CHAPTER FIVE

RELATING TWO SETS OF VARIABLES OVER A THIRD MODE

5.1 INTRODUCTION

In this chapter I present least squares methods for modeling CCA, CVA, RA and PR over a third mode. The term third mode refers to either multiple datasets or multiple occasions. These methods are generalizations of CCA, CVA, RA or PR in that they maintain some of their distinguishing features while reducing to the standard method when the number of occasions is one. The methods developed will be put in the framework of the PARAFAC (orth.) and the Tucker2 models. Like the PARAFAC (orth.) and Tucker2 models these methods are not inferential but exploratory. They can be used well in conjunction with graphical methods, although this topic is deferred until Chapter Six. The methods are flexible. One can model categorical data as well. Hence one has a generalization of correspondence analysis.

Let the variables be divided into two sets, X-variables and Y-variables. The motif running through all the models is that there are two sets of variates which model the linear relationship

between the two sets of variables. These two sets of variates are hypothesized to be common over the third mode, though the strength of the relationships is allowed to change.

Now, these models can be put in the framework of the PARAFAC (orth.) or Tucker2 models. In PARAFAC (orth.) based methods, the variates form pairs, one from each set, and the linear relationship is modeled as occurring strictly between the members of these pairs. The Tucker2 based methods model the linear relationship between all possible pairings of the members of the two sets of variates.

The chapter is organized as follows. Section **5.2** introduces the three-mode models, which I will call CCA/third, CVA/third, RA/third and PR/third. It is shown that they maximize sums of squares regression, sums of squares separation, sums of squared correlations or sums of squared covariances between the two sets of variates. Section **5.3** discusses how to evaluate the fit of a model and how to choose between the PARAFAC (orth.) model and the Tucker2. Section **5.4** presents an example based on the data from the study "Sensitivity of Stream Basins in Shenandoah National Park to Acid Deposition" (Lynch & Dise 1985). Lastly, Section **5.5** discusses several considerations of the methods presented in this chapter. These include autocorrelation, the covariances between observations at different occasions, and the invariance of solutions to different choices of transformation of the X-variables and Y-variables.

5.2 RELATING TWO SETS OF VARIABLES OVER A THIRD MODE

In this section I develop the three-mode models for extending CCA, RA and PR to three-mode data. Consider that CCA, RA and PR are defined in Section **2.2** as singular value decompositions on the transformed matrices of covariances between the X-variables and Y-variables, that is $\mathbf{S}_{XX}^{-½}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-½}$, $\mathbf{S}_{XX}^{-½}\mathbf{S}_{XY}$, or $\mathbf{S}_{XY}$. But three-mode models such as the Tucker2 and the PARAFAC (orth.) are generalizations of the SVD to three-mode data. As such they suggest a framework for modeling CCA, RA and PR for three-mode data by modeling the decomposition of $\mathbf{S}_{XX}^{-½}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-½}$, $\mathbf{S}_{XX}^{-½}\mathbf{S}_{XY}$, or $\mathbf{S}_{XY}$ over a third mode. This is the approach that shall be taken in this chapter.

Define $\mathbf{X}_k$ and $\mathbf{Y}_k$ as the $n_k \times m$ and $n_k \times p$ data matrices of the X-variables and Y-variables over the third mode, $k = 1,\ldots,g$. Then, in the most general sense what is modeled will be $\mathbf{X}_k^{+\prime}\mathbf{Y}_k^{+}$, where $\mathbf{X}_k^{+}$ and $\mathbf{Y}_k^{+}$ are $\mathbf{X}_k$ and $\mathbf{Y}_k$ multiplied by the appropriate transformations; i.e., $\mathbf{S}_{XX}^{-½}$, $\mathbf{S}_{YY}^{-½}$ or $\mathbf{I}$. The relationship between the two sets of variables is harbored in these $\mathbf{X}_k^{+\prime}\mathbf{Y}_k^{+}$ terms. On the other hand, $\mathbf{S}_{XX}$ and $\mathbf{S}_{YY}$ are viewed as suggesting the metric in which to perform the model fitting through transformations of $\mathbf{X}_k$ and $\mathbf{Y}_k$.

It is important to make the proper choice of which transformations to apply to $\mathbf{X}_k'\mathbf{Y}_k$. When relating two sets of variables at just one occasion, one has the choice of using CCA, RA or PR. The choice one makes depends on what the researcher wants to emphasize in his analysis (van de Geer 1984). However, when relating two sets of variables over a third mode, there will usually be just one appropriate three-mode analog method. The choice of transformations defines

the metric in which one wants to maximize fit and minimize lack of fit, and it defines CCA/third, CVA/third, RA/third or PR/third. This topic will be discussed in more detail in Section **5.2.5**.

5.2.1 Redundancy Analysis over a Third Mode

The first three-mode method I will discuss will be redundancy analysis over a third mode, or RA/third. RA/third directly generalizes RA to three-mode data. The RA/third model hypothesizes that there are redundancy variates which are constant or stable over occasion or group. However, the root variance explained between each pair of X-variates and Y-variates varies over occasion. To model RA/third a necessary assumption is that $\mathbf{S}_{XX}$ is common over a third mode. If $\mathbf{S}_{XX}$ is not derived from constant **X** over the third mode then it must be estimated from the data.

The RA/third model is

$$\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k = \mathbf{W}\mathbf{H}_k\mathbf{V}', \tag{5.1}$$

for $k = 1,\ldots,g$, where **W** is the $m \times q$ matrix of uncorrelated common canonical variates for the X-variables, **V** is an $p \times r$ orthonormal matrix of redundancy variates for the Y-variates, and $\mathbf{H}_k$ is $q \times r$ matrix whose elements are root of the variance explained of the Y-variates, $\mathbf{V}'\mathbf{y}$, by the X-variates, $\mathbf{W}'\mathbf{x}$ (**y** and **x** represent vectors of random variables). If the X-variables are constant over the third mode, such as may be the case with longitudinal data, replace $\mathbf{X}_k$ by **X**.

Note the choice of weight for $\mathbf{X}_k'\mathbf{Y}_k$ in (5.1) of $\frac{1}{n_k - 1}$. This weighting implies one is modeling $\mathbf{S}_{XY}$. If one is modeling data from multiple datasets with different sample sizes this gives each dataset the same weight in the analysis. However, one may just as easily model $\mathbf{X}_k'\mathbf{Y}_k$ instead, yielding an analysis that effectively weights by sample size. All of the results in this chapter and in Chapter Six apply with minor modifications to this alternative weighting. Also, if one has longitudinal data n_k may be replaced by a constant *n*.

