Order Number 9211092

Case deletion for multilinear models

Arboleda, Maria Liza De Ungria, Ph.D.

The Ohio State University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Case Deletion for Multilinear Models

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

M. Liza U. Arboleda, B.S.,M.S.

* * * * *

The Ohio State University

1991

Dissertation Committee:

Professor Sue E. Leurgans

Professor Angela M. Dean

Professor Robert T. Ross

Approved by

Sue E Leurgans

Adviser

Department of Statistics

To my father and the memory of my mother

ACKNOWLEDGMENTS

Support was received from the Ohio Supercomputer Center. The use of Robert Ross' FORTRAN program for computing full estimates for multilinear models, and help in modifying the program are acknowledged. Comments from Angela Dean were very helpful. I thank my adviser, Sue Leurgans, without whose guidance, patience, and encouragement this document would not have been completed. I am grateful to my family and friends for their unwavering support and understanding, and to all who have encouraged, cheered, and prayed for me throughout my studies.

VITA

October 1, 1959	Born-Metro-Manila, Philippines
1981	B.S., University of the Philippines Quezon City, Philippines
1985	M.S. in Statistics, University of the Philippines
1985-1986	University Fellow The Ohio State University Columbus, Ohio
1987	M.S. in Statistics, The Ohio State University
1988 to present	Graduate Research Assistant The Ohio State University

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER I

Introduction

1.1 Why Case Deletion?

Parameters are usually estimated based on a full or complete data set. Case deletion refers to leaving out one "case" or observation at a time and re-computing the parameter estimates. Why would one want to do that?

This idea of leaving out cases or observations has been used in several contexts. In data sets where there are some outlying observations, it is often desired to determine the effect that leaving out this observation will have on the results of the analysis. Or, one may wish to see how a particular estimator is affected by extreme values. Various measures which assess the influence of an observation have been constructed; such measures, quite naturally, involve case deletion.

Another area where the concept of leaving out observations appears is cross-validation. When a model is fit to a set of observations, the investigator would want some indication of how well the model fits. Or, the investigator's task may be to choose one of several models that will fit the data best. A reasonable way to do this is to test a candidate model on a set of observations different from the one on which estimates of the model parameters were obtained. Since the investigator usually has

only one data set at hand, the available observations can be divided such that only some are used in obtaining the model estimates, while the rest are set aside to be used later as a test set. Then the roles can be switched so that each set contributes to both the estimation and the testing of the model. The "limit" of such a procedure is setting aside one observation at a time, instead of a group of observations.

Case deletion is also used to obtain jackknife estimates. Like cross-validation, jackknifing involves leaving out one data point at a time, then re-computing estimates based on the rest. The goal, however, is to arrive at estimates which possess some desirable property to a larger degree (or an undesirable property to a lesser degree) than the original estimates calculated from the complete data set. Originally proposed to reduce bias, jackknife estimates have also become useful in estimating the variances of estimators when there are no cut-and-dried formulas for them. Various jackknife estimators are presented in

Now that we have seen some purpose in case deletion, we turn to the question of its implementation. When the number of observations is large, deleting one case at a time can become a tedious and computationally intensive task. In linear regression, closed-form expressions exist for the parameter estimates when the r^{th} observation is deleted; these do not require re-computing the estimates every time a case is deleted. In nonlinear regression, however, parameters are usually estimated using iterative procedures, and to get case deletion estimates, it is necessary to actually remove one observation at a time and go through the entire calculation every time. In order to avoid such difficulty, some approximations have been suggested in the literature.

How well these approximations perform, compared to the estimates obtained when the observations are actually left out one at a time is a major point of interest here.

1.2 Linear, Nonlinear, and Multilinear

Multilinear models are a class of nonlinear models which are conditionally linear in the parameters. They are defined and described in Section 2.1. These models have been found to be especially useful in spectroscopy (see Leurgans and Ross, 1991a), among other areas. A discussion of some applications is given in Section 2.2.

The quantity of results that are available for nonlinear models pales in comparison with the volumes that have been written on the linear model. But what is known about nonlinear models can already be considered extensive, relative to what has been done for multilinear models. If results for linear models can be extended to nonlinear models, all the more can we apply methods used for nonlinear models to the special case of multilinear models. In particular, measures of curvature and approximations to the actual case deletion estimates used in nonlinear regression can and will be applied in the multilinear setting.

The extension of concepts from linear to nonlinear models usually involves linear approximation. For instance, the curved expectation surface of a nonlinear model is approximated by a tangent plane at a point in order to obtain parameter estimates. The approximation is given by a first-order Taylor series expansion of the expectation function about that point. Taking this first-order expansion is the same idea behind the linear approximation estimates presented here.

"Models with conditionally linear parameters enjoy some advantageous proper-

ties, which can be exploited in nonlinear regression."(Bates and Watts, 1988) The special structure of multilinear models that arises from their conditional linearity and from the way they are defined is evident in the form of their derivative matrices and second-derivative arrays. Unlike other nonlinear models, the multilinear model derivatives have been found to have closed-form expressions and well-defined relationships. This nice structure is used to advantage in many of the calculations.

1.3 Roadmap

The main purpose of this dissertation is to explore case deletion for multilinear models. In particular, how viable some of the case deletion approximations are, when applied to multilinear models, is of interest.

A general introduction to multilinear models is given in Chapter II. The terminology and notation for multilinear models is given in Section 2.1, those for nonlinear models are given in Section 2.3.1, and two are connected in Section 2.3.2. Translation of multilinear into nonlinear language makes it easy, later on, to apply nonlinear methods to multlinear models. An example using a small number of observations is introduced in Section 2.4; it serves to illustrate the material in this and the next chapter.

In Chapter III, the structure of a particular type of multilinear model is investigated. The velocities or first derivatives, and accelerations or second derivatives of a trilinear PARAFAC model are studied in detail. Closed-form expressions for them are derived. The formulas for the various estimates in Chapters V and VI use the velocities and accelerations heavily, so the results of Secitons 3.1.1 and 3.1.2 will

prove invaluable.

The quadratic approximation to the case deletion estimates given in Chapter 6 assumes that the expectation surface is more or less planar. Thus, a measure of planarity is necessary to verify that validity of this assumption for the data set under consideration. Curvature measures, which measure planarity and uniformity of parameter lines for a nonlinear model, are discussed in Section 3.2. A couple of examples to illustrate their application to PARAFAC models, as well as to pave the way for the quadratic approximation of Section 6.1.2 are given in Section 3.4.

Calculation of the curvature measures require that the combined dimension of the tangent and acceleration spaces be known. Although this can be determined numerically given a data set, the structure of the PARAFAC models again makes it plausible that it can be found analytically. Investigating linear dependencies among the velocities and accelerations indeed yields a formula for this quantity, at least for the one-factor model. The linear dependencies derived in Section 3.3 are also of interest in themselves since they help to further define the structure of these models.

Chapter IV gives a more detailed background of cross-validation. It discusses the relevance of case deletion in the cross-validation of multilinear models and points out that this is a major area of possible applications.

Chapter V presents some alternatives to the ordinary jackknife estimates. By relying on a linear approximation, these estimators avoid having to do the actual deletions. Some of the estimators presented here are evaluated for the numerical example of the next chapter.

The roads paved by the previous chapters all lead to Chapter 6. Here, linear and quadratic approximations to case deletion estimates are presented and applied to a simulated data set. The actual and approximate estimates are compared in order to see whether the approximations are viable.

Chapter 7 summarizes the results and suggests areas of application and further study.

CHAPTER II

Introduction to Multilinear Analysis

This chapter introduces the reader to multilinear models. Section 2.1 defines a multilinear model and sets down the notation and terminology peculiar to this class of nonlinear models. Some of its applications are given in Section 2.2. In Section 2.3, nonlinear models are discussed, with emphasis on the aspects that are relevant to multilinear analysis. The terminology and notation presented in Sections 2.1 and 2.3 will be used throughout the succeeding chapters. A small example using real data is given in Section 2.4 to illustrate the material in the preceding sections. This example will also be used repeatedly in the next chapter.

2.1 Terminology and Notation

A multilinear model represents an M-way array of observed values of a response variable Y as the sum of F simpler arrays, each the outer product of M vectors, plus a random term. The dimension of the m^{th} vector in each outer product ($m = 1, .., M$) is the same as the number of levels in the m^{th} way of the array. The number of outer products, F, is called the number of *components* or *factors* of the model. The random term has mean zero and unknown variance. The array of expected values of Y is denoted by μ.

The simplest multilinear model is a one-component bilinear model:

$$\begin{aligned} \boldsymbol{Y}[i,j] &= \boldsymbol{\alpha}[i]\boldsymbol{\beta}[j] + \epsilon[i,j] \\ E(\boldsymbol{Y}[i,j]) &= \boldsymbol{\mu}[i,j] = \boldsymbol{\alpha}[i]\boldsymbol{\beta}[j],\ i = 1,...,I;\ j = 1,...,J. \end{aligned} \tag{2.1}$$

where

$\boldsymbol{\alpha}$ is an I-dimensional vector corresponding to the first way,

$\boldsymbol{\beta}$ is a J-dimensional vector corresponding to the second way.

Equation (2.1) is written using the S notation for matrices described in Appendix A. In matrix form, the equation is

$$^{I\times J}\boldsymbol{\mu} = \boldsymbol{\alpha}\boldsymbol{\beta}^T = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_I \end{bmatrix} [\beta_1 \ldots \beta_J] .$$

In outer product form, $^{I\times J}\boldsymbol{\mu} = {}^{I\times 1}\boldsymbol{\alpha} \times {}^{J\times 1}\boldsymbol{\beta}$. Note that for fixed $\boldsymbol{\alpha}$, the elements of $\boldsymbol{\mu}$ are linear in $\boldsymbol{\beta}$. Similarly, $\boldsymbol{\mu}$ is a conditionally linear function of $\boldsymbol{\alpha}$. Thus $\boldsymbol{\mu}$ is a bilinear function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

A one-factor trilinear model is of the form

$$E(\boldsymbol{Y}[i,j,k]) = \boldsymbol{\alpha}[i]\boldsymbol{\beta}[j]\boldsymbol{\gamma}[k],\ i = 1,\ldots,I;\ j = 1,\ldots,J;\ k = 1,\ldots,K. \tag{2.2}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are as before, and $\boldsymbol{\gamma}$ is a $K-$dimensional vector corresponding to the third way. The S notation for arrays (see Appendix A) is used in Equation (2.2). The array $\boldsymbol{\mu}$ can also be written as: $^{I\times J\times K}\boldsymbol{\mu} = {}^{I\times 1}\boldsymbol{\alpha} \times {}^{J\times 1}\boldsymbol{\beta} \times {}^{K\times 1}\boldsymbol{\gamma}$. For fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $\boldsymbol{\mu}$ is linear in $\boldsymbol{\gamma}$; for fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ is linear in $\boldsymbol{\beta}$; and for fixed $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ is linear in $\boldsymbol{\alpha}$. Thus $\boldsymbol{\mu}$ is a trilinear function of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.

An F-factor bilinear model has the representation

$$\mu[i,j] = \sum_{f=1}^{F} \alpha_f[i]\beta_f[j], \quad i = 1,\ldots,I;\ j = 1,\ldots,J. \tag{2.3}$$

where the F vectors α_f's are each I-dimensional and correspond to the first way, and the F vectors β_f's are each J-dimensional and correspond to the second way. In matrix form,

$$\overset{I\times J}{\mu} = \overset{I\times F}{A}\overset{F\times J}{B^T} = \begin{bmatrix} \overset{I\times 1}{\alpha_1} \overset{I\times 1}{\alpha_2} \ldots \overset{I\times 1}{\alpha_F} \end{bmatrix} \begin{bmatrix} \overset{1\times J}{\beta_1^T} \\ \overset{1\times J}{\beta_2^T} \\ \vdots \\ \overset{1\times J}{\beta_F^T} \end{bmatrix}$$

In outer product form, $\overset{I\times J}{\mu} = \sum_{f=1}^{F} \overset{I\times 1}{\alpha_f} \times \overset{J\times 1}{\beta_f}$.

An F-component trilinear model is of the form

$$\mu[i,j,k] = \sum_{f=1}^{F} \alpha_f[i]\beta_f[j]\gamma_f[k], \quad i = 1,\ldots,I;\ j = 1,\ldots,J;\ k = 1,\ldots,K. \tag{2.4}$$

where the α_f's and β_f's are as before, and the F vectors γ_f's are K-dimensional and correspond to the third way of the array of expected values. This may be expressed in outer product form as $\overset{I\times J\times K}{\mu} = \sum_{f=1}^{F} \overset{I\times 1}{\alpha_f} \times \overset{J\times 1}{\beta_f} \times \overset{K\times 1}{\gamma_f}$. For a fixed k, the model can be expressed as

$$\overset{I\times J}{\mu_k} = \overset{I\times F}{A}\overset{F\times F}{D_k}\overset{F\times J}{B^T} = \begin{bmatrix} \overset{I\times 1}{\alpha_1} \overset{I\times 1}{\alpha_2} \ldots \overset{I\times 1}{\alpha_F} \end{bmatrix} \begin{bmatrix} \gamma_{k1} & 0 & \ldots & 0 \\ 0 & \gamma_{k2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \gamma_{kF} \end{bmatrix} \begin{bmatrix} \overset{1\times J}{\beta_1^T} \\ \overset{1\times J}{\beta_2^T} \\ \vdots \\ \overset{1\times J}{\beta_F^T} \end{bmatrix}$$

Models (2.1) through (2.4) are called PARAFAC (Parallel Factors) models, and may be extended to higher-way models. (See Carroll and Chang, 1970.)

