CHAPTER TWO

BACKGROUND

In this dissertation I develop methods for certain multivariate applications when one has three-mode data; that is, data taken over multiple occasions or datasets. In this chapter I present background material that is central to the development of these methods. Section **2.1** details the singular value decomposition (SVD). The SVD is important to this dissertation for several reasons. First, it unifies the methods for relating two sets of variables, as they can be put into the framework of a SVD. Second, the three-mode PCA models are generalizations of the SVD. Lastly, the SVD is the basis for the biplot, a fundamental graphical technique. Section **2.2** describes canonical correlation analysis, canonical variate analysis, redundancy analysis and Procrustes rotation. These are kindred methods for relating two sets of variables which I will generalize to longitudinal and multiple group data. Section **2.3** delves into three-mode principal components, which is one of the approaches I take to generalizing the aforementioned multivariate methods. An advantage of three-mode PCA is that it lends itself readily to graphical displays. Lastly, Section **2.4** discusses the Campbell and Tomenson model for canonical variate analysis for data from multiple datasets. To the best of my knowledge it is the only model that accomplishes something similar to that which the models in this dissertation will accomplish. In particular, the model I develop in Chapter Eight can be viewed as an extension of Campbell and Tomenson's model.

2.1 THE SINGULAR VALUE DECOMPOSITION

An $m \times n$ matrix **X** can be decomposed as follows (Kshirsagar 1972):

$$\mathbf{X} = \mathbf{\Pi\Sigma\Omega}'$$

where $\mathbf{\Pi}$ is an $m \times m$ orthogonal matrix, $\mathbf{\Omega}$ is an $n \times n$ orthogonal matrix, and $\mathbf{\Sigma}$ is an $m \times n$ matrix with elements σ_{ij}, where the singular values are $\sigma_{\mathrm{jj}} = \sigma_{\mathrm{j}}$, and $\sigma_{\mathrm{ij}} = 0$ for $\mathrm{i} \neq \mathrm{j}$. Such a decomposition is called the singular value decomposition (SVD). An equivalent form of the SVD (assume $m \geq n$) specifies that $\mathbf{\Pi}$ be an $m \times n$ matrix with orthonormal columns and $\mathbf{\Sigma}$ an $n \times n$ diagonal matrix with $\sigma_{\mathrm{jj}} = \sigma_{\mathrm{j}}$. Which form is being referred to will be clear from the context.

The SVD has the optimality property (Eckart and Young 1936) that it yields the best low rank approximation to a matrix in a least squares sense. If **X** has rank *r* and one wants a rank *s* approximation to **X**, $s < r$, then the optimal approximation is $\hat{\mathbf{X}} = \mathbf{\Pi}_{\mathrm{s}}\mathbf{\Sigma}_{\mathrm{s}}\mathbf{\Omega}'_{\mathrm{s}}$, where $\mathbf{\Sigma}_{\mathrm{s}}$ is an $s \times s$ diagonal matrix whose diagonal elements are the *s* largest singular values of **X**, and $\mathbf{\Pi}_{\mathrm{s}}$ and $\mathbf{\Omega}_{\mathrm{s}}$ contain the columns of $\mathbf{\Pi}$ and $\mathbf{\Omega}$ corresponding to those *s* singular values. It is this optimality property that allows the SVD to be used as the basis for biplots. For a discussion of some other uses of the SVD in statistics see Good (1969).

If **X** is a symmetric matrix then the singular value decomposition is equivalent to a spectral decomposition. $\mathbf{\Pi}_{\mathrm{s}} = \mathbf{\Omega}_{\mathrm{s}} =$ the matrix of eigenvectors of **X**, while $\mathbf{\Sigma}_{\mathrm{s}}$ is a diagonal matrix of eigenvalues of **X**.