Finding **V** and **W** such that the total variance explained of $\mathbf{V}'\mathbf{y}$ when regressed on $\mathbf{W}'\mathbf{x}$ is maximized is equivalent to modeling the following in the Tucker2 or PARAFAC (orth.) framework:

$$\frac{1}{n_k - 1}\mathbf{X}_k^{*\prime}\mathbf{Y}_k = \mathbf{W}^*\mathbf{H}_k\mathbf{V}' + \text{Error}_k, \tag{5.2}$$

where $\mathbf{W}^*$ is the $m \times r$ orthonormal matrix $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}\mathbf{W}$, $\mathbf{X}_k^* = \mathbf{X}_k\mathbf{S}_{XX}^{-1/2}$, and **V** and $\mathbf{H}_k$ are defined as for (5.1). To see that minimizing the sums of squares of the error term in (5.2) maximizes the sum of the variances explained of the Y-variables, recognize first that $\mathbf{H}_k[i, j]$ is indeed the root of the variation of the j^{th} Y-variate, $\mathbf{v}_j'\mathbf{y}$ explained by the i^{th} X-variate, $\mathbf{w}_i'\mathbf{x}$.

This is so because for the Tucker2 solution $\mathbf{H}_k = \mathbf{W}^{*\prime}\mathbf{C}_k\mathbf{V}$, and for the PARAFAC (orth.) solution $\mathbf{H}_k = \text{diag}\left(\mathbf{W}^{*\prime}\mathbf{C}_k\mathbf{V}\right)$ (Kroonenberg 1983), where $\mathbf{C}_k = \frac{1}{n_k - 1}\mathbf{X}_k^{*\prime}\mathbf{Y}_k$. Thus

$$\mathbf{H}_k[i, j] = \frac{1}{n_k - 1}\mathbf{w}_i^{*\prime}\mathbf{X}_k^{*\prime}\mathbf{Y}_k\mathbf{v}_j = \frac{1}{n_k - 1}\mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k\mathbf{v}_j.$$

But the variance of $\mathbf{v}_j'\mathbf{y}$ explained in a regression against $\mathbf{w}_i'\mathbf{x}$ is

$$\frac{1}{n_k - 1}\mathbf{v}_j'\mathbf{Y}_k'\mathbf{X}_k\mathbf{w}_i(\mathbf{w}_i'\mathbf{X}_k'\mathbf{X}_k\mathbf{w}_i)^{-1}\mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k\mathbf{v}_j = \left(\frac{1}{n_k - 1}\right)^2\left(\mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k\mathbf{v}_j\right)^2,$$

noting that $(\mathbf{w}_i'\mathbf{X}_k'\mathbf{X}_k\mathbf{w}_i)^{-1} = \frac{1}{n_k - 1}$. Next, by **Proposition 3.2** the sums of squares of the $\mathbf{H}_k$ is maximized. Since the sums of squares of the $\frac{1}{n_k - 1}\mathbf{X}_k^{*\prime}\mathbf{Y}_k$ terms represent the total sums of squares of the Y-variables explainable by the X-variables, the lack of fit being minimized in (5.2) is just these sums of squares explainable by the relationship that are not being fit by the model over the third mode.

5.2.2 Canonical Variate Analysis over a Third Mode

CVA/third is a direct generalization of CVA. It is appropriate in the following situation; when the data have unchanging group structure, that is, the X-variables are group indicators; and the within-groups covariance matrix, $\mathbf{S}_{\text{YY(WITHIN)}}$, is stable over the third mode ($\mathbf{S}_{\text{YY(WITHIN)}}$ is distinguished from $\mathbf{S}_{\text{YY(TOTAL)}}$). Note that $\mathbf{S}_{\text{YY(WITHIN)}}$ has to be estimated from the data. The CVA/third scenario outlined here could also be approached by Campbell and Tomenson's (1983) model if the data are from multiple datasets or groups, or by the CVA/time model of Chapter Eight if the data are longitudinal.

Recall from Section **2.2.2** that CVA is equivalent to modeling the group means in the space of the variables transformed by the Mahalanobis transformation. CVA/third extends this conception of CVA to finding planes (components) that maximize the total dispersion over the third mode in the transformed space of the transformed variables.

The arguments for CVA/third are analogous to those of RA/third, except here one maximizes the variance explained in the transformed space of the Y-variables, i.e., the dispersion. The CVA/third model is

$$\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k = \mathbf{W}\mathbf{H}_k\mathbf{V}',$$

where $\mathbf{W}$ is the $m \times q$ matrix of uncorrelated common canonical variates for the X-variables, $\mathbf{V}$ is an $p \times r$ matrix of uncorrelated variates for the Y-variates, ($\mathbf{V}'\mathbf{S}_{\text{YY(WITHIN)}}\mathbf{V} = \mathbf{I}$) and $\mathbf{H}_k$ is a

$q \times r$ matrix of root transformed variances of the Y-variates, $\mathbf{V}'\mathbf{y}$, explained by the X-variates, $\mathbf{W}'\mathbf{x}$. If the X-variables are constant over the third mode, replace $\mathbf{X}_k$ with $\mathbf{X}$.

Finding $\mathbf{W}$ and $\mathbf{V}$ that maximize the sum of the transformed variance explained of the Y-variables by the X-variables is equivalent to modeling the following in a Tucker2 or PARAFAC (orth.) framework:

$$\frac{1}{n_k - 1}\mathbf{X}_k^{*\prime}\mathbf{Y}_k^{*} = \mathbf{W}^{*\prime}\mathbf{H}_k\mathbf{V}^{*} + \text{Error}_k, \tag{5.3}$$

where $\mathbf{W}^{*}$ is an $m \times r$ orthonormal matrix such that $\mathbf{W}^{*} = \mathbf{S}_{XX}^{1/2}\mathbf{W}$, $\mathbf{V}^{*}$ an $n \times r$ orthonormal matrix such that $\mathbf{V}^{*} = \mathbf{S}_{YY(WITHIN)}^{1/2}\mathbf{V}$, $\mathbf{H}_k$ is a $q \times r$ matrix, $\mathbf{X}_k^{*} = \mathbf{X}_k\mathbf{S}_{XX}^{-1/2}$ and $\mathbf{Y}_k^{*} = \mathbf{Y}_k\mathbf{S}_{YY}^{-1/2}$.