Note that in (2.4), if $\boldsymbol{\alpha}_f$, $\boldsymbol{\beta}_f$ and $\boldsymbol{\gamma}_f$ are each multiplied by constants whose product is one, the array $\boldsymbol{\mu}$ of expected values does not change; that is,

$$\boldsymbol{\mu}[i,j,k] = \sum_{f=1}^{F} a_f \boldsymbol{\alpha}_f[i] b_f \boldsymbol{\beta}_f[j] c_f \boldsymbol{\gamma}_f[k],$$

for any constants a_f, b_f, and c_f such that $a_f b_f c_f = 1$. There are several ways to remove this nonidentifiability. One is to require $\boldsymbol{\alpha}_f$ and $\boldsymbol{\beta}_f$ to be unit vectors with nonnegative first elements, that is, $\| \boldsymbol{\alpha}_f \| \equiv 1$ and $\| \boldsymbol{\beta}_f \| \equiv 1 \; \forall f$. (See Leurgans, Ross and Abel, 1990.) An alternative way is to set $\boldsymbol{\alpha}_f[I] \equiv 1$ and $\boldsymbol{\beta}_f[J] \equiv 1 \; \forall f$. Finding the derivative matrix is much simpler when using this latter convention. Thus, the following normalization shall be used in the subsequent discussion: In Model (2.4), divide the vectors $\boldsymbol{\alpha}_f$ and $\boldsymbol{\beta}_f$ by the constants $\boldsymbol{\alpha}_f[I]$ and $\boldsymbol{\beta}_f[J]$, respectively, $\forall f$; then multiply the vector $\boldsymbol{\gamma}_f$ by $\boldsymbol{\alpha}_f[I]\boldsymbol{\beta}_f[J] \; \forall f$, thereby making $\boldsymbol{\alpha}_f[I]$ and $\boldsymbol{\beta}_f[J]$ equal to 1 but leaving $\boldsymbol{\mu}$ unchanged. Similarly, the parameters in Model (2.2) are normalized by setting $\boldsymbol{\alpha}[I] \equiv 1$ and $\boldsymbol{\beta}[J] \equiv 1$. For the bilinear models (2.1) and (2.3), the corresponding identifiability conditions are $\boldsymbol{\alpha}[I] \equiv 1$ and $\boldsymbol{\alpha}_f[I] \; \forall f$, respectively.

2.2 Applications

One of the earliest published applications of PARAFAC models is the study made by Harshman, Ladefoged, and Goldstein (1977) on the position of the tongue while a speaker pronounces vowels. X-ray movies were taken while a speaker was talking. During the central portion of a vowel, one frame of the movie was chosen, and lines were drawn on the X-ray photograph at selected positions. The height of the upper surface of the tongue was then measured with respect to those lines.

Observations were taken for 10 American-English vowels, measured at 13 positions or lines along the tongue, spoken by 5 speakers. PARAFAC models (with varying number of factors) were fit to the 10x13x5 array of tongue heights. A two-factor model was selected, yielding three matrices of parameter estimates: a 10x2 matrix $[\alpha_{if}]$ corresponding to the 10 vowels, a 13x2 matrix $[\beta_{jf}]$ corresponding to the 13 positions along the tongue, and a 5x2 matrix $[\gamma_{kf}]$ corresponding to the 5 speakers. The first two matrices were of most interest: The first was displayed with each row as a point in F-dimensional space; i.e., the vowels were represented by 10 points in 2-dimensional space. The resulting configuration was strikingly similar to two classical representations of American-English vowels. The second matrix was displayed with each column as a curve made of 13 points (j, β_{jf}), because the 13 positions along the tongue form a natural sequence.

A major area of application of multilinear models is spectroscopy. "Spectroscopy is the measurement of the absorption of particles by a specimen, or the emission of particles from a specimen, as a function of the energy of the particles... Common spectroscopy uses electromagnetic radiation, such as visible light. The particles are photons. The wavelength of the radiation is inversely proportional to photon energy, and wavelength is often used instead of energy as the independent variable." (Leurgans and Ross, 1991b) Additional independent variables, such as concentration, time, or other environmental conditions affecting absorption or emission of light may also be included.

In fluorescence spectroscopy, the response variable is the amount of light emit-

ted by a specimen composed of one or more components. The light emission of a specimen containing only one component can be modeled as

$$\mu[i,j] = \alpha[i]\beta[j]$$

where

$\mu[i,j]$ is the amount of light emitted,

$\alpha[i]$ is the number of photons absorbed when the specimen is illuminated with light of wavelength λ_i (also called excitation wavelength),

$\beta[j]$ is the fraction of absorbed photons which lead to emission of light at wavelength τ_j (also called emission wavelength).

For a specimen containing more than one chromophore or component, we have

$$\mu[i,j] = \sum_{f=1}^{F} \alpha_f[i]\beta_f[j]$$

where

$\alpha_f[i]$ is the absorbance of component f at wavelength λ_i,

$\beta_f[j]$ is the relative emission of chromophore f at detection wavelength τ_j,

$\mu[i,j]$ is the amount of light emission measured,

and F is the number of chromophores.

The absorbance and emission of light by chromophore f are considered independent events, so the amount of light emitted by chromophore f at wavelenth τ_j when illuminated with light at wavelength λ_i is $\alpha_f[i]\beta_f[j]$. It is assumed that there is no transfer of energy among the F chromophores, so that the amount of light emitted by the specimen is just the sum of the amounts of light emitted by its individual components.

Appelloff and Davidson (1981) provided the first application of multilinear analysis to three-way arrays of fluorescence data. The fluorescence intensity of a solution containing F chromophores were observed at 30 excitation wavelengths, 30 emission wavelengths, and 10 different times (or concentrations, since the concentration of a chromophore varies with time). The model is

$$\mu[i,j,k] = \sum_{f=1}^{F} \alpha_f[i]\beta_f[j]\gamma_f[k]$$

where

$\alpha_f[i]$, $\beta_f[j]$, and F are defined as before,

$\gamma_f[k]$ is the relative concentration of chromophore f at time t_k,

$\mu[i,j,k]$ is the fluorescence intensity at excitation wavelength λ_i, emission wavelength τ_j, and time t_k.

PARAFAC models were fit to real data with two and three chromophores, and to simulated data with two, three, and four chromophores. For the two- and three-factor models, the estimated spectra were very close to the true spectra, both for the real and the simulated data. The four-factor estimates were not as close.

Many more applications of PARAFAC models can be found in Kroonenberg(1983).

2.3 Multilinear Models as a Special Case of Nonlinear Models

The literature on nonlinear models is very much more extensive than that on multilinear models. A number of books and articles have been written on nonlinear regression, covering broad areas such as parameter estimation, or specialized topics

such as curvature measures and case deletion. In order to avail of this "common knowledge" and apply it to the uncommon area of multilinear analysis, it would be useful to cast multilinear models in a nonlinear regression framework. After all, multilinear models are, strictly speaking, nonlinear in nature.

Section 2.3.1 presents some notation and geometry for the nonlinear regression model. The same notation and geometrical concepts will be used in Chapters III, V and VI. Section 2.3.2 connects the multilinear notation in Section 2.1 and the nonlinear notation in Section 2.3.1 through some examples.

2.3.1 Nonlinear Models: Notation and Geometry

A nonlinear regression model can be written as

$$Y_r = \eta(x_r, \boldsymbol{\theta}) + \epsilon_r, \quad r = 1, \ldots, N \tag{2.5}$$

where η is a nonlinear function of the unknown parameter vector $\boldsymbol{\theta}$, ϵ_i are independent with zero expectation and unknown variance. Considering the x_i's as fixed, we may rewrite the model as

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{2.6}$$

where $\boldsymbol{Y}$ is an $N \times 1$ vector of responses, $\boldsymbol{\eta}$ is an $N \times 1$ vector whose elements are functions of the $P \times 1$ vector of parameters $\boldsymbol{\theta}$. Note that $\boldsymbol{Y}$ is nonlinear in the parameters.

Let:

Ω be the subset of $\Re^p$ consisting of all possible values of $\boldsymbol{\theta}$;

$M = \{\boldsymbol{\eta}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Omega\} \in \Re^N$ be called the expectation surface;

$\boldsymbol{\epsilon}(\boldsymbol{\theta}) = \boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})$;

$S(\boldsymbol{\theta}) = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \|\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2$;

$\hat{\boldsymbol{\theta}}$ be the least squares estimate of $\boldsymbol{\theta}$, i.e., the value of $\boldsymbol{\theta}$ that minimizes $S(\boldsymbol{\theta})$;

$\hat{\boldsymbol{\eta}} = \eta(\hat{\boldsymbol{\theta}})$ be the point on M closest to $\boldsymbol{y}$;

$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{\eta}}$ be the residual vector;

$S(\hat{\boldsymbol{\theta}}) = \boldsymbol{e}^T\boldsymbol{e}$;

$\boldsymbol{V}_p = \dot{\boldsymbol{\eta}}_p = \frac{\partial \boldsymbol{\eta}}{\partial \theta_p}$, $1 \leq p \leq P$, the partial derivative of $\boldsymbol{\eta}$ with respect to the p^{th} component of $\boldsymbol{\theta}$, also called *tangent vector* or *velocity vector* ;

$\ddot{\boldsymbol{\eta}}_{pq} = \frac{\partial^2 \boldsymbol{\eta}}{\partial \theta_p \partial \theta_q}$, $1 \leq p, q \leq P$, called an *acceleration vector* ;

$\boldsymbol{V}$ be the $N \times P$ matrix whose columns are $\boldsymbol{V}_p$, called the *derivative matrix*;

$\ddot{\boldsymbol{\eta}}$ be the $N \times P \times P$ array whose n^{th} face is $(\ddot{\boldsymbol{\eta}}_{npq})$;

$\boldsymbol{G} = \boldsymbol{V}^T\boldsymbol{V} = (g_{pq})$.

The problem of finding the least squares estimates can be stated geometrically (Bates and Watts, 1988) as: (1) finding the point $\hat{\boldsymbol{\eta}}$ on the expectation surface which is closest to $\boldsymbol{y}$, then (2) determining the parameter vector $\hat{\boldsymbol{\theta}}$ which corresponds to $\hat{\boldsymbol{\eta}}$.

For a linear model, an explicit expression for $\hat{\boldsymbol{\eta}}$ exists. Also, the parameter plane maps linearly to the expectaion plane, so if we know where we are on one plane, the corresponding point on the other plane can be found. For a nonlinear model, it is not easy to find $\hat{\boldsymbol{\eta}}$, and even if $\hat{\boldsymbol{\eta}}$ were known, it would be difficult to find the $\hat{\boldsymbol{\theta}}$ corresponding to it. To overcome these difficulties, iterative methods are used.

One such method, the Gauss-Newton method, consists of expanding the expec-

tation function in a first-order Taylor series about an initial guess for $\boldsymbol{\theta}$, $\boldsymbol{\theta}^0$. This initial guess is iteratively improved until the estimates stabilize. The approximation of $\boldsymbol{\eta}(\boldsymbol{\theta})$ by a Taylor series expansion at $\boldsymbol{\eta}^0 = \boldsymbol{\eta}(\boldsymbol{\theta}^0)$ involves the two distinct approximations:

(1) approximating the expectation surface by its tangent plane at $\boldsymbol{\eta}(\boldsymbol{\theta}^0)$ (called the planar assumption), and

(2) approximating the true parameter coordinate system by a uniform coordinate system, i.e., imposing a linear coordinate system on the approximating tangent plane (called the uniform coordinate assumption).