2.2 CANONICAL CORRELATION AND RELATED MODELS

Canonical correlation analysis (CCA) is the most widely known of the multivariate methods that relate two sets of variables. CCA relates *p* X-variables to *q* Y-variables. The assignment of variables into the X-set and the Y-set is always performed before the analysis and is based on the nature and purpose of the study. For example, a medical researcher may want to relate lifestyle variables, X-variables, such as daily caloric intake and exercise, with cardiovascular variables, Y-variables, such as cholesterol level and blood pressure. Canonical correlation analysis (CCA), redundancy analysis (RA) and Procrustes rotation (PR) all require this a priori division of the variables.

As a preliminary, define the covariance matrix **S**:

$$\mathbf{S} = \left(\frac{1}{n-1}\right)\begin{bmatrix}\mathbf{Y}'\mathbf{Y} & \mathbf{Y}'\mathbf{X}\\ \mathbf{X}'\mathbf{Y} & \mathbf{X}'\mathbf{X}\end{bmatrix} = \begin{bmatrix}\mathbf{S}_{\mathrm{YY}} & \mathbf{S}_{\mathrm{YX}}\\ \mathbf{S}_{\mathrm{XY}} & \mathbf{S}_{\mathrm{XX}}\end{bmatrix},$$

where **X** is an $n \times p$ matrix of measurements of *p* variables on *n* units, and **Y** is an $n \times q$ matrix of measurements of *q* variables on *n* units. I shall assume throughout the discussion that **X** and **Y** are centered by variable. If the variables are also scaled to unit variance $\mathbf{S}$ is a correlation matrix.

2.2.1 Canonical Correlation

Hotelling (1935) proposed canonical correlation analysis as a model to relate two sets of variables measured on the same units. He derived linear combinations of the X-variables and linear combinations the Y-variables that were maximally correlated, subject to the constraints that each derived variate was uncorrelated with the other variates in its set and that each variate had a variance of one. Denote the vector of X-variables by **x**, and the vector of Y-variables by **y**. CCA finds linear compounds of **x** and **y**, a_i and b_i:

$$a_i = \mathbf{w}_i'\mathbf{x}, \qquad b_i = \mathbf{v}_i'\mathbf{y},$$

choosing $\mathbf{w}_i$ and $\mathbf{v}_i$ to maximize the correlation between a_i and b_i, subject to the constraints:

$$\mathbf{w}_i'\mathbf{S}_{XX}\mathbf{w}_j = 0, \qquad \mathbf{v}_i'\mathbf{S}_{YY}\mathbf{v}_j = 0 \qquad \forall\ i \neq j$$

and

$$\mathbf{w}_i'\mathbf{S}_{XX}\mathbf{w}_i = 1, \qquad \mathbf{v}_i'\mathbf{S}_{YY}\mathbf{v}_i = 1.$$

The canonical coefficients, $\mathbf{w}_i$ and $\mathbf{v}_i$, can be obtained by a spectral analysis. The $\mathbf{w}_i$ are eigenvectors of the matrix $\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}$ and the $\mathbf{v}_i$ are eigenvectors of $\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$. The eigenvalues of these two matrices are equal, and they are the canonical correlations between the pairs of canonical variates, $\mathbf{w}_i$ and $\mathbf{v}_i$.

Let **W** represent the matrix whose columns consist of $\mathbf{w}_i$ and **V** the matrix whose columns consist of $\mathbf{v}_i$. For a positive definite matrix **M** define the unique symmetric positive definite square root matrix $\mathbf{M}^{1/2}$ such that $\mathbf{M} = \mathbf{M}^{1/2}\mathbf{M}^{1/2}$. Clearly, $\mathbf{M}^{1/2} = \mathbf{L}\boldsymbol{\Lambda}^{1/2}\mathbf{L}'$, where $\mathbf{M} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}'$ is the spectral decomposition of **M**. Then it is also possible to get the canonical coefficients from a singular value decomposition of the matrix $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$ (Gittins 1985) as follows:

$$\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2} = \mathbf{W}^*\mathbf{E}\mathbf{V}^{*\prime},$$

where $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}\mathbf{W}$, $\mathbf{V}^* = \mathbf{S}_{YY}^{1/2}\mathbf{V}$, **E** is a diagonal matrix with the canonical correlations as its elements.