To see that minimizing the sums of squares error term of (5.3) maximizes the sum of the explained of the transformed variance, note first that $\mathbf{H}_k[i, j]$ is indeed the root of the transformed variation of the j^{th} Y-variate, $\mathbf{v}_j'\mathbf{y}$, explained by the i^{th} X-variate, $\mathbf{w}_i'\mathbf{x}$. This is so because for the Tucker2 solution $\mathbf{H}_k = \mathbf{W}'\mathbf{C}_k\mathbf{V}$, and for the PARAFAC (orth.) solution $\mathbf{H}_k = \text{diag}(\mathbf{W}'\mathbf{C}_k\mathbf{V})$ (Kroonenberg 1983), where $\mathbf{C}_k = \frac{1}{1 - n_k}\mathbf{X}_k^{*\prime}\mathbf{Y}_k^{*}$. Thus

$$\mathbf{H}_k[i, j] = \mathbf{w}_i^{*\prime}\mathbf{X}_k^{*\prime}\mathbf{Y}_k^{*}\mathbf{v}_j = \mathbf{w}_i'\mathbf{X}_k^{*\prime}\mathbf{Y}_k^{*}\mathbf{v}_j.$$

Now the variance (in the transformed space) of $\mathbf{v}_j'\mathbf{y}$ explained by $\mathbf{w}_i'\mathbf{x}$ is

$$\frac{1}{1 - n_k}\mathbf{v}_j'\mathbf{Y}_k^{*\prime}\mathbf{X}_k\mathbf{w}_i(\mathbf{w}_i'\mathbf{X}_k'\mathbf{X}_k\mathbf{w}_i)^{-1}\mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k^{*}\mathbf{v}_j = \left(\frac{1}{1 - n_k}\right)^2\left(\mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k^{*}\mathbf{v}_j\right)^2,$$

noting that $(\mathbf{w}_i'\mathbf{X}_k'\mathbf{X}_k\mathbf{w}_i)^{-1} = \frac{1}{n_k - 1}$. Next, by **Proposition 3.2** the sums of squares of the $\mathbf{H}_k$ are being maximized. Hence the error being minimized is the sums of squares explainable by the relationship that is not being fit by the model over a third mode.

5.2.3 Canonical Correlation Analysis over a Third Mode

This section defines two generalizations of CCA, based on two different distinguishing features of CCA. CCA generates variates $\mathbf{V}$, which are both uncorrelated with respect to the total covariance of the Y-variables, and with respect to the matrix of error terms, where the error is the total variation less the sums of squares explained by a multivariate regression. That is: $\mathbf{V}'\mathbf{S}_{YY}\mathbf{V} = \mathbf{I}$ and $\mathbf{V}'(\mathbf{S}_{YY} - \mathbf{S}_{XY}'\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY})\mathbf{V} = \mathbf{I} - \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix with the squared canonical correlations. Thus one can generalize CCA to the third mode by defining Y-variates which are uncorrelated with respect to a stable total variation, or by defining Y-variates which are uncorrelated with respect to a stable error term. On the other hand, taking the former approach leads to a method which can be shown to maximize the sums of the squared correlations between the variates. However it requires the awkward assumption that both $\mathbf{S}_{XX}$ and $\mathbf{S}_{YY}$ are stable over the third mode. Taking the latter approach leads to a method which has

the identical form of that of CVA/third (5.3) in Section **5.2.2**, except that the X-variables are no longer restricted to be indicators, and the error is defined as above. This method requires one to assume a stable error matrix and maximizes the sum of a weighted variance explained.

5.2.4 Procrustes Rotation over a Third Mode

Procrustes rotation is traditionally defined as a method that finds an orthogonal transformation **Q** such the point configuration of **YQ** is similar to that of **X** (see Section **2.2.4**). However, Procrustes rotation can also be defined as a method that finds orthogonal X-variates and orthogonal Y-variates such that the sum of the squared covariances between the pairs of corresponding X-variates and Y-variates is maximized (see Section **2.2.4)**. In this section I generalize PR to three-mode data along the lines of the former definition. Putting PR/third in the PARAFAC and Tucker2 frameworks allows me to maximize the sum of the squared covariances over the third mode. Note that one can also define Procrustes rotation as a method that finds pairs of X-variates and Y-variates such that sum (not squared) of the covariances is maximized. However, this definition has the disadvantage that covariances of differing signs would cancel each other out in a summation, so that a strong relationship that changed in sign would receive less weight. The attractive feature about PR/time is that it is not necessary to assume either $\mathbf{S}_{XX}$ or $\mathbf{S}_{YY}$ is constant.

The PR/third model is

$$\frac{1}{n_k - 1}\mathbf{X}_K'\mathbf{Y}_k = \mathbf{W}\mathbf{H}_k\mathbf{V}' + \text{Error}_k$$

where **W** is a $m \times q$ orthonormal matrix of X-variates and **V** is a $n \times r$ orthonormal matrices and $\mathbf{H}_k$ is a $q \times r$ matrix.

To see that PR/third maximizes the sums of the squared covariances, note that $\mathbf{H}_k[i, j]$ is indeed the root of the covariance of the j^{th} Y-variate, $\mathbf{v}_j'\mathbf{y}$, and the i^{th} X-variate, $\mathbf{w}_i'\mathbf{x}$. This is so because for the Tucker2 solution, $\mathbf{H}_k = \mathbf{W}'\mathbf{C}_k\mathbf{V}$, and for the PARAFAC (orth.) solution $\mathbf{H}_k = \text{diag}(\mathbf{W}'\mathbf{C}_k\mathbf{V})$ (Kroonenberg 1983), where $\mathbf{C}_k = \frac{1}{1 - n_k}\mathbf{X}_k'\mathbf{Y}_k$. Thus $\mathbf{H}_k[i, j] = \mathbf{w}_i'\mathbf{X}_k'\mathbf{Y}_k\mathbf{v}_j$. Now by **Proposition 3.2** $\sum_{k=1}^{g} \text{trace}(\mathbf{H}_k^2)$ is maximized.

For comparison's sake, if one wanted to generalize Procrustes rotation in the sense of maximizing the sum of the covariances, then it is simple to show that one obtains **W** and **V** by performing a singular value decomposition on the sum of the $\mathbf{X}_k'\mathbf{Y}_k$ matrices. That is, $\mathbf{W}\mathbf{J}\mathbf{V}' = \sum_{k=1}^{g}\mathbf{X}_k'\mathbf{Y}_k$. Further, say one wanted to generalize PR less in the spirit of finding common variates, but more in the spirit of finding an orthogonal rotation **Q** by which to rotate $\mathbf{Y}_k$

to maximum similarity to $\mathbf{X}_k$. That is, one wished to find $\mathbf{Q}$ that minimizes the following expression:

$$\sum_{k=1}^{g} \text{trace}(\mathbf{QY}_k - \mathbf{X}_k)'(\mathbf{QY}_k - \mathbf{X}_k).$$

Then again it is easy to show that one performs a singular value decomposition on the sum of the $\mathbf{X}_k' \mathbf{Y}_k$ matrices.

