psychometrika*, 57, 415–422.
- Kiers, H. A. L., & Takane, Y. (1993). Constrained DEDICOM. *Psychometrika*, 58, 339–355.
- Lebart, L., Morineau, A., & Warwick, K. (1984). *Multivariate descriptive statistical analysis*. New York: Wiley.
- Lohmöller, J. (1989). *Latent variable path modelling with partial least squares*. Heidelberg: Physica Verlag.
- Meredith, W., & Millsap, R. E. (1985). On component analysis. *Psychometrika*, 50, 495–507.
- Meulman, J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
- Meulman, J., & Heiser, W. (1984). Constrained multidimensional scaling: More directions than dimensions. In T. Havranek, et al. (Eds.), *COMPSTAT 84* (pp. 137–142). Vienna: Physica Verlag.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Ramsay, J. O. (1980). Joint analysis of direct ratings, pairwise preferences and dissimilarities. *Psychometrika*, 45, 149–165.
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya A*, 26, 329–358.
- Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 3–22). Amsterdam: North Holland.
- Takane, Y. (1978). A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent theory and evaluations. *Japanese Psychological Research*, 20, 7–17 and 105–114.
- Takane, Y. (1981). Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, 46, 9–28.
- Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, 56, 97–120.
- Takane, Y., Yanai, H., & Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, 56, 667–684.
- Takane, Y., Young, F. W., & de Leeuw, J. (1980). An individual differences additive model: An alternating least squares method with optimal scaling features. *Psychometrika*, 45, 183–209.
- ten Berge, J. M. F. (1986). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behavioral Research*, 21, 41–64.
- ten Berge, J. M. F., & Kiers, H. A. L. (1991). Some clarifications of the CANDECOMP algorithm applied to INDSCAL. *Psychometrika*, 56, 317–326.
- ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167–1179.
- ter Braak, C. J. F. (1990). Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika*, 55, 519–531.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika*, 42, 207–219.
- van der Burg, E., de Leeuw, J., & Verdegaal, R. (1988). Homogeneity analysis: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, 177–197.
- van der Lans, I. A. (1992). *Nonlinear multivariate analysis for multivariate preference data*. Leiden: DSWO Press.
- van der Leeden, R. (1990). *Reduced rank regression with structured residuals*. Leiden, The Netherlands: DSWO Press.
- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25, 1–28.
- Velu, R. P. (1991). Reduced rank models with two sets of regressors. *Applied Statistics*, 40, 159–170.
- Wold, H. (1982). Soft modelling: The basic design and some extensions. *Systems under indirect observations, II* (pp. 1–54). Amsterdam: North-Holland.
- Young, G., & Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3, 19–22.

Manuscript received 7/15/91

Final version received 5/1/94