The second-order partial derivatives $\ddot{\boldsymbol{\eta}}_{pq}$ can be decomposed into their *tangential* and *normal* components, denoted by $\ddot{\boldsymbol{\eta}}^T_{pq}$ and $\ddot{\boldsymbol{\eta}}^N_{pq}$, respectively (Bates and Watts, 1980). The space spanned by the vectors $\ddot{\boldsymbol{\eta}}^N_{pq}$ $(1 \leq p, q \leq P)$ is called the *acceleration space.* Let $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_m$ $(m \leq p(p+1)/2)$ be an orthonormal basis for the acceleration space. Then each $\ddot{\boldsymbol{\eta}}_{pq}$ can be written as a linear combination of the vectors $\dot{\boldsymbol{\eta}}_s$ $(s = 1, \ldots, P)$ and $\boldsymbol{\delta}_a$ $(a = 1, \ldots, m)$:

$$\ddot{\boldsymbol{\eta}}_{pq} = \Gamma^s_{pq}\dot{\boldsymbol{\eta}}_s + b^a_{pq}\boldsymbol{\delta}_a, \; 1 \leq p, q \leq P \tag{2.7}$$

The above equation uses the Einstein summation convention, where a repeated index, once as a superscript and once as a subscript, means summation over the range of that index, i.e.,

$$\begin{aligned} \Gamma^s_{pq}\dot{\boldsymbol{\eta}}_s &= \sum_{s=1}^{P} \Gamma^s_{pq}\dot{\boldsymbol{\eta}}_s \\ b^a_{pq}\boldsymbol{\delta}_a &= \sum_{a=1}^{m} b^a_{pq}\boldsymbol{\delta}_a \, . \end{aligned} \tag{2.8}$$

The coefficients Γ^s_{pq} are called *connection coefficients* by Ross (1987), and correspond to the elements of the *parameter-effects curvature array* of Bates and Watts (1980) in a standardized parameterization. The coefficients b^a_{pq} correspond to the elements of their *intrinsic curvature array*, and are referred to as *coefficients of the second fundamental form* by Ross. The connection coefficients reflect the degree of parameter nonuniformity, while the coefficients of the second fundamental form indicate the degree of nonplanarity of the model. The faces of the intrinsic curvature array are given by the $P \times P$ matrices $\boldsymbol{B}_a = (b^a_{pq})$. Section 3.2 discusses how the curvature arrays are obtained.

2.3.2 Some Multilinear Models in Nonlinear Notation

The class of nonlinear regression models contains the class of multilinear models, and in particular the PARAFAC models. The following examples illustrate this:

1. Consider a one-component bilinear model (2.1) with $I = 2$ and $J = 2$. Here, $N = 4 = I \cdot J$, $P = 3 = I - 1 + J$, and

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha}[1] \\ \boldsymbol{\beta}[1] \\ \boldsymbol{\beta}[2] \end{bmatrix}.$$

In the form of (2.6), we can write the model as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}[1]\boldsymbol{\beta}[1] \\ \boldsymbol{\alpha}[1]\boldsymbol{\beta}[2] \\ 1 \cdot \boldsymbol{\beta}[1] \\ 1 \cdot \boldsymbol{\beta}[2] \end{bmatrix} + \boldsymbol{\epsilon}$$

or

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = vec(\boldsymbol{\alpha} \times \boldsymbol{\beta})^T = \boldsymbol{\alpha} \otimes \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\theta}[1] \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{\theta}[2] \\ \boldsymbol{\theta}[3] \end{bmatrix},$$

where the operator $\otimes$ represents a Kronecker product and *vec* denotes the vectorization of a matrix or array. Kronecker products and vectorization are defined in Appendix B.

2. In a one-factor trilinear model (2.2) with $I = 2$, $J = 3$, $K = 2$ there are $N = 12 = I \cdot J \cdot K$ observations and $P = 5 = I - 1 + J - 1 + K$ elements in the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha}[1] \\ \boldsymbol{\beta}[1] \\ \boldsymbol{\beta}[2] \\ \boldsymbol{\gamma}[1] \\ \boldsymbol{\gamma}[2] \end{bmatrix} . \tag{2.9}$$

This model can be written in the form of (2.6) as

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}[1]\boldsymbol{\beta}[1]\boldsymbol{\gamma}[1] \\ \boldsymbol{\alpha}[1]\boldsymbol{\beta}[1]\boldsymbol{\gamma}[2] \\ \boldsymbol{\alpha}[1]\boldsymbol{\beta}[2]\boldsymbol{\gamma}[1] \\ \boldsymbol{\alpha}[1]\boldsymbol{\beta}[2]\boldsymbol{\gamma}[2] \\ \boldsymbol{\alpha}[1] \cdot 1 \cdot \boldsymbol{\gamma}[1] \\ \boldsymbol{\alpha}[1] \cdot 1 \cdot \boldsymbol{\gamma}[2] \\ 1 \cdot \boldsymbol{\beta}[1]\boldsymbol{\gamma}[1] \\ 1 \cdot \boldsymbol{\beta}[1]\boldsymbol{\gamma}[2] \\ 1 \cdot \boldsymbol{\beta}[2]\boldsymbol{\gamma}[1] \\ 1 \cdot \boldsymbol{\beta}[2]\boldsymbol{\gamma}[2] \\ 1 \cdot 1 \cdot \boldsymbol{\gamma}[1] \\ 1 \cdot 1 \cdot \boldsymbol{\gamma}[2] \end{bmatrix} + \boldsymbol{\epsilon} \tag{2.10}$$

or

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = vec(\boldsymbol{\alpha} \times \boldsymbol{\beta} \times \boldsymbol{\gamma})^T = \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\theta}[1] \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{\theta}[2] \\ \boldsymbol{\theta}[3] \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{\theta}[4] \\ \boldsymbol{\theta}[5] \end{bmatrix} .$$

3. An F-component trilinear model (2.4) with $I = 2$, $J = 3$, $K = 2$, $F = 2$ has $N = 2 \cdot 3 \cdot 2 = 12$ observations and

$$P = (I - 1 + J - 1 + K)F = (1 + 2 + 2)2 = 10$$

elements in the parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\alpha}_1[1] \\ \boldsymbol{\alpha}_2[1] \\ \boldsymbol{\beta}_1[1] \\ \boldsymbol{\beta}_1[2] \\ \boldsymbol{\beta}_2[1] \\ \boldsymbol{\beta}_2[2] \\ \boldsymbol{\gamma}_1[1] \\ \boldsymbol{\gamma}_1[2] \\ \boldsymbol{\gamma}_2[1] \\ \boldsymbol{\gamma}_2[2] \end{bmatrix}. \tag{2.11}$$

In the form of (2.6), the model can be written as

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1[1]\boldsymbol{\beta}_1[1]\boldsymbol{\gamma}_1[1] + \boldsymbol{\alpha}_2[1]\boldsymbol{\beta}_2[1]\boldsymbol{\gamma}_2[1] \\ \boldsymbol{\alpha}_1[1]\boldsymbol{\beta}_1[1]\boldsymbol{\gamma}_1[2] + \boldsymbol{\alpha}_2[1]\boldsymbol{\beta}_2[1]\boldsymbol{\gamma}_2[2] \\ \boldsymbol{\alpha}_1[1]\boldsymbol{\beta}_1[2]\boldsymbol{\gamma}_1[1] + \boldsymbol{\alpha}_2[1]\boldsymbol{\beta}_2[2]\boldsymbol{\gamma}_2[1] \\ \boldsymbol{\alpha}_1[1]\boldsymbol{\beta}_1[2]\boldsymbol{\gamma}_1[2] + \boldsymbol{\alpha}_2[1]\boldsymbol{\beta}_2[2]\boldsymbol{\gamma}_2[2] \\ \boldsymbol{\alpha}_1[1] \cdot 1 \cdot \boldsymbol{\gamma}_1[1] + \boldsymbol{\alpha}_2[1] \cdot 1 \cdot \boldsymbol{\gamma}_2[1] \\ \boldsymbol{\alpha}_1[1] \cdot 1 \cdot \boldsymbol{\gamma}_1[2] + \boldsymbol{\alpha}_2[1] \cdot 1 \cdot \boldsymbol{\gamma}_2[2] \\ 1 \cdot \boldsymbol{\beta}_1[1]\boldsymbol{\gamma}_1[1] + 1 \cdot \boldsymbol{\beta}_2[1]\boldsymbol{\gamma}_2[1] \\ 1 \cdot \boldsymbol{\beta}_1[1]\boldsymbol{\gamma}_1[2] + 1 \cdot \boldsymbol{\beta}_2[1]\boldsymbol{\gamma}_2[2] \\ 1 \cdot \boldsymbol{\beta}_1[2]\boldsymbol{\gamma}_1[1] + 1 \cdot \boldsymbol{\beta}_2[2]\boldsymbol{\gamma}_2[1] \\ 1 \cdot \boldsymbol{\beta}_1[2]\boldsymbol{\gamma}_1[2] + 1 \cdot \boldsymbol{\beta}_2[2]\boldsymbol{\gamma}_2[2] \\ 1 \cdot 1 \cdot \boldsymbol{\gamma}_1[1] + 1 \cdot 1 \cdot \boldsymbol{\gamma}_2[1] \\ 1 \cdot 1 \cdot \boldsymbol{\gamma}_1[2] + 1 \cdot 1 \cdot \boldsymbol{\gamma}_2[2] \end{bmatrix} + \boldsymbol{\epsilon} \tag{2.12}$$

or

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = vec(\sum_{f=1}^{2} \boldsymbol{\alpha}_f \times \boldsymbol{\beta}_f \times \boldsymbol{\gamma}_f)^T = \sum_{f=1}^{2} \boldsymbol{\alpha}_f \otimes \boldsymbol{\beta}_f \otimes \boldsymbol{\gamma}_f.$$

2.4 A Small Example: Plastocyanin Absorption

We now consider an illustration based on real data. This will serve as a running example to help clarify the material in the preceding as well as the forthcoming sections. The data is taken from Durell, Draheim and Gross (1988). To keep the

illustration simple and tractable, we shall look only at a subset of the original data set.

Plastocyanin is a protein, found in plant chloroplasts, which participates in electron transport. The data consists of amounts of light absorbed by this substance measured at different wavelengths, species of plastocyanin, oxidation states and environmental pH. Absorption measurements were made at 30 different wavelengths, for 4 biological sources or species (spinach, poplar, lettuce, parsley), and 4 chemical treatments (combinations of oxidation state and pH variables). The original data can thus be arranged in a 30 by 4 by 4 array of observations, from which we take a 2 by 3 by 2 subset. The selected observations correspond to wavelengths of 250 nm and 262 nm, the first 3 species, and the first two treatment levels. Below is the $2 \times 3 \times 2$ data array $\boldsymbol{Y}[i, j, k]$ in S array notation:

$\boldsymbol{Y}[,,1]$:

	$\boldsymbol{Y}[,1,]$	$\boldsymbol{Y}[,2,]$	$\boldsymbol{Y}[,3,]$
$\boldsymbol{Y}[1,,]$	4.2	4.4	4.3
$\boldsymbol{Y}[2,,]$	4.3	4.1	4.2

$\boldsymbol{Y}[,,2]$:

	$\boldsymbol{Y}[,1,]$	$\boldsymbol{Y}[,2,]$	$\boldsymbol{Y}[,3,]$
$\boldsymbol{Y}[1,,]$	3.4	3.7	3.9
$\boldsymbol{Y}[2,,]$	3.9	3.4	3.8

In vectorized form,

$$vec(\boldsymbol{Y}[i,j,k])^T = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{bmatrix} = \begin{bmatrix} 4.2 \\ 3.4 \\ 4.4 \\ 3.7 \\ 4.3 \\ 3.9 \\ 4.3 \\ 3.9 \\ 4.1 \\ 3.4 \\ 4.2 \\ 3.8 \end{bmatrix} \tag{2.13}$$

where $vec(\boldsymbol{Y}[i,j,k])^T$ denotes the modified vectorization of an array defined in Appendix A.

Previous analysis of the original 30 by 4 by 4 data set indicated that at least 2 factors were present. (See Leurgans and Ross, 1991a.) To keep the example simple, however, we shall use only a one-factor model. The multilinear model for the 2 by 3 by 2 data set is therefore

$$E(\boldsymbol{Y}[i,j,k]) = \boldsymbol{\mu}[i,j,k] = \boldsymbol{\alpha}[i]\boldsymbol{\beta}[j]\boldsymbol{\gamma}[k], \tag{2.14}$$

$$i = 1,2;\ j = 1,2,3;\ k = 1,2. \tag{2.15}$$

A 2-factor trilinear PARAFAC model was fit to the 12 observations using R.T. Ross' FORTRAN program (see Leurgans, Ross and Abel, 1990) and the least squares estimates for the second of the factors was chosen to represent the estimated parameter values of the one-factor model. These, rounded to 4 decimal places, are:

$$\hat{\boldsymbol{\alpha}} = \begin{bmatrix} 0.7231 \\ 0.6907 \end{bmatrix},\ \hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.5710 \\ 0.5716 \\ 0.5893 \end{bmatrix},\ \text{and}\ \hat{\boldsymbol{\gamma}} = \begin{bmatrix} 10.3918 \\ 8.9853 \end{bmatrix}.$$

Using the normalization convention described in Section 2.1, we divide α by 0.6907, we divide β by 0.5893, and we multiply γ by both these constants to obtain the following normalized least squares estimates:

Parameter	Original LSE	Normalized LSE
$\alpha[1]$	0.7231	1.0469
$\alpha[2]$	0.6907	1.0000
$\beta[1]$	0.5710	0.9689
$\beta[2]$	0.5716	0.9700
$\beta[3]$	0.5893	1.0000
$\gamma[1]$	10.3918	4.2299
$\gamma[2]$	8.9853	3.6574

Note that in this example, $I = 2$, $J = 3$, $K = 2$, and $F = 1$, just as in the second example of Section 2.3.2. So the 5×1 vector θ is given by Equation 2.9 and its least squares estimate is

$$\hat{\theta} = \begin{bmatrix} 1.0469 \\ 0.9689 \\ 0.9700 \\ 4.2299 \\ 3.6574 \end{bmatrix}.$$

Equation 2.10 gives the nonlinear regression form of 2.15. So the estimate of $\eta(\theta)$ is

$$\hat{\eta} = \eta(\hat{\theta}) = \begin{bmatrix} 4.2906 \\ 3.7099 \\ 4.2953 \\ 3.7140 \\ 4.4284 \\ 3.8290 \\ 4.0983 \\ 3.5436 \\ 4.1028 \\ 3.5475 \\ 4.1028 \\ 3.5475 \\ 4.2299 \\ 3.6574 \end{bmatrix}.$$

CHAPTER III

Curvature Measures of Nonlinearity for Trilinear PARAFAC Models

For linear models, the first derivatives of the expectation function with respect to the parameters are constants, and the second derivatives are identically zero. This is not the case for nonlinear models, in general, and for multilinear models, in particular. Thus, the acceleration vectors $\ddot{\eta}_{pq}$ have been used to measure nonlinearity of a model (Bates and Watts, 1980). Measures of intrinsic curvature (how planar the expectation surface M is, or how well M can be approximated by its tangent plane), and of parameter effects curvature (how uniform the parameter lines are on the tangent plane) have been derived using the $\ddot{\eta}_{pq}$'s.