Canonical correlation has the appealing property of biorthogonality. Biorthogonality is the property that each canonical variate in the X-domain is uncorrelated with the canonical variates in the Y-domain except the corresponding Y-variate. This is equivalent to requiring that **W** and **V** diagonalize $\mathbf{S}_{XY}$, that is;

$$\mathbf{W}'\mathbf{S}_{XY}\mathbf{V} = \mathbf{E}.$$

where **E** is a diagonal matrix. Biorthogonality implies that the relationship between the X-variables and Y-variables can be partitioned by the pairs of canonical variates, enhancing the interpretability of the analysis.

As an aid to interpreting the canonical variates, researchers often examine structure coefficients (Meredith 1964). The structure coefficients for the X-variables are the correlations between the X-variables and the canonical variates for the X-domain. The matrix of these terms is $\mathbf{S}_{XX}\mathbf{W}$. Similarly, the structure coefficients for the Y-variables are the correlations between the Y-variables and the canonical variates of the Y-domain, and the matrix of these terms is $\mathbf{S}_{YY}\mathbf{V}$.

2.2.2 Canonical Variate Analysis

Canonical variate analysis (CVA) is a widely used method for analyzing group structure in multivariate data. It is mathematically equivalent to a one-way multivariate analysis of variance (MANOVA) and often goes by the name canonical discriminant analysis. CVA can be interpreted as a special case of canonical correlation analysis where one set of variables consists of group indicators (Gittins 1985). This formulation of CVA as a canonical correlation analysis will be exploited in Chapters Five and Six. A geometrical formulation of CVA given by Campbell and Atchley (1981) will be exploited in Chapter Eight.

To start, however, it is useful to review the traditional formulation of CVA. Krzanowski (1988, page 291) summarizes that the objective of CVA is to, "provide a low-dimensional representation of the data that highlights as accurately as possible the true differences existing between the *m* subsets of points in the full configuration". One finds a weighted sum of the variables whose between-groups variation is maximized with respect to its within-groups variation. That is, find the $p \times 1$ vector $\mathbf{v}_1$ maximizing $\frac{\mathbf{v}_1'\mathbf{B}\mathbf{v}_1}{\mathbf{v}_1'\mathbf{C}\mathbf{v}_1}$, subject to $\mathbf{v}_1'\mathbf{C}\mathbf{v}_1 = 1$, where $\mathbf{B} = \sum_{g=1}^{m} n_g(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})(\bar{\mathbf{x}}_g - \bar{\mathbf{x}})'$, $\mathbf{C} = \sum_{g=1}^{m}\sum_{i=1}^{n_g}(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)(\mathbf{x}_{gi} - \bar{\mathbf{x}}_g)'$, $\bar{\mathbf{x}}_g$ is vector of sample means for the g^{th} group and $\bar{\mathbf{x}}$ is the vector means for the entire dataset. **C** is referred to as the within-groups covariance matrix and **B** as the between-groups covariance matrix. Find further $\mathbf{v}_i$, $i = 2, \ldots, r$, which maximize $\frac{\mathbf{v}_i\text{\`{}}\mathbf{B}\mathbf{v}_i}{\mathbf{v}_i\text{\`{}}\mathbf{C}\mathbf{v}_i}$ subject to $\mathbf{V}'\mathbf{C}\mathbf{V} = \mathbf{I}$, where $\mathbf{v}_i$ is the i^{th} canonical variate, and $\mathbf{V}$ is a $p \times r$ matrix whose i^{th} column is $\mathbf{v}_i$. Note that $r \leq \min(m, p)$ and that $\mathbf{B} + \mathbf{C} = \mathbf{S}_{XX}$.