5.2.5 Which Transformations to Use

The decision of whether to transform the X-variables by $\mathbf{S}_{XX}^{-1/2}$ and/or to transform the Y-variables by $\mathbf{S}_{YY}^{-1/2}$ is a central one. It determines the metric in which the fit is being maximized and the lack of fit minimized. In the standard case it is equivalent to choosing between CCA, RA and PR. With three-mode data it is equivalent to choosing between CVA/third, CCA/third, RA/third and PR/third. This choice is determined by several factors; whether a set of variables stays constant; whether a covariance matrix of either the X-variables or Y-variables is hypothesized to be stable over the third mode; the hypothesized nature of cause and effect between the X-variables and the Y-variables; and what one wants to emphasize in his analysis.

For example, one is probably not interested in modeling the variation of a set of variables which are constant. Suppose the X-variables are group indicators that are constant over time. To choose PR/third over RA/third or CVA/third would imply one is maximizing the covariance between the two sets. The covariance, however, involves variation of the X-variables, and one is likely not interested in modeling the variation of the X-variables since they are constant over the third mode anyway. One is likely to be more interested in modeling just the variation of the Y-variables. Thus RA/third or CVA/third would be more appropriate. A similar logic would hold if the X-variables were not constant but the covariance of the X-variables, $\mathbf{S}_{XX}$, were stable.

Also, one may think in terms of cause and effect. One may hypothesize the X-variables cause variation in the Y-variables in a regression sense. If the X-variables are not constant over time, one may nevertheless choose to estimate of $\mathbf{S}_{XX}$ and to transform the X-variables by $\mathbf{S}_{XX}^{-1/2}$. This reduces the importance of the variation of the X-variables in the estimation, though in an uneven way; that is, since $\mathbf{S}_{XXk} \neq \mathbf{S}_{XX}$, for $k = 1,\ldots,g$, $\mathbf{S}_{X^*X^*k} \neq \mathbf{I}$ where $\mathbf{X}_k^* = \mathbf{X}_k \mathbf{S}_{XX}^{-1/2}$, though for some occasions $\mathbf{S}_{X^*X^*k}$ may be closer to $\mathbf{I}$ than others.

What one wishes to emphasize may also play a role in choosing the transformation. For example, with grouped data with stable within-group covariances, one could use the within-groups covariance to transform the Y-variables to get scale invariance. However, if the Y-variables are directly comparable one may prefer not to scale the data, but rather perform a RA/third which analyzes the Y-variables in their raw form. As another example, RA/third would usually be preferred over CVA/third if the within-groups covariance matrix was not hypothesized to be stable. In this case one may wish to standardize the variance to make each Y-variable of equal importance by scaling each variable such that they all have an equal total variation.

However, if the within-group covariances were not too dissimilar, one could estimate a common within-groups covariance matrix anyway and transform the Y-variables to achieve a crude scale invariance.

In summary, the methods introduced in this chapter are exploratory. When deciding what kind of analysis to perform the researcher needs to consider the nature of his data and the phenomena, and what he wants to bring out in an analysis.

5.3 HOW TO EVALUATE THE FIT OF THE MODEL

The choices in fitting the model are whether to use the PARAFAC (orth.) or Tucker2 framework, and how many components to include in the model. While there is no systematic approach to fitting the model such as hypothesis tests, there are certain principles and pointers to guide in the decision. Basically, one evaluates the sums of squares lack of fit and the interpretability of the model terms.

It may help to view the modeling frameworks as a hierarchy going from the most complex model that explains the relationship perfectly to the least complex which would have the greatest lack of fit. The most complex model is a separate CCA, RA or PR at each occasion. These will explain the relationship perfectly at each occasion or for each dataset. The next most complex model is the Tucker2. The Tucker2 models two sets of stable variates. A linear relationship which varies in strength is assumed to exist between each variate of the X-set and each variate of the Y-set. Then there is the PARAFAC (orth.) model, which hypothesizes pairs of stable variates with the linear relationship varying in strength. Unlike the Tucker2 the relationship is modeled strictly between members of a pair. Lastly, the simplest model would be to model identical relationships at each occasion, with identical variates and equal strength of relationship.

As an aid to the evaluation and interpretation of a three-mode model one can examine what I shall call the matrix of explained sums of squares. This matrix shows the sums of squares explained of each Y-variable by each Y-component. To determine how much of the sums of squares of the i^{th} Y-variable is explained by the j^{th} Y-component, $\mathbf{v}_j'\mathbf{y}$, first consider the weight of the (orthonormal) component corresponding to that variable. If one is performing RA/third or PR/third this is $\mathbf{v}_j[i]$; if CVA/third or CCA/third this is $\mathbf{v}_j^*[i]$. Square this weight and multiply it by the sum of the squared core elements corresponding to the j^{th} Y-component. The matrix one obtains allows one to see in a simple way what variables are well explained by what components. The interpretation of what the sums of squares means depends on the method. For example, for RA/third, the sums of squares represent the variance of a given variable explained by a given component.

When examining fit to compare the PARAFAC (orth.) and Tucker2, the values of the off-diagonal elements are worth looking at. Small off-diagonal elements suggest the PARAFAC (orth.) model is more appropriate. The Tucker2 will always explain more sums of square, but the PARAFAC (orth.) requires fewer parameters to be estimated, and it has an advantage in interpretability because each X-variate is related only to one Y-variate.

When deciding on the number of components to include one can use a scree plot such as that used in multidimensional scaling. One plots the sums of squares lack of fit against the number of components. The point where the curve levels out suggests the number of components to include in the model.

Other points to consider when evaluating the fit of a three-mode model are the nestedness and scale invariance of the solutions. In Chapter Three I show that the PARAFAC (orth.) solutions are nested. The nestedness property implies that a rank-*f* solution is always a subset of a rank-(*f*+1) solution. This allows one to evaluate the fit contributed by each pair of components when comparing solutions of differing rank. The Tucker2 solutions do not have this property. In Chapter Nine I discuss the topic of scale invariance. While three-mode methods are generally not scale invariant, the Tucker2 and PARAFAC (orth.) models do have approximate scale invariance properties if the fit is very good.