In Section 6.1.2, the quadratic approximation to case deletion estimates of the parameters is a simple expression that has been derived assuming that the expectation surface is more or less planar. Bates and Watts (1988) studied real data sets using different nonlinear models (67 data set-model combinations in all) and found that in most of them, the expectation surface could be well-approximated by a tangent plane. Their results strongly support the assumption of planarity. It would be of great interest to us to see whether the planarity assumption also seems to hold for PARAFAC models.

Measures of nonplanarity and parameter nonuniformity for nonlinear models are discussed in Section 3.2. In Section 3.4, these measures are applied to a trilinear PARAFAC model. In Section 3.1 expressions for the derivative matrix $\boldsymbol{V}$, its submatrices, the velocity vectors $\boldsymbol{V}_p$, and the acceleration vectors $\ddot{\boldsymbol{\eta}}_{pq}$ are derived. Besides being of interest in themselves, $\boldsymbol{V}$ and $\ddot{\boldsymbol{\eta}}$ are necessary ingredients in the curvature measures of Section 3.2, in the jackknife estimators of Chapter V, and in the approximate case deletion estimators of Chapter VI. Because they have a number of uses, the velocities and accelerations are represented in various forms, some more convenient for certain purposes than others. Linear relationships among the velocities and accelerations are investigated in Section 3.3. The results of that section, which are interesting in themselves, are also used in the calculation of curvature measures.

3.1 Velocity and Acceleration

3.1.1 Velocity

The special structure of PARAFAC models yields "nice" forms for the matrix of partial derivatives $\boldsymbol{V}$. For instance, for the one-factor bilinear model, $\boldsymbol{V}$ has the form

$$\underset{IJ\times I-1+J}{\boldsymbol{V}} = \begin{bmatrix} \underset{J\times I-1}{\boldsymbol{I}_{I-1}\otimes\boldsymbol{\beta}} & \boldsymbol{\alpha}\otimes\boldsymbol{I}_J \\ \underset{}{\boldsymbol{0}} & \end{bmatrix} = \begin{bmatrix} \underset{J\times I-1}{\boldsymbol{I}_{I-1}\otimes\boldsymbol{\beta}} & \boldsymbol{\alpha}^*\otimes\boldsymbol{I}_J \\ \underset{}{\boldsymbol{0}} & \boldsymbol{I}_J \end{bmatrix}$$

where $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}[-I]$ is $\boldsymbol{\alpha}$ with the last row deleted.

For the one-factor trilinear model, $\boldsymbol{V}$ is given by

$$\underset{IJK\times I+J+K-2}{\boldsymbol{V}} = \left[\begin{array}{c|c|c} \underset{JK\times I-1}{\boldsymbol{I}_{I-1}\otimes\boldsymbol{\beta}\otimes\boldsymbol{\gamma}} & \boldsymbol{\alpha}\otimes\begin{bmatrix}\underset{K\times J-1}{\boldsymbol{I}_{J-1}\otimes\boldsymbol{\gamma}}\\ \underset{}{\boldsymbol{0}}\end{bmatrix} & \boldsymbol{\alpha}\otimes\boldsymbol{\beta}\otimes\boldsymbol{I}_K \end{array}\right] \tag{3.1}$$

The first $I-1$ columns of $\boldsymbol{V}$ contain the derivative of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\alpha}$; the next $J-1$ columns, the derivative with respect to $\boldsymbol{\beta}$; and the last K columns, the derivative with respect to $\boldsymbol{\gamma}$.

For Example 2 in Section 2.3.2 and the plastocyanin example of Section 2.4, $\boldsymbol{V}$ is the 12×5 matrix,

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{\beta}[1]\boldsymbol{\gamma}[1] & \boldsymbol{\alpha}[1]\boldsymbol{\gamma}[1] & 0 & \boldsymbol{\alpha}[1]\boldsymbol{\beta}[1] & 0 \\ \boldsymbol{\beta}[1]\boldsymbol{\gamma}[2] & \boldsymbol{\alpha}[1]\boldsymbol{\gamma}[2] & 0 & 0 & \boldsymbol{\alpha}[1]\boldsymbol{\beta}[1] \\ \boldsymbol{\beta}[2]\boldsymbol{\gamma}[1] & 0 & \boldsymbol{\alpha}[1]\boldsymbol{\gamma}[1] & \boldsymbol{\alpha}[1]\boldsymbol{\beta}[2] & 0 \\ \boldsymbol{\beta}[2]\boldsymbol{\gamma}[2] & 0 & \boldsymbol{\alpha}[1]\boldsymbol{\gamma}[2] & 0 & \boldsymbol{\alpha}[1]\boldsymbol{\beta}[2] \\ \boldsymbol{\gamma}[1] & 0 & 0 & \boldsymbol{\alpha}[1] & 0 \\ \boldsymbol{\gamma}[2] & 0 & 0 & 0 & \boldsymbol{\alpha}[1] \\ 0 & \boldsymbol{\gamma}[1] & 0 & \boldsymbol{\beta}[1] & 0 \\ 0 & \boldsymbol{\gamma}[2] & 0 & 0 & \boldsymbol{\beta}[1] \\ 0 & 0 & \boldsymbol{\gamma}[1] & \boldsymbol{\beta}[2] & 0 \\ 0 & 0 & \boldsymbol{\gamma}[2] & 0 & \boldsymbol{\beta}[2] \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.2}$$

Observe that (3.2) can be written as

$$\underset{12\times 5}{\boldsymbol{V}} = \left[\begin{array}{c|c|c} \underset{6\times 1}{\boldsymbol{\beta}\otimes\boldsymbol{\gamma}} & \boldsymbol{\alpha}\otimes\begin{bmatrix} \underset{2\times 2}{\boldsymbol{I}_2\otimes\boldsymbol{\gamma}} \\ \mathbf{0} \end{bmatrix} & \boldsymbol{\alpha}\otimes\boldsymbol{\beta}\otimes\boldsymbol{I}_2 \end{array} \right], \tag{3.3}$$

which is just (3.1) with $I=2$, $J=3$, and $K=2$. For the plastocyanin data, the derivative matrix evaluated at $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{V}} = \begin{bmatrix} 4.0983 & 4.4284 & 0 & 1.0144 & 0 \\ 3.5436 & 3.8290 & 0 & 0 & 1.0144 \\ 4.1028 & 0 & 4.4284 & 1.0155 & 0 \\ 3.5475 & 0 & 3.3290 & 0 & 1.0155 \\ 4.2299 & 0 & 0 & 1.0469 & 0 \\ 3.6574 & 0 & 0 & 0 & 1.0469 \\ 0 & 4.2299 & 0 & 0.9689 & 0 \\ 0 & 3.6574 & 0 & 0 & 0.9689 \\ 0 & 0 & 4.2299 & 0.9700 & 0 \\ 0 & 0 & 3.6574 & 0 & 0.9700 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}. \tag{3.4}$$

For an F-factor trilinear model, $\boldsymbol{V}$ is an IJK by $(I-1+J-1+K)F$ matrix which may be partitioned into submatrices, each of which corresponds to the partial derivatives of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\alpha}_f^*$, $\boldsymbol{\beta}_f^*$, and $\boldsymbol{\gamma}_f$. The vectors $\boldsymbol{\alpha}_f^*$ and $\boldsymbol{\beta}_f^*$ are $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with the I^{th} and J^{th} rows deleted, respectively. We have

$$\boldsymbol{V} = \left[\begin{array}{ccc} \boldsymbol{V}_A & \boldsymbol{V}_B & \boldsymbol{V}_C \end{array} \right]$$

where, for $f = 1, 2, \ldots, F$,

$$\begin{aligned}
\overset{IJK\times(I-1)F}{\boldsymbol{V}_A} &= (\boldsymbol{V}_{Af}), \\
\overset{IJK\times(J-1)F}{\boldsymbol{V}_B} &= (\boldsymbol{V}_{Bf}), \\
\overset{IJK\times KF}{\boldsymbol{V}_C} &= (\boldsymbol{V}_{Cf}), \\
\overset{IJK\times I-1}{\boldsymbol{V}_{Af}} &= \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\alpha}_{if}}\right), i = 1, \ldots, I-1, \\
\overset{IJK\times J-1}{\boldsymbol{V}_{Bf}} &= \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}_{jf}}\right), j = 1, \ldots, J-1, \\
\overset{IJK\times K}{\boldsymbol{V}_{Cf}} &= \left(\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\gamma}_{kf}}\right), k = 1, \ldots, K,
\end{aligned} \tag{3.5}$$

so that $\boldsymbol{V}_{Af}$, $\boldsymbol{V}_{Bf}$, and $\boldsymbol{V}_{Cf}$ are the matrices of partial derivatives of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\alpha}_f^*$, $\boldsymbol{\beta}_f^*$, and $\boldsymbol{\gamma}_f$, respectively. The expression for $\boldsymbol{V}$ is essentially (3.1) with $\boldsymbol{\alpha}_f$, $\boldsymbol{\beta}_f$, and $\boldsymbol{\gamma}_f$ in place of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$. Thus for $f = 1, \ldots, F$,

$$\begin{aligned}
\boldsymbol{V}_{Af} &= \left[\begin{array}{c} \boldsymbol{I}_{I-1} \otimes \boldsymbol{\beta}_f \otimes \boldsymbol{\gamma}_f \\ \boldsymbol{0} \end{array} \right] \\
\boldsymbol{V}_{Bf} &= \boldsymbol{\alpha}_f \otimes \left[\begin{array}{c} \boldsymbol{I}_{J-1} \otimes \boldsymbol{\gamma}_f \\ \boldsymbol{0} \end{array} \right] \\
\boldsymbol{V}_{Cf} &= \boldsymbol{\alpha}_f \otimes \boldsymbol{\beta}_f \otimes \boldsymbol{I}_K.
\end{aligned}$$

The i^{th} column of $\boldsymbol{V}_{Af}$, the j^{th} column of $\boldsymbol{V}_{Bf}$, and the k^{th} column of $\boldsymbol{V}_{Cf}$ can

be written, respectively, as

$$\boldsymbol{V}_{Af}[,i] = \begin{bmatrix} \overset{(i-1)JK\times 1}{\mathbf{0}} \\ \boldsymbol{\beta}_f \otimes \boldsymbol{\gamma}_f \\ \underset{(I-i)JK\times 1}{\mathbf{0}} \end{bmatrix}, \; i = 1, \ldots, I-1 \tag{3.6}$$

$$\boldsymbol{V}_{Bf}[,j] = \boldsymbol{\alpha}_f \otimes \begin{bmatrix} \overset{(j-1)K\times 1}{\mathbf{0}} \\ \boldsymbol{\gamma}_f \\ \underset{(J-j)K\times 1}{\mathbf{0}} \end{bmatrix}, \; j = 1, \ldots, J-1 \tag{3.7}$$

$$\boldsymbol{V}_{Cf}[,k] = \boldsymbol{\alpha}_f \otimes \boldsymbol{\beta}_f \otimes \begin{bmatrix} \overset{(k-1)\times 1}{\mathbf{0}} \\ 1 \\ \underset{(K-k)\times 1}{\mathbf{0}} \end{bmatrix} = \boldsymbol{\alpha}_f \otimes \boldsymbol{\beta}_f \otimes \boldsymbol{e}_k, \; k = 1, \ldots, K, \tag{3.8}$$

where $\boldsymbol{e}_k$ is the vector of length K with 1 in the k^{th} position, 0 elsewhere.