The matrix of canonical variates is obtained by finding the eigenvectors of $\mathbf{C}^{-1}\mathbf{B}$. However, it is easier to take the eigenvectors of $\mathbf{C}^{-\frac{1}{2}}\mathbf{B}\mathbf{C}^{-\frac{1}{2}}$. Denote the matrix of eigenvectors of $\mathbf{C}^{-\frac{1}{2}}\mathbf{B}\mathbf{C}^{-\frac{1}{2}}$ by **U**. Then $\mathbf{V} = \mathbf{C}^{-\frac{1}{2}}\mathbf{U}$. There are two hypothesis tests of interest. The first is the test that the vectors of group means are equal. The second is the test of dimensionality; i.e., how many canonical variates are statistically significant. For details on these tests see Kshirsagar (1972).

To put CVA in the framework of canonical correlation analysis, create m - 1 binary variables, $x_1, \ldots, x_{m\text{-}1}$. If a subject belongs to the s^{th} group, then $x_s = 1$, otherwise $x_s = 0$. Now one can obtain canonical variates as described in Section **2.2.1**.

Campbell and Atchley (1981) give a geometrical formulation of CVA. They model the group means to lie in the subspace defined by the canonical variates, in particular, by $\mathbf{\Sigma V}$, as seen below:

$$\boldsymbol{\mu}_g = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}\mathbf{V}\mathbf{e}_g, \quad (2.1)$$

for $g = 1, \ldots, m$, where *m* is the number of groups, *p* is number of variables, $\boldsymbol{\mu}_g$ is a $p \times 1$ vector of means for the g^{th} group, $\boldsymbol{\mu}_0$ is a $p \times 1$ vector of overall means, $\boldsymbol{\Sigma}$ is a $p \times p$ within-groups covariance matrix, and $\mathbf{e}_g$ is an $f \times 1$ vector of scores of group means on each canonical variate, i.e., $\mathbf{e}_g = \mathbf{X}'\mathbf{V}$. When one assumes multivariate normality and estimates by maximum likelihood, the solution to **V** is the same as given in Section **2.2.1**.

Campbell and Atchley argue that one can view CVA as a principal components analysis performed on the group means in the space obtained by transforming the variables by the Mahalanobis transformation; that is, $\mathbf{x}^* = \mathbf{S}_{XX}^{-1/2}\mathbf{x}$. In this space Euclidean distance equals Mahalanobis distance, where the Mahalanobis distance between two group means $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ is defined as $(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$. Further, in this space $\mathbf{S}_{X^*X^*} = \mathbf{I}$. The principal components of the group means in this transformed space correspond to the canonical variates. To illustrate these points, consider **Figure 2.1**, which shows a scatterplot of the data from two groups for two variables. Compare **Figure 2.1** to **Figure 2.2**, which shows a scatterplot for the same data in the space of the transformed variables.

To see that this view of CVA leads to (2.1) consider that the positions of the group means in the transformed space is $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_g$, for $g = 1, \ldots, m$. A principal components analysis of the positions of these group means leads to the spectral decomposition of

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_g - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_g - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1/2} = \tfrac{n}{g-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{B}\boldsymbol{\Sigma}^{-1/2}.$$

The matrix principal components is **U**, where $\mathbf{U} = \boldsymbol{\Sigma}^{1/2}\mathbf{V}$. Now reverse the transformation by multiplying **U** by $\boldsymbol{\Sigma}^{1/2}$, and one has the group means in the space defined by $\boldsymbol{\Sigma}\mathbf{V}$.