In summary, it should be clear from the discussion that how much complexity the researcher decides to model will be in part subjective.

5.4 AN EXAMPLE

I will go into more details about the models in Section **5.5**, and about the interpretation of the models in Chapter Six. But first I present an example to illustrate what has been laid out so far. The following example is an application of RA/third. The U.S. Geological Survey in cooperation with the University of Virginia's Department of Environmental Sciences performed a study, "Sensitivity of Stream Basins in Shenandoah National Park to Acid Deposition" (Lynch & Dise 1985). This study investigated the acidification of the streams in the said park. The acid presumably came in the form of acid rain from man-made sources such as sulfur bearing pollutants. The purpose of the study was to identify and evaluate geological factors relating to the sensitivity of basins to acid deposition. There were 56 streams located throughout the park, which was divided into 56 corresponding basins. Per recommendation of the authors I discarded three streams which ran parallel to roads, leaving 53 streams in my analysis.

Geological measurements were taken once over the whole basin and are assumed to be unchanging over time. These I designate the X-variables. The geological measurements taken included the classification of the types of underlying bedrock. Since a stream basin frequently contained more than one type of underlying bedrock, the basins are assigned a percentage for each bedrock type. The bedrock classifications included Catoctin, Pedlar, Old Rag, Hampton, Antietam, Swift Run and Weverton. To avoid colinearity among classification variables, I did not assign variables for the Swift Run and Weverton bedrocks, which taken together were still less common than any of the other bedrock types. Other geological measurements were: the percent of the basin above 2400 feet in elevation; an indicator variable for whether the site was on the east or west slope (E/W); an indicator variable for whether 5% or more of the basin was developed; and the drainage density (DD), calculated by dividing the length of the stream by the area of the basin.

For each stream, streamwater measurements were taken on six occasions, always at the same site. These were the Y-variables. The measurements taken were pH; alkalinity, which was

defined as a measure of ability to neutralize or buffer strong acids; conductivity, which is an indirect measure of how much ionization has occurred; temperature; stream discharge; the base cations calcium (Ca^{++}), magnesium (Mg^{++}), potassium (K^{+}), sodium (Na^{+}) and ammonium (NH_4^{+}); and the acidic anions chloride (Cl^{-}), sulfate ($SO_4^{=}$), nitrate (NO_3^{-}), and silica ($SiO_4^{=}$).

The researchers have analyzed their data carefully. The intention of my analysis is not to replicate their work, but to illustrate my method, which will emphasize the investigation of what is changing or not changing over time. The researchers' statistical analysis was largely concentrated in two parts. First, they averaged the data over the six occasions and performed a multiple regression for each chemical measurement against the geological measures. The multiple regression with alkalinity as a response showed that the geological variables explained 0.962 of the variation. Bedrock alone had an R^2 of 0.947. These were the key statistical findings of their study. Geology was also important for predicting amount of base cations and silica, with R^2's ranging from 0.855 to 0.944. Sulfate, nitrate and chlorine concentrations were modeled with R^2's of 0.535, 0.695 and 0.600 respectively.

Next, in an attempt to examine time differences with respect to different bedrocks the researchers subtracted the data for January 1982 from that of May 1982, and then that of June 1982 from September 1981. They do this for only the 28 sites that are classified 75% or more as one type of bedrock, and for only four of the more common bedrocks. Then they compare the mean differences between the sites classified to the various bedrocks. These pairwise comparisons revealed that in warmer months there was in general more alkalinity as well as more base cations, and in particular there was relatively more in certain bedrocks susceptible to carbonic weathering. These increases are explained by more carbonic acid weathering due to higher levels of carbon dioxide in the soil that result from greater microbial and plant activity during these times of year.

My analysis attempts to model the relationship between the geological variables and the stream measurements over time. Since the geological measurements are constant it is appropriate to transform these measurements to uncorrelated variates by multiplying by the Mahalanobis transformation ($\mathbf{S}_{XX}^{-1/2}$). **Table 5.1** shows the error variances of the fourteen water variables at each occasion. (I define error variance as the sum of squared residuals obtained if one performs a separate multivariate regression at each occasion). One sees the error variances of the streamwater measurements are not constant over time. Note for example the changes in the variation in pH and alkalinity. To transform based on an averaged covariance matrix would put the between-group differences in a metric that may not make sense. Hence I choose not to transform the water chemistry variables. (See Section **5.2.5** for a discussion on the appropriate choice of transformations.) However, the responses are

Table 5.1. Error Variance of Responses

Measurement	Aug. 1981	Sept. 1981	Jan. 1982	March 1982	May 1982	June 1982
discharge	0.08	0.01	0.07	0.07	2.19	0.11
conductivity	0.13	0.55	0.05	0.05	0.48	0.10
pH	0.01	0.01	0	0	**4.19**	0.01
temperature	0.99	0.64	0.28	0.28	0.69	0.54
Ca^{++}	0.19	0.60	0.05	0.05	0.25	0.09
Mg^{++}	0.14	0.36	0.05	0.05	0.14	0.09
Na^{+}	0.13	0.24	0.12	0.12	0.28	0.12
K^{+}	0.10	0.14	0.03	0.03	1.90	0.05
alkalinity	**0.22**	**0.61**	0.06	0.06	0.06	0.08
$SO_4^{=}$	0.18	0.35	0.30	0.30	0.59	0.30
Cl^{-}	0.21	0.28	0.24	0.24	2.04	0.16
$SiO_4^{=}$	0.19	0.27	0.10	0.10	0.18	0.19
NO_3^{-}	0.08	0.06	0.30	0.30	2.96	0.12
NH_4^{+}	0.01	0	0	0	4.50	0.01

standardized to have an equal variance over the six occasions. This standardization gives each of the variables equal weight in the analysis while allowing the measurements to vary over time. The choice of transforming the geological variables but not the water measurement variables defines the analysis as an RA/third analysis. The RA/third model finds uncorrelated weighted sums of the geological variables that explain the maximum amount of variation of the streamwater variables over time.