For Example 2 in Section 2.3.2 and the plastocyanin example, since there is only one factor, $\boldsymbol{V}$ can be partitioned into the 12×1 matrix $\boldsymbol{V}_{A1} = \boldsymbol{V}_A$, the 12×2 matrix $\boldsymbol{V}_{B1} = \boldsymbol{V}_B = (\boldsymbol{V}_B[,1], \; \boldsymbol{V}_B[,2])$, and the 12×2 matrix $\boldsymbol{V}_{C1} = \boldsymbol{V}_C = (\boldsymbol{V}_C[,1], \; \boldsymbol{V}_C[,2])$. Equations (3.6), (3.7), and (3.8) yield

$$\boldsymbol{V}_A[,1] = \boldsymbol{\beta} \otimes \boldsymbol{\gamma} \; ;$$

$$\boldsymbol{V}_B[,1] = \alpha \otimes \begin{bmatrix} \boldsymbol{\gamma} \\ \underset{4\times 1}{\mathbf{0}} \end{bmatrix}, \qquad \boldsymbol{V}_B[,2] = \alpha \otimes \begin{bmatrix} \overset{2\times 1}{\mathbf{0}} \\ \boldsymbol{\gamma} \\ \underset{2\times 1}{\mathbf{0}} \end{bmatrix} \; ;$$

$$\boldsymbol{V}_C[,1] = \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{V}_C[,2] = \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

It is easily verified that the vectors above are indeed the columns of the derivative matrix (3.2).

An alternative expression for $\boldsymbol{V}$ can be obtained as follows: Let

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \ldots & \boldsymbol{\alpha}_F \end{bmatrix}$$

$$\boldsymbol{B} = \left[\begin{array}{cccc} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_F \end{array} \right]$$

$$\boldsymbol{C} = \left[\begin{array}{cccc} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_F \end{array} \right]$$

$$\begin{aligned} \overset{IJ\times F}{\boldsymbol{AB}} &= \boldsymbol{A} \odot \boldsymbol{B} \\ \overset{JK\times F}{\boldsymbol{BC}} &= \boldsymbol{B} \odot \boldsymbol{C} \\ \overset{IK\times F}{\boldsymbol{AC}} &= \boldsymbol{A} \odot \boldsymbol{C}, \end{aligned}$$

where $\odot$ is the circle product of Khatri and Rao (1968), defined in Appendix B. For $f = 1, \ldots, F$, define ABf to be the f^{th} block of length IJ of the vectorized matrix AB; BCf to be the f^{th} block of length JK of $vec(BC)$, and ACf to be the f^{th} block of length IK of $vec(AC)$. Using the S notation for matrices, we have

$$\begin{aligned} \boldsymbol{ABf} &= vec(\boldsymbol{AB})[IJ(f-1)+1 : IJf] \\ \boldsymbol{BCf} &= vec(\boldsymbol{BC})[JK(f-1)+1 : JKf] \\ \boldsymbol{ACf} &= vec(\boldsymbol{AC})[IK(f-1)+1 : IKf] \end{aligned}$$

Then,

$$\begin{aligned} \boldsymbol{V}_{Af} &= \boldsymbol{I}_I \otimes \boldsymbol{BCf} \\ \boldsymbol{V}_{Cf} &= \boldsymbol{ABf} \otimes \boldsymbol{I}_K \\ \boldsymbol{V}_{Bf} &= (\boldsymbol{I}_j \otimes \boldsymbol{ACf}_i)[,-J] \\ \boldsymbol{ACf}_i &= \boldsymbol{ACf}[K(i-1)+1 : Ki], \quad f = 1, \ldots, F;\ i = 1, \ldots, I, \end{aligned}$$

so that V_{Bf} consists essentially of permutations of the ACf's.

We now set some notation that will especially be useful in Section 3.3 and that will give more compact expressions for the velocity and acceleration vectors.

Let $\mathcal{I}_f$ be the vector whose elements pick out the elements in θ corresponding to α_f^*:

$$\mathcal{I}_f = (f-1)(I-1)\mathbf{1} + \begin{bmatrix} 1 \\ \vdots \\ I-1 \end{bmatrix} \tag{3.9}$$

where $\mathbf{1}$ is a vector of 1's of conformable length. Setting

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 \\ \vdots \\ \mathcal{I}_F \end{bmatrix},$$

we have

$$\begin{aligned} \theta[\mathcal{I}_f] &= \alpha_f^*, \\ \theta[\mathcal{I}] &= vec(A[-I,]), \\ V_{Af} &= V[,\mathcal{I}_f], \\ V_A &= V[,\mathcal{I}]. \end{aligned}$$

To illustrate the above notation, consider Example 3 of Section 2.3.2, where $I = 2$, $J = 3$, $K = 2$, $F = 2$. Then (3.9) yields

$$\begin{aligned} \mathcal{I}_1 &= (1-1)(2-1)1 + [1] = 1, \\ \mathcal{I}_2 &= (2-1)(2-1)1 + [1] = 2, \text{ and} \\ \mathcal{I} &= \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \\ \text{So } \theta[\mathcal{I}_1] &= \theta[1] = \alpha_1^* = \alpha_1[1], \end{aligned}$$

$$\begin{aligned}\boldsymbol{\theta}[\mathcal{I}_2] &= \boldsymbol{\theta}[2] = \boldsymbol{\alpha}_2^* = \boldsymbol{\alpha}_2[1]\,, \text{ and} \\ \boldsymbol{\theta}[\mathcal{I}] &= \begin{bmatrix} \boldsymbol{\alpha}_1[1] \\ \boldsymbol{\alpha}_2[1] \end{bmatrix}.\end{aligned}$$

Comparing the above with Equation (2.11) verifies that $\mathcal{I}_f$ indeed picks out the elements of $\boldsymbol{\theta}$ corresponding to $\boldsymbol{\alpha}_f^*$. In that same example,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1[1] & \boldsymbol{\alpha}_2[1] \\ \boldsymbol{\alpha}_1[2] & \boldsymbol{\alpha}_2[2] \end{bmatrix}.$$

So $vec(\boldsymbol{A}[-I,])$ is the vectorization of the first row of $\boldsymbol{A}$, which is indeed $\boldsymbol{\theta}[\mathcal{I}]$. Observe also that $\boldsymbol{V}[,\mathcal{I}_1] = \boldsymbol{V}[,1]$, the first column of $\boldsymbol{V}$; $\boldsymbol{V}[,\mathcal{I}_2] = \boldsymbol{V}[,2]$, the second column of $\boldsymbol{V}$; and $\boldsymbol{V}[,\mathcal{I}] = \boldsymbol{V}[,1:2]$, the first two columns of $\boldsymbol{V}$, which agree with the definitions given by the set of equations (3.5).

Let $\mathcal{J}_f$ and $\mathcal{K}_f$ be the vectors whose elements give the positions, respectively, of $\boldsymbol{\beta}_f^*$ and of $\boldsymbol{\gamma}_f^*$ in $\boldsymbol{\theta}$:

$$\mathcal{J}_f = (F(I-1)+(f-1)(J-1))\mathbf{1} + \begin{bmatrix} 1 \\ \vdots \\ J-1 \end{bmatrix}, \tag{3.10}$$

$$\mathcal{K}_f = (F(I-1)+F(J-1)+(f-1)K)\mathbf{1} + \begin{bmatrix} 1 \\ \vdots \\ K \end{bmatrix}, \tag{3.11}$$

and set

$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_1 \\ \vdots \\ \mathcal{J}_F \end{bmatrix}, \text{ and } \mathcal{K} = \begin{bmatrix} \mathcal{K}_1 \\ \vdots \\ \mathcal{K}_F \end{bmatrix}.$$

Then

$$\begin{aligned}\boldsymbol{\theta}[\mathcal{J}_f] &= \boldsymbol{\beta}_f^*\,, \\ \boldsymbol{V}_{Bf} &= \boldsymbol{V}[,\mathcal{J}_f]\,,\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{V}_{\boldsymbol{B}} &= \boldsymbol{V}[,\mathcal{J}]\,, \\
\boldsymbol{\theta}[\mathcal{K}_f] &= \boldsymbol{\gamma}_f\,, \\
\boldsymbol{V}_{\mathcal{C}f} &= \boldsymbol{V}[,\mathcal{K}_f]\,, \\
\text{and } \boldsymbol{V}_{\mathcal{C}} &= \boldsymbol{V}[,\mathcal{K}]\,.
\end{aligned}
$$

To illustrate, consider again Example 3 in Section ??. There,

$$
\begin{aligned}
\mathcal{J}_1 &= (2(2-1)+(1-1)(3-1))\mathbf{1} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}\,, \\
\mathcal{J}_2 &= (2(2-1)+(2-1)(3-1))\mathbf{1} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}\,, \text{ and} \\
\mathcal{J} &= \begin{bmatrix} 3 \\ \vdots \\ 6 \end{bmatrix}\,. \\
\text{So } \boldsymbol{\theta}[\mathcal{J}_1] &= \boldsymbol{\theta}[3:4] = \boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1[1:2]\,, \\
\boldsymbol{\theta}[\mathcal{J}_2] &= \boldsymbol{\theta}[5:6] = \boldsymbol{\beta}_2^* = \boldsymbol{\beta}_2[1:2]\,, \text{ and} \\
\boldsymbol{\theta}[\mathcal{J}] &= \begin{bmatrix} \boldsymbol{\beta}_1[1:2] \\ \boldsymbol{\beta}_2[1:2] \end{bmatrix}\,.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathcal{K}_1 &= (2(2-1)+2(3-1)+(1-1)2)\mathbf{1} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}\,, \\
\mathcal{K}_2 &= (2(2-1)+2(3-1)+(2-1)2)\mathbf{1} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 10 \end{bmatrix}\,, \text{ and} \\
\mathcal{K} &= \begin{bmatrix} 7 \\ \vdots \\ 10 \end{bmatrix}\,. \\
\text{So } \boldsymbol{\theta}[\mathcal{K}_1] &= \boldsymbol{\theta}[7:8] = \boldsymbol{\gamma}_1\,, \\
\boldsymbol{\theta}[\mathcal{K}_2] &= \boldsymbol{\theta}[9:10] = \boldsymbol{\gamma}_2\,, \text{ and} \\
\boldsymbol{\theta}[\mathcal{K}] &= \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix}\,.
\end{aligned}
$$

Again, it is easy to see that the above are consistent with Equation (2.11).

Now define $\boldsymbol{\xi}$ to be a three-way array whose $[i,j,k]$ entry is the position of $\boldsymbol{\mu}[i,j,k]$ in $\boldsymbol{\eta} = vec(\boldsymbol{\mu}^T)$, so that

$$\boldsymbol{\xi}[i,j,k] = (i-1)JK + (j-1)K + k. \tag{3.12}$$

The nonzero velocities can be written as

$$\boldsymbol{V}\left[vec(\boldsymbol{\xi}[i,,])^T, \mathcal{I}_f[i]\right] = \boldsymbol{\beta}_f \otimes \boldsymbol{\gamma}_f,\ i = 1,\ldots,I-1, \tag{3.13}$$

$$\boldsymbol{V}\left[vec(\boldsymbol{\xi}[,j,])^T, \mathcal{J}_f[j]\right] = \boldsymbol{\alpha}_f \otimes \boldsymbol{\gamma}_f,\ j = 1,\ldots,J-1, \tag{3.14}$$

$$\text{and } \boldsymbol{V}\left[vec(\boldsymbol{\xi}[,,k])^T, \mathcal{K}_f[k]\right] = \boldsymbol{\alpha}_f \otimes \boldsymbol{\beta}_f,\ k = 1,\ldots,K. \tag{3.15}$$

In Model (2.10) and in the plastocyanin example, consider $\boldsymbol{\mu}[2,2,1] = 1 \cdot \boldsymbol{\beta}[2]\boldsymbol{\gamma}[1]$, which is the ninth entry in $\boldsymbol{\eta}(\boldsymbol{\theta})$. The formula (3.12) should therefore yield a value of 9, which it does:

$$\boldsymbol{\xi}[2,2,1] = (2-1)(3)(2) + (2-1)(2) + 1 = 9.$$

It can be verified that (3.12) holds for other values of i, j, and k, as well. Since $F = 1$, Equations (3.9), (3.10), and (3.11) give

$$\begin{aligned} \mathcal{I} &= 1, \\ \mathcal{J} &= \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \text{ and} \\ \mathcal{K} &= \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \end{aligned}$$

which correspond to the correct columns of the derivative matrix (3.2). For-

mula (3.13) is worked out as follows:

$$vec(\boldsymbol{\xi}[1,,])^T = \begin{bmatrix} \boldsymbol{\xi}[1,1,1] \\ \boldsymbol{\xi}[1,1,2] \\ \boldsymbol{\xi}[1,2,1] \\ \boldsymbol{\xi}[1,1,2] \\ \boldsymbol{\xi}[1,3,1] \\ \boldsymbol{\xi}[1,3,2] \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}.$$

$$\text{So } \boldsymbol{V}\left[vec(\boldsymbol{\xi}[1,,])^T, \mathcal{I}[1]\right] = \boldsymbol{V}[1:6,1] = \boldsymbol{\beta} \otimes \boldsymbol{\gamma},$$

which is exactly the first column of $\boldsymbol{V}$ in (3.2). Formulas (3.14) and (3.15) are similarly worked out below.