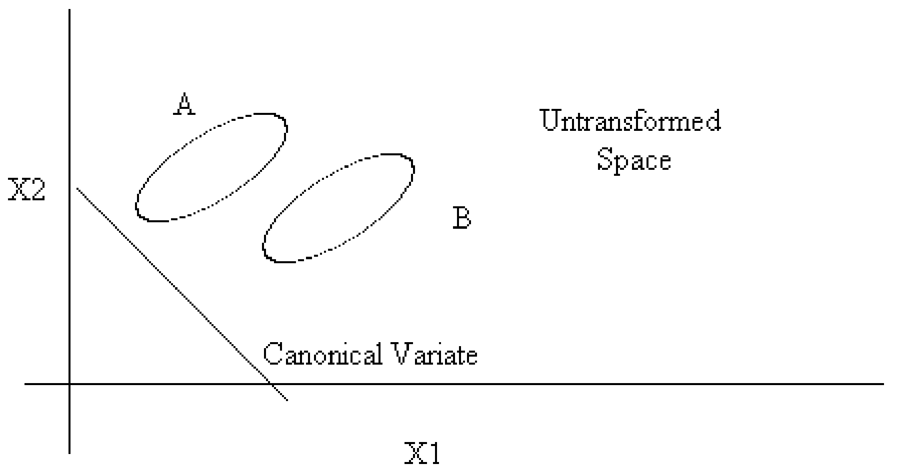

Figure 2.1 Scatterplots in the Untransformed Space

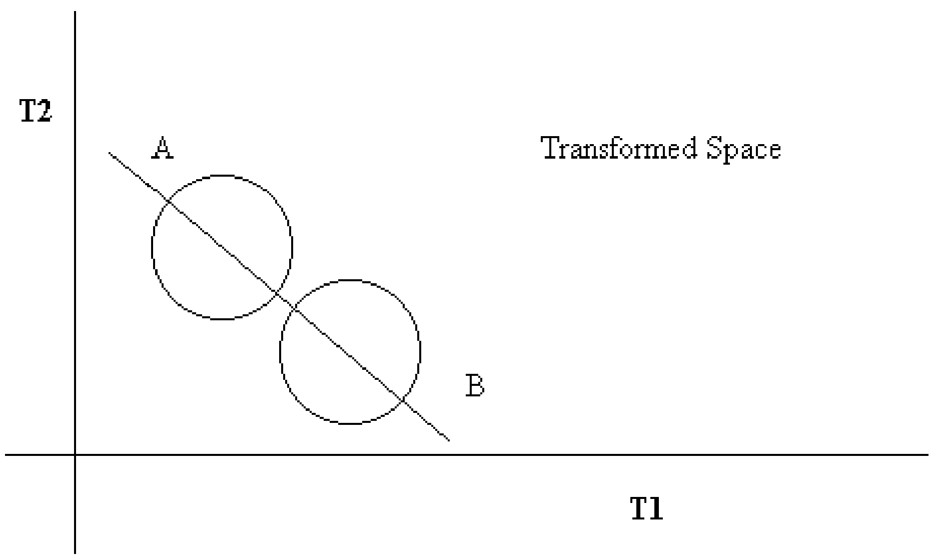

Figure 2.2 Scatterplots in the Transformed Space

2.2.3 Correspondence Analysis

Another method which can be viewed as a special case of CCA is correspondence analysis. Correspondence analysis can be interpreted as a canonical correlation analysis on data where both sets of variables are group indicators; that is, the data are in the form of a two-way contingency table. Correspondence analysis is an alternative to loglinear models that lends itself well to graphical displays (Greenacre 1984). Some of the methods developed for CCA will be applicable to correspondence analysis, providing a method for generalizing correspondence analysis to longitudinal or multiple group data.

2.2.4 Redundancy Analysis

CCA sometimes finds variates that are correlated but not of practical interest to the researcher because they explain little variance. Redundancy analysis (RA) was devised by Van den Wollenberg (1978) as an alternative to CCA that avoids this problem. RA derives uncorrelated compounds of the X-variables, called redundancy variates, which maximize the variance explained of the Y-variables. Rao (1964) had earlier proposed the same method.

The weights for the redundancy variates, $\mathbf{w}_i$, are determined by the following eigenvalue equation:

$$(\mathbf{S}_{XY}\mathbf{S}_{YX} - \lambda_i \mathbf{S}_{XX})\mathbf{w}_i = 0,$$

where λ_i is the variance explained by the i^{th} variate.