The next step in the analysis is to determine whether the PARAFAC (orth.) or Tucker2 model is appropriate and how many pairs of components are appropriate. I do this by comparing the fit and the interpretability of the competing models. I start by presenting the estimates of the core elements PARAFAC (orth.) estimates in **Table 5.2**. To save space they are not presented as

Table 5.2 PARAFAC (orth.) Core

Date / Component	Aug. 1981	Sept. 1981	Jan. 1982	Mar. 1982	May 1982	June 1982
1st	2.25	2.52	1.83	1.61	2.2	2.06
2nd	0.93	0.95	0.95	0.99	1.03	1.15
3rd	0.90	1.05	0.56	0.76	0.71	0.76
4th	-0.03	0.01	0.01	-1.61	0.02	0.02
5th	-0.13	-0.20	-0.30	-0.83	-0.25	-0.3
6th	-0.23	-0.27	-0.28	-0.35	-0.31	-0.42
7th	0.14	0.06	0.16	0.3	0.28	0.27
8th	0.22	0.28	0.07	-0.1	-0.07	0.09
9th	0.18	0.13	0.12	0.03	0.13	0.12

a diagonal matrix for each occasion, rather as vectors for each occasion. Only the first four components have core elements greater than one in magnitude. Also, as will be discussed later, the first four components have interpretations that lend credence to their selection in the model. Thus the PAFARAC (orth.) model with four pairs of components is a competitive model. The SAS programming codes for the Tucker2 and PARAFAC (orth.) models are found in Appendix Three.

The four-component Tucker2 solution is examined for comparison. The estimates of its core matrices are shown in **Table 5.3**. The sums of squares explained is equal to the sums of squares of the core matrix. The Tucker2 explains 41.8 (81.3%) of the variation, which is only modestly more than the

Table 5.3 Core Matrices for the Tucker2

August 1981			
2.26	-0.02	-0.11	-0.21
-0.63	**0.90**	0.13	0.17
-0.01	0.23	**0.81**	-0.26
0.29	-0.10	0.22	**-0.10**

September 1981			
2.54	-0.13	-0.09	-0.22
-0.10	**0.89**	-0.01	0.20
-0.06	0.31	**0.96**	-0.34
-0.10	-0.14	0.13	**0.01**

January 1982			
1.82	0.03	-0	0.16
-0.02	**0.91**	0.11	0.18
-0.03	0.16	**0.45**	-0.26
0.24	-0.1	0.2	**-0.05**

March 1982			
1.55	-0.35	0.36	**1.25**
0.50	**1.05**	-0.17	-0.10
-0.26	-0.30	**0.90**	-0.01
0.41	0.27	-0.54	**-1.13**

May 1982			
2.19	-0.03	0.13	-0.22
0.42	**0.99**	0.05	0.21
-0.02	0.13	**0.59**	-0.32
-0.03	-0.10	0.20	**-0.08**

June 1982			
2.06	0.06	0.1	-0.12
0.37	**1.12**	-0.08	0.29
0.07	0.01	**0.63**	-0.35
0.04	-0.1	0.24	**-0.08**

rank-four PARAFAC (orth.) solution which explains 39.0 (75.9%). Note that the total variation that can be explained by the relationship by separate multivariate regressions at each occasion is 51.37. Further, the components are similar to those of the PARAFAC (orth.) model, as one can

see by comparing the weights for geological variables in **Table 5.4** for the PARAFAC (orth.) model to those in **Table 5.5** for the Tucker2, and the weights for the water measurements for the PARAFAC (orth.) model in **Table 5.6** to those for the Tucker2 in **Table 5.7**.

Table 5.4 Canonical Variate X-weights for PARAFAC (orth.)

Geological Variable	First Comp.	Second Comp.	Third Comp.	Fourth Comp.
Antietam	-0.084	-0.311	0.682	0.588
Hampton	0.188	-0.938	1.167	1.025
Catoctin	1.229	-0.750	1.327	1.373
Pedlar	0.838	-0.489	1.888	1.475
Old Rag	0.436	-0.238	1.542	1.052
ab2400	-0.445	0.406	-0.586	-0.468
DD	0.124	-0.131	0.345	-0.947
E/W	0.228	-0.396	0.101	0.585
Dev.	0.256	-0.173	-0.261	-0.313

Table 5.5 Canonical Variate X-weights for the Tucker2

Geological Variable	First Comp.	Second Comp.	Third Comp.	Fourth Comp.
Antietam	-0.105	-0.314	0.540	0.695
Hampton	0.148	-0.907	0.858	1.331
Catoctin	1.169	-0.688	0.943	1.668
Pedlar	0.766	-0.433	1.489	1.871
Old Rag	0.382	-0.284	1.299	1.393
ab2400	-0.425	0.342	-0.409	-0.648
DD	0.193	-0.246	0.527	-0.851
E/W	0.193	-0.382	-0.031	0.597
Dev.	0.271	-0.204	-0.189	-0.250

Also, the core elements are similar. The diagonal elements of the Tucker2 core in **Table 5.3** are close to that of the PARAFAC (orth.) of **Table 5.2**. For example, for the first occasion they are 2.26, 0.9, 0.81 and -0.1 versus 2.25, 0.93, 0.9 and -0.03. One can see in **Table 5.3** that the off-diagonal elements of the core matrices for the Tucker2 are generally small, except for those at occasion four. Given the similarity in fit and interpretation, the PARAFAC (orth.) model is preferable to the Tucker2 because it has less terms and is simpler to interpret. For the rest of this section I discuss the PARAFAC (orth.) model and its estimates.

The interpretation of the first four components of the PARAFAC (orth.) model is consistent with the analysis given by the researchers in their study. The first component relates to the level of alkalinity and the concentrations of base cations. First inspect the matrix of sums of squares explained, **Table 5.8**, which indicates how much variance a given geological component explains of a given water measurement variable. Note that the total variance of alkalinity explained by the first component is 4.11. Compare this with the total variance explainable of 4.86 (if separate regression were performed at each occasion), and a total variation of 6.0. Also, there are large variances explained for the base cations Ca^{++}, Mg^{++} and Na^{+}, and the acid silica ($SiO_4^{=}$). These are all products of the same process, the carbonic weathering of minerals high in silica and in base anions. This process tends to increase the alkalinity through the production of carbonic acid, which is a buffer against strong acids. The corresponding geology variate, the X-variate or

Table 5.6 Y-weights for PARAFAC

Measurement	First Comp.	Second Comp.	Third Comp.	Fourth Comp.
discharge	0.036	0.297	-0.271	-0.113
conductivity	**0.369**	-0.293	-0.108	-0.156
pH	0.113	0.058	-0.002	0.593
temperature	0.132	-0.08	0.351	0.048
Ca^{++}	**0.383**	0.018	-0.241	0.002
Mg^{++}	**0.365**	-0.261	-0.43	-0.006
Na^{+}	**0.364**	0.256	0.419	-0.063
K^{+}	-0.158	-0.566	0.182	0.308
alkalinity	**0.394**	0.012	-0.235	0.12
$SO_4^{=}$	0.14	-0.543	0.22	-0.231
Cl^{-}	0.265	-0.06	0.133	-0.065
$SiO_4^{=}$	**0.363**	0.229	0.458	0.003
NO_3^{-}	0.133	0.098	-0.089	0.321
NH_4^{+}	0.035	-0.029	0.034	0.579