$$vec(\boldsymbol{\xi}[,1,])^T) = \begin{bmatrix} \boldsymbol{\xi}[1,1,1] \\ \boldsymbol{\xi}[1,1,2] \\ \boldsymbol{\xi}[2,1,1] \\ \boldsymbol{\xi}[2,1,2] \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 7 \\ 8 \end{bmatrix}, \quad vec(\boldsymbol{\xi}[,2,])^T) = \begin{bmatrix} \boldsymbol{\xi}[1,2,1] \\ \boldsymbol{\xi}[1,2,2] \\ \boldsymbol{\xi}[2,2,1] \\ \boldsymbol{\xi}[2,2,2] \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 9 \\ 10 \end{bmatrix};$$

$$vec(\boldsymbol{\xi}[,,1])^T) = \begin{bmatrix} \boldsymbol{\xi}[1,1,1] \\ \boldsymbol{\xi}[1,2,1] \\ \boldsymbol{\xi}[1,3,1] \\ \boldsymbol{\xi}[2,1,1] \\ \boldsymbol{\xi}[2,2,1] \\ \boldsymbol{\xi}[2,3,1] \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \\ 11 \end{bmatrix}, \quad vec(\boldsymbol{\xi}[,,2])^T) = \begin{bmatrix} \boldsymbol{\xi}[1,1,2] \\ \boldsymbol{\xi}[1,2,2] \\ \boldsymbol{\xi}[1,3,2] \\ \boldsymbol{\xi}[2,1,2] \\ \boldsymbol{\xi}[2,2,2] \\ \boldsymbol{\xi}[2,3,2] \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}.$$

We then have

$$\begin{aligned}
\boldsymbol{V}\left[vec(\boldsymbol{\xi}[,1,])^T, \mathcal{J}[1]\right] &= \boldsymbol{V}[(1,2,7,8),1] &&= \boldsymbol{\alpha} \otimes \boldsymbol{\gamma}, \\
\boldsymbol{V}\left[vec(\boldsymbol{\xi}[,2,])^T, \mathcal{J}[2]\right] &= \boldsymbol{V}[(3,4,9,10),2] &&= \boldsymbol{\alpha} \otimes \boldsymbol{\gamma}, \\
\boldsymbol{V}\left[vec(\boldsymbol{\xi}[,,1])^T, \mathcal{K}[1]\right] &= \boldsymbol{V}[(1,3,5,7,9,11),1] &&= \boldsymbol{\alpha} \otimes \boldsymbol{\beta}, \text{ and} \\
\boldsymbol{V}\left[vec(\boldsymbol{\xi}[,,2])^T, \mathcal{K}[2]\right] &= \boldsymbol{V}[(2,4,6,8,10),2] &&= \boldsymbol{\alpha} \otimes \boldsymbol{\beta},
\end{aligned}$$

where the S notation $\boldsymbol{V}[(r_1, r_2, \ldots, r_R), c]$ denotes the elements in rows $r_1, r_2, \ldots, r_R$, and column c of the matrix $\boldsymbol{V}$. (See Appendix A.) Inspection of $\boldsymbol{V}$ in (3.2) shows that the above are a correct representation of the columns of $\boldsymbol{V}$.

We see that, owing to PARAFAC models' conditional linearity, the partial derivative matrix $\boldsymbol{V}$ can be "easily" obtained, relatively speaking, from the least squares

parameter estimates α, β, and γ for bilinear and trilinear models. We expect that for higher-way models, this will also be the case.

3.1.2 Acceleration Vectors

Again, the special structure of PARAFAC models enables us to find patterns in the $\ddot{\eta}_{pq}$'s. Using the notation (3.9), (3.10), and (3.11) of the previous section, $\ddot{\eta}[,\mathcal{I}_f,\mathcal{J}_f]$ is the three-way array of order $IJK \times (I-1)F \times (J-1)F$ whose columns are mixed partial derivatives of η with respect to α_f^* and β_f^*. Similarly, the columns of $\ddot{\eta}[,\mathcal{J}_f,\mathcal{K}_f]$ are the mixed partial derivatives of η with respect to β_f^* and γ_f, and the columns of $\ddot{\eta}[,\mathcal{I}_f,\mathcal{K}_f]$, those with respect to α_f^* and γ_f.

We now derive expressions for the acceleration vectors:

1. All second derivatives with respect to a single element of θ are 0 since μ is a linear function of every individual element of θ.

$$\ddot{\eta}[,\mathcal{I}_f[i],\mathcal{I}_f[i]] = \ddot{\eta}[,\mathcal{J}_f[j],\mathcal{J}_f[j]] = \ddot{\eta}[,\mathcal{K}_f[k],\mathcal{K}_f[k]] = \mathbf{0}.$$

2. Because μ is conditionally linear in A given B and C, in B given A and C, and in C given A and B,

$$\ddot{\eta}[,\mathcal{I},\mathcal{I}] = \ddot{\eta}[,\mathcal{J},\mathcal{J}] = \ddot{\eta}[,\mathcal{K},\mathcal{K}] = \mathbf{0}.$$

3. For PARAFAC models the factors do not "cross," so

$$\ddot{\eta}[,\mathcal{I}_{f_1}[i],\mathcal{J}_{f_2}[j]] = \mathbf{0}, \quad \text{for } f_1 \neq f_2;$$

$$\ddot{\eta}[,\mathcal{J}_{f_2}[j],\mathcal{K}_{f_3}[k]] = \mathbf{0}, \quad \text{for } f_2 \neq f_3;$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}_{f_1}[i],\mathcal{K}_{f_3}[k]] = \mathbf{0}, \quad \text{for } f_1 \neq f_3.$$

4. For fixed f, the mixed partial derivatives with respect to parameters for two of the ways, say $\boldsymbol{\alpha}_f^*$ and $\boldsymbol{\beta}_f^*$, involve only the parameters for the third remaining way, $\boldsymbol{\gamma}_f$.

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}_f[i],\mathcal{J}_f[j]] = \begin{bmatrix} \overset{(i-1)JK\times 1}{\mathbf{0}} \\ \overset{(j-1)K\times 1}{\mathbf{0}} \\ \overset{K\times 1}{\boldsymbol{\gamma}_f} \\ \overset{(J-j)K\times 1}{\mathbf{0}} \\ \overset{(I-i)JK\times 1}{\mathbf{0}} \end{bmatrix}, \tag{3.16}$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}_f[i],\mathcal{K}_f[k]] = \begin{bmatrix} \overset{(i-1)JK\times 1}{\mathbf{0}} \\ \overset{J\times 1}{\boldsymbol{\beta}_f} \otimes \begin{bmatrix} \overset{(k-1)\times 1}{\mathbf{0}} \\ 1 \\ \overset{(K-k)\times 1}{\mathbf{0}} \end{bmatrix} \\ \overset{(I-i)JK\times 1}{\mathbf{0}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\beta}_f \otimes \boldsymbol{e}_k \\ \mathbf{0} \end{bmatrix}, \tag{3.17}$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{J}_f[j],\mathcal{K}_f[k]] = \overset{I\times 1}{\boldsymbol{\alpha}_f} \otimes \begin{bmatrix} \overset{(j-1)K\times 1}{\mathbf{0}} \\ \begin{bmatrix} \overset{(k-1)\times 1}{\mathbf{0}} \\ 1 \\ \overset{(K-k)\times 1}{\mathbf{0}} \end{bmatrix} \\ \overset{(J-j)K\times 1}{\mathbf{0}} \end{bmatrix} = \boldsymbol{\alpha}_f \otimes \begin{bmatrix} \mathbf{0} \\ \boldsymbol{e}_k \\ \mathbf{0} \end{bmatrix}. \tag{3.18}$$

For Example 2 of Section 2.3.2 and the plastocyanin example, the nonzero ac-

celeration vectors are given by

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{J}[1]=\begin{bmatrix}\boldsymbol{\gamma}[1]\\ \boldsymbol{\gamma}[2]\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix},\quad \ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{J}[2]=\begin{bmatrix}0\\ 0\\ \boldsymbol{\gamma}[1]\\ \boldsymbol{\gamma}[2]\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix}; \tag{3.19}$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{K}[1]=\begin{bmatrix}\boldsymbol{\beta}[1]\\ 0\\ \boldsymbol{\beta}[2]\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix},\quad \ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{K}[2]=\begin{bmatrix}0\\ \boldsymbol{\beta}[1]\\ 0\\ \boldsymbol{\beta}[2]\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix}; \tag{3.20}$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{J}[1],\mathcal{K}[1]=\begin{bmatrix}\boldsymbol{\alpha}[1]\\ 0\\ 0\\ 0\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\end{bmatrix},\quad \ddot{\boldsymbol{\eta}}[,\mathcal{J}[1],\mathcal{K}[2]=\begin{bmatrix}0\\ \boldsymbol{\alpha}[1]\\ 0\\ 0\\ 0\\ 0\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\end{bmatrix}, \tag{3.21}$$

$$\ddot{\eta}[,\mathcal{J}[2],\mathcal{K}[1]=\begin{bmatrix}0\\0\\\alpha[1]\\0\\0\\0\\0\\1\\0\\0\\0\end{bmatrix},\quad \ddot{\eta}[,\mathcal{J}[2],\mathcal{K}[2]=\begin{bmatrix}0\\0\\0\\\alpha[1]\\0\\0\\0\\0\\0\\1\\0\\0\end{bmatrix}. \tag{3.22}$$

The accelerations (3.19), (3.22), and (3.20) were obtained by actually taking partial derivatives. It can be seen by closer inspection that the same vectors are obtained if formulas (3.16), (3.18), and (3.17) are applied to the plastocyanin example.

Using the notation of Equation (3.12), the nonzero entries of $\ddot{\eta}$ can be written in more compact form as

$$\ddot{\eta}\,[\xi[i,j,],\mathcal{I}[i],\mathcal{J}[j]] \;=\; \boldsymbol{C},\;\; i=1,\ldots,I-1,\; j=1,\ldots J-1, \tag{3.23}$$

$$\ddot{\eta}\,[\xi[i,,k],\mathcal{I}[i],\mathcal{K}[k]] \;=\; \boldsymbol{B},\;\; i=1,\ldots,\; I-1, k=1,\ldots K, \tag{3.24}$$

$$\ddot{\eta}\,[\xi[,j,k],\mathcal{J}[j],\mathcal{K}[k]] \;=\; \boldsymbol{A},\;\; j=1,\ldots,\; J-1, k=1,\ldots K. \tag{3.25}$$

An illustration for each of the formulas above is given below.

$$\xi[1,2,1]=3,\;\; \xi[1,2,2]=4,\;\; \text{so, from (3.19),}$$

$$\ddot{\eta}\,[\xi[1,2,],\mathcal{I}[1],\mathcal{J}[2]]=\ddot{\eta}\,[(3,4),\mathcal{I}[1],\mathcal{J}[2]]=\begin{bmatrix}\gamma[1]\\\gamma[2]\end{bmatrix}=\gamma\,.$$

$$\xi[1,1,2]=2,\;\; \xi[1,2,2]=4,\;\; \xi[1,3,2]=6,\;\; \text{so, from (3.20),}$$

$$\ddot{\eta}\,[\xi[1,,2],\mathcal{I}[1],\mathcal{K}[2]]=\ddot{\eta}\,[(2,4,6),\mathcal{I}[1],\mathcal{K}[2]]=\begin{bmatrix}\beta[1]\\\beta[2]\\1\end{bmatrix}=\beta\,.$$

$$\boldsymbol{\xi}[1,2,2] = 4, \quad \boldsymbol{\xi}[2,2,2] = 10, \quad \text{so, from (3.19),}$$

$$\ddot{\boldsymbol{\eta}}\,[\boldsymbol{\xi}[,2,2], \mathcal{J}[2], \mathcal{K}[2]] = \ddot{\boldsymbol{\eta}}\,[(4,10), \mathcal{J}[2], \mathcal{K}[2]] = \begin{bmatrix} \boldsymbol{\alpha}[1] \\ 1 \end{bmatrix} = \boldsymbol{\alpha}\,.$$

For a one-factor model, $\boldsymbol{C}$, $\boldsymbol{B}$, and $\boldsymbol{A}$ are just $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$, so that formulas (3.23), (3.24), and (3.25) hold for these three examples. It is easily verified that they hold for other values of i, j, and k, too, as well as for the case when $F > 1$.

3.2 Curvature Measures in Nonlinear Models

We now describe how the curvature measures of Bates and Watts are obtained, with a view to computing these measures for PARAFAC models.

Recall our notation for velocity and acceleration vectors:

$$\overset{N\times P}{\boldsymbol{V}} = \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^T} = (\dot{\boldsymbol{\eta}}_p) = (\boldsymbol{V}_p),\ p = 1,..,P,$$

where the $\boldsymbol{V}_p$ are called the tangent or velocity vectors, and $\overset{N\times P\times P}{\ddot{\boldsymbol{\eta}}} = \frac{\partial^2 \boldsymbol{\eta}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$ is the array whose n^{th} face is $\overset{P\times P}{\ddot{\boldsymbol{\eta}}_n} = (\ddot{\boldsymbol{\eta}}_{npq})$, $p, q = 1, \ldots, P$, where

$$\ddot{\boldsymbol{\eta}}_{npq} = \frac{\partial \dot{\boldsymbol{\eta}}_p}{\partial \boldsymbol{\theta}_q} = \frac{\partial \boldsymbol{V}_p}{\partial \boldsymbol{\theta}_q}$$

are the acceleration vectors.