In Van den Wollenberg's original development of RA, only redundancy variates for the X-variables are found. Johansson (1981) extended redundancy analysis by deriving variates in the Y-set that correspond to the redundancy variates in the X-set. Linear combinations of the Y-variables, $\mathbf{v}_i'\mathbf{y}$, are extracted such that the absolute value of $\mathbf{w}_i'\mathbf{S}_{XY}\mathbf{v}_i$ is maximized subject to the constraints $\mathbf{v}_i'\mathbf{v}_i = 1$ and $\mathbf{v}_i'\mathbf{v}_j = 0$ for $i \neq j$. The resulting solution is:

$$\mathbf{v}_i = \lambda_i^{-1/2}\mathbf{S}_{YX}\mathbf{w}_i$$

where $\lambda_i = \mathbf{w}_i'\mathbf{S}_{XY}\mathbf{S}_{XY}'\mathbf{w}_i$ (Tyler 1982). An alternate solution is to perform a singular value decomposition on $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}$,

$$\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY} = \mathbf{W}^*\mathbf{E}\mathbf{V}' \qquad (2.2)$$

where $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}\mathbf{W}$, $\mathbf{V}$ is the matrix of variates for the Y-variables and $\mathbf{E}$ is the diagonal matrix whose elements are the square roots of λ_i.

RA has the property of biorthogonality, which was defined in Section **2.2.1** for CCA. However, Tyler (1982) showed that RA has an even stronger property. The pairs of redundancy variates additively partition the total variation of the Y-variables that is explained by the X-variables. In other words, all of the variance explained by a redundancy variate $\mathbf{w}_i'\mathbf{x}$ is associated with one vector in the Y-set, $\mathbf{v}_i'\mathbf{y}$. This property also holds for CCA. However, the redundancy variates partition the variance in an optimal way, as they successively maximize the variance explained in the Y-variables.

Redundancy analysis (RA) has analogues to canonical variate analysis. When the X-variables are dummy variables indicating group membership, as in canonical variate analysis, RA yields a procedure similar to canonical variate analysis. However, it differs in that RA determines pairs of variates that maximize the between-group variance, whereas CCA determines pairs of variates that maximize the ratio of the between-group variance to the within-group variance. One can also look at RA with this kind of data as a principal components analysis on the group means in the untransformed space, with the principal components corresponding to the redundancy variates.

2.2.5 Procrustes Rotation

Procrustes analysis (Gower 1975) is an analysis where two sets of variables measured on the same units are translated, dilated and rotated such that the point configurations are as similar as possible in a least squares sense. The interest in this section is in the rotation part of the Procrustes analysis, which shall be referred to simply as Procrustes rotation (PR). PR is closely related to CCA and RA. In later chapters it will be seen to be more suitable than CCA or RA for particular kinds of data when extended to a three-mode model.

A Procrustes analysis usually starts with a translation of the data. But that is obviated in this discussion by the centering of both sets of variables. Since PR can be performed independently of the dilation, as will be shown shortly, I start with the rotation. If **X** and **Y** are mean centered, Procrustes rotation finds the orthogonal matrix **Q** such that m is minimized, where: $m = \sum_i \sum_j \left(x_{ij} - y^*_{ij} \right)^2$ and $\mathbf{Y}^* = \left[y^*_{ij} \right] = \mathbf{YQ}$. **Q** can be shown to be: $\mathbf{Q} = \mathbf{VW}'$, where one derives **W** and **V** by performing a SVD on $\mathbf{S}_{XY}$, so that

$$\mathbf{S}_{XY} = \mathbf{WEV}'.$$

Now one sees that if one performs a dilation on either **x** or **y** by multiplying by a scalar c, one obtains the same **W** and **V** as c is factored into the matrix of singular values, **E**.

Like CCA and RA, Procrustes rotation can be viewed as a method which finds pairs of variates that relate two sets of variables. These pairs of variates are orthogonal and maximize the covariance. To see this, note that the SVD of $\mathbf{S}_{XY}$ is equivalent to successively finding pairs of $\mathbf{w}_i$ and $\mathbf{v}_i$ such that $\mathbf{w}_i'\mathbf{S}_{XY}\mathbf{v}_i$ is maximized; but $\mathbf{w}_i'\mathbf{S}_{XY}\mathbf{v}_i$ is just the covariance of $\mathbf{w}_i'\mathbf{x}$ and $\mathbf{v}_i'\mathbf{y}$. Thus **E** is a diagonal matrix with the covariances of $\mathbf{w}_i'\mathbf{x}$ and $\mathbf{v}_i'\mathbf{y}$ in the i^{th} diagonal position. For the rest of the dissertation when Procrustes rotation is referred to it shall be in the sense of finding orthogonal variate pairs such that the covariance of the pairs is maximized.