Table 5.7. Y-weights for Tucker2

Measure-ment	First Comp.	Second Comp.	Third Comp.	Fourth Comp.
discharge	0.057	0.311	-0.256	0.260
conductivity	**0.373**	-0.274	-0.177	-0.143
pH	0.094	-0.050	0.239	0.513
temperature	0.134	-0.065	0.311	-0.269
Ca^{++}	**0.381**	-0.003	-0.206	0.074
Mg^{++}	**0.363**	-0.267	-0.410	0.087
Na^{+}	**0.367**	0.258	0.379	-0.139
K^{+}	-0.173	-0.592	0.243	0.079
alkalinity	**0.393**	-0.019	-0.193	0.059
$SO_4^{=}$	0.129	-0.508	0.138	-0.144
Cl^{-}	0.268	-0.049	0.116	-0.002
$SiO_4^{=}$	**0.361**	0.234	0.437	-0.125
NO_3^{-}	0.122	0.023	0.101	0.529
NH_4^{+}	0.017	-0.129	0.255	0.467

Table 5.8. Matrix of Sums of Squares Explained by Variable and Component

Measurement	First Comp.	Second Comp.	Third Comp.	Fourth Comp.	Total
discharge	0.03	0.53	0.29	0.03	0.88
conductivity	**3.60**	0.52	0.05	0.06	4.23
pH	0.34	0.02	0	**0.91**	1.27
temperature	0.46	0.04	0.48	0.01	0.99
Ca^{++}	**3.88**	0	0.23	0	4.11
Mg^{++}	**3.53**	0.41	**0.72**	0	4.63
Na^{+}	**3.51**	0.39	**0.68**	0.01	4.49
K^{+}	0.66	**1.93**	0.13	0.25	2.95
alkalinity	**4.11**	0	0.21	0.04	4.36
$SO_4^{=}$	0.52	**1.78**	0.19	0.14	2.66
Cl^{-}	**1.86**	0.02	0.07	0.01	1.96
$SiO_4^{=}$	**3.48**	0.31	**0.81**	0	4.60
NO_3^{-}	0.47	0.06	0.03	0.27	0.83
NH_4^{+}	0.03	0	0	**0.87**	0.90
Total Variation Explained	26.47	6.02	3.88	2.59	39.0

predictor variate, is seen in **Table 5.4**. This variate is uncorrelated with the other X-variates. Its weights are interpreted in the same sense that weights in a regression equation are. First, one sees the weights for type of bedrock are ordered as Catoctin (1.22), Pedlar (0.84), Old Rag (0.44), Hampton (0.18) and Antietam (-0.08). This ordering is the same as that of the regression weights predicted by the researchers and found in their regression equation for alkalinity. The other weights are also consistent with their regression equation for alkalinity.

This first and largest component has several implications. First it affirms the researchers' decision to average the data over time. Analyzing data averaged over time is after all a crude way to get common components. Indeed, it only makes sense if there are common components, otherwise averaging muddles the analysis. Second, it shows the advantage of the multivariate approach over the univariate approach in that it models the responses simultaneously. Alkalinity, the base cations Ca^{++}, Mg^{++}, Na^{+} and silica, which are all modeled individually as univariate responses in the researchers' analysis, are modeled in RA/time in a way that reveals their interrelationship. Further, what the analysis over time reveals is that although this process was roughly stable in strength over time, there were differences. At occasion four the total variance explained is 2.6 (one squares the core element to obtain the variance explained, which is 1.61 from **Table 5.2**). This small value is likely due to heavy rain during that month which would increase the proportion of runoff in the stream as opposed to ground discharge. Runoff has less of the chemicals that are formed by reactions in the soil and bedrock than does ground discharge, weakening the relative strength of these alkalinity and base cations and silica. The occasion where

this process is strongest is September 1981 where the variate explains a variation of 6.36 (2.52 squared), followed by August 1981 with 5.06. These higher values relate to the fact that in warmer weather there is more microbial and plant activity in the soil creating carbon dioxide and initiating carbonic acid weathering.

The second component pair is likewise interpretable as a process predicted and observed by the researchers in their analysis. It is related to precipitation of sulfurous compounds and the ions that result from the ensuing reactions with the bedrock. This is the process that researchers refer to as "acidification". In the table of variance explained, **Table 5.8**, one sees that 1.78 of the variation of sulfate, $SO_4^=$, is explained, and also 1.93 of potassium (K). From **Table 5.4** one sees that there is more sulfate on the western slope and less at higher altitudes. This can be accounted for by the effect of the prevailing winds bearing pollution from the west. Also, higher elevations have more rain and consequently more acid deposition. The high concentration of potassium may be due to greater reactivity with sulfur in bedrock consisting of minerals with high potassium content such as Hampton. What the analysis over time reveals is first, that the researchers' averaging of data over time was again plausible. Second, though the strength of the relationship does seem to be relatively stable, there is a weak increasing trend over time.

The variance explained by the third variate pair is a little less than that explained by the second (**Table 5.2**). The geological variate shows higher weights for bedrock that is granitous, such as Pedlar and Old Rag. The streamwater variate has larger weights for Silica and Na^{++}, which are byproducts of the plagioclastic weathering of granites, which also happen to be low in Mg^{++}. Hence this variate pair is interpretable as indicating plagioclastic weathering. The differences over time in the strength of the variances would need to be interpreted by the researchers for significance.

The fourth component pair is related to runoff effects of rainwater. It is one that would not fair well in an averaging over time. Indeed it was not mentioned by the researchers. One sees that the variance explained at the fourth occasion is 2.59 (-1.61 squared, from **Table 5.2**), but at other occasions it is close to zero. This is likely due to the unusually heavy runoff during March due to heavy rains and perhaps to melting snows. The mean stream flow was 3.2 cubic feet per second per square mile in March, versus 0.2 to 1.3 at the other occasions. The salient weight among the geology variates is drainage density, though altitude, east/west and bedrock type all play a role. In areas with poor drainage, the proportion of runoff will be greater, hence there will be greater runoff effects. Also, basins at higher altitudes receive more rain and have a greater runoff. The streamwater variables explained by this canonical variate are pH and NH_4^+. These both reflect the fact that streamwater with a higher proportion of runoff from rain or snow melts is more like rainwater. Rainwater has a pH of 4.22, lower than the lowest soils in the park which is about pH of 5. Also, ammonia is not found in ground discharge but rather derives strictly from atmospheric precipitation.