The dimension of the tangent space is P, and the combined dimension of the tangent and acceleration spaces is, say, $(P + P')$. So P' is the dimension of the space orthogonal to the tangent space but spanned by the acceleration vectors. (The maximum possible dimension is $P + P(P+1)/2 = P(P+3)/2$.)

Let $\overset{N\times\frac{P(P+1)}{2}}{\ddot{W}}$ be the matrix of distinct acceleration vectors, and define

$$\overset{N\times\frac{P(P+3)}{2}}{D} = \left[\overset{N\times P}{V}, \overset{N\times\frac{P(P+1)}{2}}{\ddot{W}}\right] \tag{3.26}$$

If we perform a QR decomposition (See Dongarra, 1979) on D:

$$D = \left[\overset{N\times P}{Q_1} \mid \overset{N\times P'}{Q_1'} \mid \overset{N\times(N-P-P')}{Q_2}\right] \overset{N\times\frac{P(P+3)}{2}}{R}$$

the *acceleration array* is given by

$$\overset{(P+P')\times P\times P}{\ddot{A}} = \left[(Q_1|Q_1')^T\right][\ddot{\eta}], \tag{3.27}$$

where the brackets operator [ ][ ] denotes the matrix-array multiplication defined in Appendix B. The array $\ddot{A}$ has P faces in the tangent space and P' faces in the acceleration space: The first P faces of $\ddot{A}$, A_θ, is called the *parameter effects acceleration array*, which measures the nonuniformity of parameter lines. The last P' faces of $\ddot{A}$, A_a, called the *intrinsic acceleration array*, measures nonplanarity or intrinsic nonlinearity.

The accelerations are dependent on scaling of data and of parameters. Thus, measures of relative curvature are necessary. From the QR decomposition on D, we get

$$R_1 = \left[(Q_1|Q_1')^T\right][D] = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix},$$

where $\overset{P\times P}{R_{11}}$ is the R matrix from the QR decomposition of V.

The *relative curvature array* is given by

$$\overset{(P+P')\times P\times P}{B} = (R_{11}^T)^{-1}\ddot{A}R_{11}^{-1}s\sqrt{P}, \tag{3.28}$$

where $s = \sqrt{\frac{S(\hat{\theta})}{N-P}}$. The first P faces of $\boldsymbol{B}$, $\overset{P\times P\times P}{\boldsymbol{B}_\theta}$, is the *parameter effects relative curvature array.* The last P' faces of $\boldsymbol{B}$, $\overset{P'\times P\times P}{\boldsymbol{B}_a}$, is the *intrinsic relative curvature array* introduced in Section 2.3.1. To measure nonplanarity and nonuniformity, we look at how large (i.e. far from 0) terms in $\boldsymbol{B}_\theta$ and $\boldsymbol{B}_a$ are.

3.3 Linear Dependencies

To find measures of relative curvature for the PARAFAC models, we need to know $(P + P')$, the combined dimension of the tangent and acceleration spaces, or simply P' (since we know what P is). We, therefore, need to determine whether any linear dependencies exist among the tangent and acceleration vectors, and if so, how many. Although $(P + P')$ can always be obtained numerically by finding the rank of the matrix of combined tangent and acceleration vectors, it is hoped that this value can be found in closed form for some, if not all, of the PARAFAC models, owing to their special structure.

In the sections below, we shall give the structures for a one-factor trilinear model. Linear dependencies will be investigated (1) within $\ddot{\boldsymbol{\eta}}$; (2) pairwise, between columns of $\boldsymbol{V}$ and of $\ddot{\boldsymbol{\eta}}$; and (3) mutually, among the columns of $\boldsymbol{V}$ and $\ddot{\boldsymbol{\eta}}$. What is known, so far, for the general F-factor model will be stated.

3.3.1 Independence Among Acceleration Vectors

For a trilinear model, we wish to determine whether the columns of $\ddot{\boldsymbol{\eta}}$ have any linear dependencies, i.e., if there exist nonzero coefficients q_{ijf}, r_{jkf} and s_{ikf} such

that

$$\sum_{i,j,f} q_{ijf}\ddot{\eta}[,\mathcal{I}_f[i],\mathcal{J}_f[j]] + \sum_{j,k,f} r_{jkf}\ddot{\eta}[,\mathcal{J}_f[j],\mathcal{K}_f[k]] + \sum_{i,k,f} s_{ikf}\ddot{\eta}[,\mathcal{I}_f[i],\mathcal{K}_f[k]] = \overset{IJK\times 1}{\mathbf{0}} \tag{3.29}$$

Lemma III.1 *For $F = 1$, Equation 3.29 implies that $q_{ijf} = r_{jkf} = s_{ikf} = 0$. For $F > 1$, we have the constraints*

$$\sum_{f=1}^{F}\{q_{i'jf}\gamma_f[k] + r_{jkf}\alpha_f[i'] + s_{i'kf}\beta_f[j]\}$$

$$\textit{for } i' = 1,\ldots,I-1,\ j = 1,\ldots,J-1,\ k = 1,\ldots,K.$$

Thus, for a one-factor model, the columns of $\ddot{\eta}$ are linearly independent. The lemma, of course, implies that linear independence fails when $F > 1$. For an F-factor model, there are

$$(I-1)(J-1)(F-1) + (J-1)K(F-1) + (I-1)K(F-1)$$

constraints. The proof is given in Appendix C.

For our example of a 2 by 3 by 2 one-factor model, the columns of $\ddot{\eta}$ are given by formulas (3.19), (3.20), and (3.22). Observe that they are indeed linearly independent.

3.3.2 Pairwise Independence Between Tangent and Acceleration Vectors

We wish to know whether any linear dependencies exist between the columns of V and those of $\ddot{\eta}$ for a trilinear model. For each of the three submatrices of

$\boldsymbol{V} = (\boldsymbol{V}_A, \boldsymbol{V}_B, \boldsymbol{V}_C)$, we determine whether its columns are linearly independent of $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$, of $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]$, and of $\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]]$.

Lemma III.2 *For* $F = 1$,

1. $\boldsymbol{V}_A[,i]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$ *are linearly independent.*
2. $\boldsymbol{V}_A[,i]$ *is a linear combination of* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]$.
3. $\boldsymbol{V}_A[,i]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]]$ *are linearly independent.*
4. $\boldsymbol{V}_B[,j]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$ *are linearly independent.*
5. $\boldsymbol{V}_B[,j]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]$ *are linearly independent.*
6. $\boldsymbol{V}_B[,j]$ *is a linear combination of* $\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]]$.
7. $\boldsymbol{V}_C[,k]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$ *are linearly independent.*
8. $\boldsymbol{V}_C[,k]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]$ *are linearly independent.*
9. $\boldsymbol{V}_C[,k]$ *and* $\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]]$ *are linearly independent.*

Thus, for a one-factor model, columns of blocks of $\ddot{\boldsymbol{\eta}}$ are not linearly independent of columns of blocks of $\boldsymbol{V}$. See Appendix C for the proof.

To illustrate this lemma, compare the columns of the derivative matrix (3.2) with those of $\ddot{\boldsymbol{\eta}}$ given by (3.19), (3.20), and (3.22). Observe the following relationships :

$$\boldsymbol{V}_A[,1] = \boldsymbol{\gamma}[1]\ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{K}[1]] + \boldsymbol{\gamma}[2]\ddot{\boldsymbol{\eta}}[,\mathcal{I}[1],\mathcal{K}[2]]$$

$$\boldsymbol{V}_B[,1] = \boldsymbol{\gamma}[1]\ddot{\boldsymbol{\eta}}[,\mathcal{J}[1],\mathcal{K}[1]] + \boldsymbol{\gamma}[2]\ddot{\boldsymbol{\eta}}[,\mathcal{J}[1],\mathcal{K}[2]]$$

$$\boldsymbol{V}_B[,2] = \boldsymbol{\gamma}[1]\ddot{\boldsymbol{\eta}}[,\mathcal{J}[2],\mathcal{K}[1]] + \boldsymbol{\gamma}[2]\ddot{\boldsymbol{\eta}}[,\mathcal{J}[2],\mathcal{K}[2]]$$

The above equations correspond to parts (2) and (6) of Lemma III.2. Note also that no other dependencies exist besides these, just as the rest of Lemma III.2 states.

3.3.3 Mutual Independence Among Tangent and Acceleration Vectors

We wish to determine if there exist nonzero coefficients a_{if}, b_{jf}, c_{kf}, q_{ijf}, r_{jkf} and s_{ikf} such that

$$\begin{aligned} \overset{IJK\times 1}{\mathbf{0}} = \sum_{if} a_{if} \boldsymbol{V}_{Af}[,i] \;+\; \sum_{jf} b_{jf} \boldsymbol{V}_{Bf}[,j] \\ +\sum_{kf} c_{kf} \boldsymbol{V}_{Cf}[,k] \;+\; \sum_{ijf} q_{ijf} \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}_f[i],\mathcal{J}_f[j]] \\ +\sum_{jkf} r_{jkf} \ddot{\boldsymbol{\eta}}\,[,\mathcal{J}_f[j],\mathcal{K}_f[k]] \;+\; \sum_{ikf} s_{ikf} \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}_f[i],\mathcal{K}_f[k]] \end{aligned} \tag{3.30}$$

Lemma III.3 *For $F = 1$, Equation 3.30 implies that*

1. $c_k = 0$ *for* $k = 1, \ldots, K$
2. $q_{ij} = 0$ *for* $i = 1, \ldots, I-1$, $j = 1, \ldots, J-1$
3. $a_i \boldsymbol{\gamma} = -\sum_{k=1}^{K} s_{ik} \boldsymbol{e}_k$ *for* $i = 1, \ldots, I-1$
4. $b_j \boldsymbol{\gamma} = -\sum_{k=1}^{K} r_{jk} \boldsymbol{e}_k$ *for* $j = 1, \ldots, J-1$

Part (1) of this lemma says that the first derivative with respect to the $\boldsymbol{\gamma}[k]$ is independent of everything else; so are the second derivatives with respect to $\boldsymbol{\alpha}[i]$ and $\boldsymbol{\beta}[j]$, according to part (2). Parts (3) and (4) say that the linear dependencies that exist are between the first derivative with respect to $\boldsymbol{\alpha}[i]$ and the second derivative with respect to $\boldsymbol{\alpha}[i]$ and $\boldsymbol{\gamma}[k]$, and between the first derivative with respect to $\boldsymbol{\beta}[j]$ and the second derivative with respect to $\boldsymbol{\beta}[j]$ and $\boldsymbol{\gamma}[k]$. Thus, there exist nontrivial

solutions to Equation 3.30, given by parts (3) and (4). These define $I-1+J-1$ dependencies among the velocities and accelerations. Note that the above results agree with those of the previous two sections. The proof is given in Appendix C.

For the 2 by 3 by 2 one-factor example, it was seen in the previous section that the only linear relationships which exist are those between $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]$, and between $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]]$, and that there $3 = 2-1+3-1$ such dependencies, just as Lemma III.3 indicates.

Now that the number of linear dependencies among the tangent and acceleration vectors has been determined, the value of P', the number of acceleration vectors which are not in the tangent plane, can be obtained, and the relative curvature measures of Section 3.2 calculated.

3.4 Relative Curvatures: Some Examples

In this section, the curvature measures of Section 3.2 and the results of Section 3.3 are illustrated for the one-factor trilinear model.

For the model (2.2), there are $P = I+J+K-2$ velocity vectors. Section 3.1.2 shows there are

$$w = (I-1)(J-1)+(I-1)K+(J-1)K < \frac{P(P+1)}{2}$$

nonredundant, nonzero acceleration vectors. (The zero vectors shall be excluded from the matrix $\ddot{\boldsymbol{W}}$ defined in Section 3.2 since they will eventually be consigned to the part of $\boldsymbol{Q}$ in the QR decomposition of $\boldsymbol{D}$ which is not used in getting the curvatures.) For multilinear models in general, w will be much less than $P(P+1)/2$,

which is the maximum value for nonlinear models, because of the number of zero acceleration vectors.

The matrix $\boldsymbol{D}$ in Equation (3.26) will have

$$P + w = IJ + IK + JK - K - 1 < \frac{P(P+3)}{2}$$

columns. To partition $\boldsymbol{Q}$ in the QR decomposition of $\boldsymbol{D}$, we need to know the value of P'. From Lemma III.1, we know that the columns of $\ddot{\boldsymbol{W}}$ are linearly independent, so that the rank of $\ddot{\boldsymbol{W}}$ is w. Lemmas III.2 and III.3 tell us that there are $I + J - 2$ dependencies among the velocities and accelerations. Thus, $\boldsymbol{D}$ will have rank

$$rank(\boldsymbol{D}) = (IJ+IK+JK-K-1)-(I+J-2) = IJ+IK+JK-I-J-K+1, \tag{3.31}$$

and

$$P' = rank(\boldsymbol{D}) - rank(\boldsymbol{V}) = IJ + IK + JK - 2(I + J + K) + 3 \tag{3.32}$$

since $\boldsymbol{V}$ is of rank P.