At this point it is worth emphasizing the unity of form of CCA, RA and PR. All are derived from a SVD of a transformation of $\mathbf{S}_{XY}$. In CCA both the X-variables and Y-variables are transformed by their respective Mahalanobis transformations and $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$ is decomposed. In RA only the X-variables are transformed and $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}$ is decomposed. In PR neither the X-variables nor the Y-variables are transformed before decomposition. This unity of form will be exploited in the generalizations of CCA, RA and PR to three-mode data in Chapter Four.

2.3 THREE-MODE PRINCIPAL COMPONENT ANALYSIS

Three-mode PCA is a method which I will use to generalize CCA, RA and PR to three-mode data. The discussion on three-mode PCA in this section is derived from Kroonenberg (1983). A mode is an index for the data. Traditional PCA has two modes, typically the subject and variable modes. Each measurement is indexed by subject and variable (the units shall be referred to as subjects, as is the convention in three-mode PCA). A third mode is a third index for the data, such as conditions or occasions. An example of a three-mode observation would be to measure a subject on a variable on a given occasion (or under a given condition).

2.3.1 The Tucker3 Model

Tucker (1966) generalized the SVD to three-mode data with his Tucker3 and Tucker2 models. The more general of the two models is the Tucker3, which decomposes a three-way array into a set of orthonormal vectors (components) for each mode and a three-dimensional core matrix ("box") of values relating these components. The components are a weighted sum of subjects, variables or occasions, and are interpreted the same way as are principal components, that is, as summaries of variation. The core box is analogous to the matrix of eigenvalues in a SVD, although the core box is generally not diagonal. Let the modes be indexed by i, j, k. Typically, x_{ijk} could stand for the value of variable j on condition (or time) k for subject i; hence $\underline{\mathbf{X}} = [x_{ijk}]$. Let *n* indicate the number of measurements in the i-mode, i.e. the number of subjects; let *m* indicate the number of measurements in the j-mode, i.e. the number of variables; and let *l* indicate the number measurements in the k-mode, i.e. the number of conditions or occasions. Further, let *s* indicate the number of components for the i-mode, *t* the number of components for the j-mode, and *u* the number of components for the k-mode. Let **G** denote the $n \times s$ matrix of components for the i-mode, **H** the $m \times t$ matrix of components for the j-mode, **E** the $l \times u$ matrix of components for the k-mode, and $\underline{\mathbf{C}}$ the $s \times t \times u$ core box. Without loss of generality the matrices **G**, **H** and **E** are specified to be columnwise orthonormal to identify the solution.

The Tucker3 model is expressed in terms of a single observation as follows:

$$x_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} .$$

To express the three-mode decomposition in matrix form it is necessary to reformulate $\underline{\mathbf{X}}$ and $\underline{\mathbf{C}}$. Let **X** be the $n \times lm$ matrix formed by laying out the l $n \times m$ subject$\times$variable data matrices side by side. Let **C** be the $s \times ut$ matrix formed by laying out the u $s \times t$ core matrices side by side. Then

$$\mathbf{X} = \mathbf{GC}(\mathbf{H}' \otimes \mathbf{E}'),$$

where $\otimes$ is the Kronecker product (see Appendix One).