In summary, the RA/time analysis confirms the researchers conclusion found using conventional methods. However, it gave a more integrated view of the processes by using both a multivariate approach and modeling over time. It also raised some questions about the relationships over time that the researchers might profitably address.

5.5 SOME FURTHER CONSIDERATIONS

Having presented the CCA, RA and PR/third models and gone over in detail an application, I discuss some issues that shed further light on the nature of the modeling and the problems it solves. These include autocorrelation, the covariances between the Y-variables at different occasions and the (lack of) invariance of solutions to non-singular transformations.

5.5.1 Autocorrelation

Autocorrelation is a common phenomena with measurements made over time. In this section I attempt to answer the question of what effect autocorrelation has on models relating two sets of variables over time. The situation is clearest when one examines RA/third where the X-variables indicate group membership. Then $\frac{1}{n_k - 1}\mathbf{X}'\mathbf{Y}_k$ is a matrix whose i^{th} row is the group means for each Y-variate for the k^{th} occasion. Now it is easy to see that the autocorrelation for $\overline{Y}$ equals the autocorrelation for Y, and that the effect of autocorrelation weakens over time, i.e., $\text{corr}(\mathbf{Y}_k, \mathbf{Y}_{k+m}) = \sigma^m$. Thus a strong autocorrelation tends to make the structure of the data static; that is, little changes over time. Otherwise its presence is observed in the within-groups covariance matrix. The effect of autocorrelation is similar if the X-variables are continuous. In summary, the possible presence of autocorrelation does not require any extra model terms when modeling CCA, RA and PR over time.

5.5.2 Cross Occasion Covariances

When one has longitudinal data one will observe covariances of variables at different occasions. Take, for example, $\mathbf{Y}_r'\mathbf{Y}_s$, $r \neq s$. By modeling only $\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k$ or its transformations, it seems that one is ignoring information by not modeling these cross-occasion covariances. However, for certain important situations this is not so. Consider when one has constant **X** over time and is modeling RA/third. Then the sums of squares regression of the Y-variables explained by the X-variables at occasions r and s is $\mathbf{Y}_r'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_s$. But this is just the product of $(n_k - 1)\mathbf{S}_{xx}^{-½}\mathbf{X}'\mathbf{Y}_r$ and $(n_k - 1)\mathbf{S}_{xx}^{-½}\mathbf{X}'\mathbf{Y}_s$, two matrices which are already modeled in RA/third (with a different weighting). Further, $\mathbf{Y}_r'\mathbf{Y}_s - \mathbf{Y}_r'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_s$ is the sum of squares error, which is not modeled in RA/third. In this sense the covariance between variables at different occasions offers no new information.

An analogous argument can be made for CVA/third with longitudinal data. However, with data where **X** is not constant the situation is not clear. Later chapters approach the issue of modeling cross occasion covariances. Chapter Seven presents least squares methods that attempt to model some of these cross-occasion terms. Chapter Eight presents maximum likelihood methods that model the error terms.

5.5.3 Invariance of the Rank of the Solution to Non-Singular Transformations

In Section **5.2.5** I discussed which transformations to apply. In this section I add to that a brief discussion on the invariance of the solutions to the choice of transformation. It is known that the rank of a matrix is invariant to non-singular transformations, and that the column space of a matrix is invariant to non-singular transformations of the row space, and vice versa. Since CCA, RA and PR are based on the SVD of $\mathbf{S}_{XY}$ with the appropriate transformations (see Section **2.2**), one can draw two implications; first, that for a given dataset the solutions for CCA, RA and PR all are of the same rank; second, the X-variates for CCA and RA span the same space, as do the Y-variates for RA and PR.

These features hold true for the three-mode extensions of CCA, RA and PR if they are put in the framework of the Tucker2, but not if they are in the PARAFAC (orth.) framework. To see this, consider a series of *g* matrices, $\mathbf{Q}_1,\ldots,\mathbf{Q}_g$, which are modeled exactly by the Tucker2. That is: $\mathbf{Q}_1 = \mathbf{R}\mathbf{S}_1\mathbf{T}'$, $\mathbf{Q}_2 = \mathbf{R}\mathbf{S}_2\mathbf{T}'$, with **R** and **T** orthonormal, etc. Now consider an arbitrary, non-singular transformation **A** for the row space. Perform a SVD on **A**, $\mathbf{A} = \mathbf{M}\mathbf{N}\mathbf{P}'$. Then $\mathbf{A}\mathbf{Q}_k = \mathbf{M}\mathbf{N}\mathbf{P}'\mathbf{R}\mathbf{S}_k\mathbf{T}'$, for $k = 1,\ldots,g$. Now one can perform a SVD on $\mathbf{M}\mathbf{N}\mathbf{P}'\mathbf{R}$ to get $\mathbf{M}\mathbf{N}\mathbf{P}'\mathbf{R} = \mathbf{D}\mathbf{E}\mathbf{F}'$ and consequently $\mathbf{A}\mathbf{Q}_k = \mathbf{D}\mathbf{E}\mathbf{F}'\mathbf{S}_k\mathbf{T}'$, or $\mathbf{A}\mathbf{Q}_k = \mathbf{D}\mathbf{S}_k^*\mathbf{T}'$, where $\mathbf{S}_k^* = \mathbf{E}\mathbf{F}'\mathbf{S}_k$. One sees that one has a Tucker2 solution with the same column space **T**. On the other hand, if $\mathbf{S}_k$ is now restricted to be diagonal, one cannot generally find $\mathbf{S}_k^*$ that is diagonal to yield a PARAFAC (orth.) solution.

The main implication of this to modeling is that if one is uncertain about which transformation to apply to the data, then the Tucker2 is a safer model than the PARAFAC (orth.). Also, if the Tucker2 has a lower rank model that fits better than a higher rank PARAFAC (orth.) model, one might consider another transformation.

5.5.4 Concluding Comments

First note that categorical data are handled the same way as continuous data. Hence correspondence analysis is generalized to the third mode. See Section **2.2.2** for the interpretation of categorical data with CCA type analyses.

In summary, the methods of this chapter are flexible and exploratory. They require some subjective choices as to the choice of transformation, choosing the Tucker2 versus the PARAFAC (orth.) models and determining the number of components in the solution. Ultimately, as seen in the example in Section **5.4**, one may choose a model which does not explain all of what is going on in the data, but explains some of what is going on; that is, finds some structure to the data over the third mode.