2.3.2 Special Cases of Three-Mode Principal Components

A special case of the Tucker3 model is the Tucker2 model (Kroonenberg 1983), which restricts **E** to be the identity matrix. One can specify the Tucker2 model in terms of matrices as follows:

$$\mathbf{X}_k = \mathbf{GC}_k\mathbf{H}', \qquad k = 1, \ldots, l.$$

If one makes the further restriction that the $\mathbf{C}_k$ matrices be diagonal, one has the PARAFAC (PARAllel FACtor analysis, Harshman 1970) or Candecomp (CANonical DECOMPosition, Carroll and Chang 1970) model, henceforth referred to simply as PARAFAC. The PARAFAC model does not assume orthonormal columns for the components matrices **G** and **H**, and, in contrast to the Tucker2 and Tucker3 models, to require such results in a loss in generality. However, the PARAFAC model with the restriction of orthonormality on the two sets of components will play an important role in future chapters. It shall be henceforth referred to as the PARAFAC (orth.) model.

Three-mode PCA models were envisaged for the situation where one wants to perform a PCA or factor analysis on data that was not only multivariate but had measurements taken over multiple occasions or conditions. However, they can also model data from multiple datasets or groups. One way to accomplish this is to model the crossproduct matrices, typically the covariance matrices, for each dataset (Kiers 1991). This is the approach taken in Chapter Three on common principal components. Modeling covariances implies symmetry between the **G** and **H** components. If the PARAFAC model is restrained so that $\mathbf{G} = \mathbf{H}$ one has the INDSCAL (INDividual SCALing, Carroll and Chang 1970) model.

2.4 THE CAMPBELL AND TOMENSON MODEL

Campbell and Tomenson (1983) hypothesize common canonical variates over multiple datasets. Their model is both a competitor of some of the models to be presented later and a starting point for them. In Campbell and Tomenson's formulation the common canonical variates build a reduced space in which the group means for each dataset are located. They present a hierarchy of models from most general to most specific. Briefly outlined, the most general model is that the canonical variates differ in each dataset. The next most general model is that the canonical variates are common in each dataset, but the location of the group means on them differ. The most specific model is that the variates are the same and the group means have common coordinates on them. A more detailed statement of these models follows.

The model in general form is:

$$\boldsymbol{\mu}_{gk} = \boldsymbol{\mu}_{0k} + \boldsymbol{\Sigma}\mathbf{V}_k \mathbf{e}_{gk}$$

where p indicates the number of variables, $\boldsymbol{\mu}_{gk}$ is the $p \times 1$ vector of means for the g^{th} group in the k^{th} dataset, $\boldsymbol{\mu}_{0k}$ is the $p \times 1$ vector of overall group means for the k^{th} dataset, $\boldsymbol{\Sigma}$ is the $p \times p$ within-group covariance matrix assumed common over group and dataset, $\mathbf{V}_k$ is the $p \times r$ matrix whose r columns are the canonical variates for dataset t, and $\mathbf{e}_{gk}$ is an $r \times 1$ vector specifying the coordinates on the r canonical variates for the g^{th} group mean in the k^{th} dataset. The three models described above are:

1. $\mathbf{V}_i \neq \mathbf{V}_j$ and $\mathbf{e}_{gi} \neq \mathbf{e}_{gj}$ for $i \neq j$; the canonical variates differ over dataset.
2. $\mathbf{V}_i = \mathbf{V}_j$ but $\mathbf{e}_{gi} \neq \mathbf{e}_{gj}$ for $i \neq j$; the canonical variates are common to each dataset, but the coordinates of the group means on the variates differ.
3. $\mathbf{V}_i = \mathbf{V}_j$ and $\mathbf{e}_{gi} = \mathbf{e}_{gj}$, but $\boldsymbol{\mu}_{0i} \neq \boldsymbol{\mu}_{0j}$ for $i \neq j$; the canonical variates are common to each dataset, the coordinates of the group means on the variates are the same, but the overall center of the means is different.

Campbell and Tomenson assume that the data are normally distributed, then derive maximum likelihood estimates of the parameters and hypothesis tests based on likelihood ratios.

2.5 SUMMARY

This chapter provides the starting point for the rest of the dissertation. It presents those multivariate methods I want to extend to longitudinal data, which are CCA, CVA, RA and PR. It presents two models that I will build on to achieve this: three-mode PCA, which is a flexible least squares method which is suitable for graphical displays such as joint plots: and Campbell and Tomenson's model, which is based on maximum likelihood methods.