For the 2 by 3 by 2 one-factor model, $w = (2-1)(3-1)+(2-1)2+(3-1)(2) = 8$, using the formula for w given above. It was seen in Section 3.1.2 that the number of nonredundant, nonzero acceleration vectors is indeed 8. The matrix $\boldsymbol{D}$ has $P + w = (2)(3) + (2)(2) + (3)(2) - 2 - 1 = 13$ columns consisting of the 5 tangent vectors and the 8 acceleration vectors. From (3.31) above, the rank of $\boldsymbol{D}$ should be

$$rank(\boldsymbol{D}) = (2)(3) + (2)(2) + (3)(2) - 2 - 3 - 2 + 1 = 10 = 13 - 3,$$

which is the number of columns of $\boldsymbol{D}$ minus the number of linear dependencies.

Thus,

$$P' = 10 - 5 = (2)(3) + (2)(2) + (3)(2) - 2(2 + 3 + 2) + 3 = 5,$$

using formula (3.32) above. So $\ddot{\boldsymbol{A}}$ and $\boldsymbol{B}$ will be $10 \times 5 \times 5$ arrays, with 5 faces in the tangent space and 5 faces in the acceleration space.

For the plastocyanin data, $\boldsymbol{D}$ is the matrix whose first 5 columns are exactly $\boldsymbol{V}$ in (3.4) and whose other 8 columns are

$$\ddot{\boldsymbol{W}} = \begin{bmatrix} 4.2299 & 0 & 1.0469 & 0 & 0 & 0 & 0.9689 & 0 \\ 3.6574 & 0 & 0 & 1.0469 & 0 & 0 & 0 & 0.9689 \\ 0 & 4.2299 & 0 & 0 & 1.0469 & 0 & 0.9700 & 0 \\ 0 & 3.6574 & 0 & 0 & 0 & 1.0469 & 0 & 0.9700 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3.33}$$

A numerical check on the matrix $\boldsymbol{D}$ confirmed that its rank is indeed 10. The $10 \times 5 \times 5$ relative curvature array $\boldsymbol{B}$ was calculated using the procedure described in Section 3.2. The value of $S(\hat{\boldsymbol{\theta}})$ was 0.3476, so s=0.2228. Summary statistics for the first 5 faces of $\boldsymbol{B}$, which measure parameter effects curvature, and for the other 5 faces, which measure intrinsic curvature, are presented in Table 1. The values of the intrinsic relative curvatures are close to zero, indicating that the expectation surface is well-approximated by the tangent plane at $\hat{\boldsymbol{\theta}}$. The parameter effects relative curvatures are not too far from zero either, which means that projections of the parameter curves on the tangent plane would be more or less uniform.

Table 1: Relative Curvatures for Plastocyanin Data

	Minimum	Maximum	Mean	SD
Parameter Effects	-0.0680	0.1172	-0.0015	0.0245
Intrinsic	-0.0283	0.0253	-0.0001	0.0103

We now consider a second example with a much larger number of observations, using the NFAK1A data of Leurgans and Ross (1991a). The data is a 10 by 12 by 5 simulated arrray with 1 factor present. Parameters for a three-factor model were chosen to be reasonable values from a biophysical perspective. The parameters of the third factor of the three-factor model were then used as the parameters of the one-factor model. Independent normal random variables were generated with mean 0 and standard deviation 100 (1 % of the range of the expectation array). These were added to the expectation arrays to obtain simulated data.

For this data set, $I = 10$, $J = 12$, $K = 5$, P=25, $N = 600$, $w = 199$, and $\boldsymbol{D}$ has 224 columns. There are 20 linear dependencies so the rank of $\boldsymbol{D}$ is 204. Thus, $\boldsymbol{B}$ is a $204 \times 25 \times 25$ array with the first 25 faces containing the parameter effects relative curvatures, and the last $P' = 179$ faces containing the intrinsic relative curvatures.

Summary statistics for the relative curvatures of the NFAK1A data are given in Table 2. Note that the relative intrinsic curvatures are again small. This is not surprising, considering that in various examples given by Ratkowsky (1983), Bates and Watts (1980), and Bates and Watts (1988), the expectation surface was found to be nearly planar at $\hat{\boldsymbol{\eta}}$, while parameter-effects curvatures were found to be

Table 2: Relative Curvatures for NFAK1A Data

	Minimum	Maximum	Mean	SD
Parameter Effects	-0.1291	1.2930	0.0008	0.0177
Intrinsic	-0.0174	0.0174	0.0000	0.0010

substantial.

We conclude this section with some comments on the computations involved in getting the curvature arrays. The *qr* function in the *S* language (see Becker, et al.,1988), as in most other packages that do a QR decomposition, does not report the Q and R matrices explicitly, but returns an object representing the decomposition, from which Q and R have to be extracted. It is simple to get R; calculating Q directly from the given object is very inefficient, however. Instead, applying the decomposition twice yields $(\boldsymbol{Q}_1, \boldsymbol{Q}_1')$ rather easily. The columns of $\boldsymbol{D}$ are pivoted (see Dongarra, 1979) according to the result of the first decomposition; a second decomposition is done on the pivoted matrix; then one inversion and one multiplication of the appropriate matrices yield the desired $(\boldsymbol{Q}_1, \boldsymbol{Q}_1')$. Pivoting is more easily done interactively. It may also be necessary to specify a larger tolerance when using the *qr* procedure since this *S* function appears to be very sensitive to tiny values when detecting linear dependencies. It is good practice to check if the rank given by this function agrees with the rank of $\boldsymbol{D}$ given by the formula (3.31).

Due to the large number of observations for the NFAK1A data, operations such as multiplication of the 204×600 matrix $(\boldsymbol{Q}_1, \boldsymbol{Q}_1')^T$ and the $600 \times 25 \times 25$ array $\ddot{\boldsymbol{\eta}}$

to get the $204 \times 25 \times 25$ array $\ddot{\boldsymbol{A}}$ took a few seconds longer. Contructing one long function that calculates the curvatures in one big step was also avoided because dynamic memory problems were sometimes run into. This, of course, could just be a local problem. Thus, one may need to working interactively, or increase the allotted dynamic memory, if possible, when using S to compute curvature arrays for large data sets.

CHAPTER IV

Introduction to Cross-Validation

Cross-validation is a technique mainly developed for problems involving model selection or assessment of the performance of a predictor. Like the jackknife and the bootstrap, this nonparametric method requires minimal assumptions and can be applied in an automatic way to complicated situations. Although cross-validation and jackknifing both involve omission of items one or more at a time, Stone (1974) distinguishes between the two by noting that jackknifing manufactures pseudovalues for the reduction of bias. Efron and Gong (1983) give several examples of using the bootstrap, the jackknife, and cross-validation in the estimation of statistical error (i.e., bias, standard error of an estimator, or error rate of a prediction rule).

In Section 4.1, two very general cross-validation methods are discussed. Section 4.2 deals with cross-validation techniques that have been used for multilinear models, or in settings similar to the multilinear one. The final section discusses how case deletion in multilinear models is also useful for cross-validation purposes.

4.1 Some Cross-Validation Techniques

Stone (1974) gives various descriptions and applications of cross-validation.

4.1.1 Leave-out Groups

The concept of cross-validation, in its most primitive form, involves dividing the sample into two subsamples, one on which the choice of a statistical model, predictor, or estimator is based, and the other on which performance of the chosen model, predictor, or estimator is assessed by measuring its predictions against the actual observed values. Division of the sample may be controlled, as when items are randomly assigned to one or the other subsample, or uncontrolled, e.g., in studies where a "construction sample" and a "validation sample" are collected at separate times.

This idea of dividing the sample into two groups and leaving out one group while calculations are done on the other has been extended to dividing the sample into several groups and leaving out one group at a time. A more detailed discussion is given in Section 4.2 below.

4.1.2 Leave-out One Observation

Mosteller and Tukey (1977) describe "simple cross-validation" as setting aside one item, optimizing for the rest of the observations, then testing on the set-aside item. Realizing that repetition of this process for every single item may be computationally difficult, they suggest that we make one optimization for all the data, then do a possibly simpler calculation of the effect of dropping each item, then compare the adjusted optimized result with the values for the omitted item.

4.2 Cross-Validation Methods in Multilinear Settings

In this section, some cross-validation techniques used in factor analysis and principal component analysis are given. These are relevant to cross-validation for multilinear models in that the problem of choosing the number of factors is common to all three settings.

Wold (1978) uses cross-validation to estimate the number of significant components, F, in Factor Analysis and Principal Components Analysis (PCA) of a set of data. The data is divided into G groups. For a starting value of F, say F_o , the first group is deleted and the model parameters are estimated using the reduced data set and some goodness-of-fit criterion (e.g., least squares). Predicted values of observations in the deleted group are then calculated. These are compared with the actual values via the sum of squares of prediction errors (also called PRESS, for predicted residual sum of squares), or some other measure corresponding to the goodness-of-fit criterion.

After the first group is restored to the data set, the second group of observations is deleted, parameters are estimated using the new reduced data set, and a second sum of squared prediction errors is calculated. Then the third group is deleted, and the procedure is repeated until all the groups have been deleted once. Summing all the (partial) sums of squares of prediction errors over all the groups gives a total PRESS for $F = F_o$, PRESS(F_o).

The whole procedure is repeated for different values of F, say $F_1, F2, \ldots$, yielding PRESS(F_1), PRESS(F_2), and so on. The value of F is varied systematically to find

the minimum value of PRESS(F). The value of F which gives the smallest total PRESS is the optimal choice for the number of components.

Wold's method is carried out using the NIPALS (nonlinear iterative partial least squares) algorithm, which has the advantage of converging very rapidly when good starting values are given, and of being able to work with incomplete data. It is, however, not as universally available as, say, algorithms using singular value decomposition (SVD) of a matrix.

Eastment and Krzanowski (1982) describe a method for choosing the number of components in a PCA using algorithms due to Bunch and Nielsen (1978) for updating a singular value decomposition:

Suppose that p variables are observed on each of n individuals and the data is displayed in an $n \times p$ matrix $\boldsymbol{Y}$ with SVD $\boldsymbol{Y} = USV^T$, where U is an $n \times n$ orthogonal matrix, V is a $p \times p$ orthogonal matrix, and S is an $n \times p$ upper diagonal matrix. Based on the cross-validation principle of not using each data point in both the prediction and assessment stages, but of nevertheless using as much of the original data as possible in predicting each Y_{ij}, Eastment and Krzanowski suggest that Y_{ij} should be predicted from all the data except the i^{th} row and the j^{th} columns of $\boldsymbol{Y}$.

If the i^{th} row is omitted, the SVD of the reduced data matrix will have left singular vectors with length $n-1$, and the required values of u_{it} for forming the predictor

$$\hat{Y}_{ij}^{(F)} = \sum_{t=1}^{F} u_{it} s_t v_{tj}$$

are not available. Denote the updated values of U, V, and S by $\hat{U}$, $\hat{V}$, and $\hat{S}$ when

the j^{th} column of $\mathbf{Y}$ is deleted, and by $\bar{U}$, $\bar{V}$, and $\bar{S}$ when the i^{th} row of $\mathbf{Y}$ is deleted. If the complete matrix is available,

$$\hat{Y}_{ij}^{(F)} = \sum_{t=1}^{F} u_{it} s_t v_{tj}.$$

Since u_{it} requires information on the i^{th} row, we use $\hat{U}$. Similarly, v_{tj} requires information on the j^{th} column, so use $\bar{V}$. The matrix S can come from either $\hat{S}$ or $\bar{S}$ so it is reasonable to use both. Hence, Y_{ij} is predicted by

$$\hat{Y}_{ij}^{(F)} = \sum_{t=1}^{F} (\hat{u}_{it}\sqrt{\hat{s}_t})(\sqrt{\bar{s}_t}\bar{v}_{tj}).$$

The SVD of the matrix $\mathbf{Y}$ can be found using any number of algorithms currently available. Then PRESS(F) can be obtained and the best value of F chosen based on a suitable function of PRESS(F). Eastment and Krzanowski use

$$W_F = \frac{PRESS(F-1) - PRESS(F)}{D_F} \div \frac{PRESS(F)}{D_R},$$

where D_F is the number of degrees of freedom required to fit the F^{th} component and D_R is the number of degrees of freedom remaining after fitting the F^{th} component. So W_F represents the ratio of the increase in predictive information supplied by the F^{th} component to the average information in each of the remaining components. The best choice for F is the last value of F at which $W_F > 1$. Krzanowski (1986) presents some simulation results for bilinear models.

Leurgans and Ross (1991a) extended the idea of deleting different groups and also Eastment and Krzanowski's technique of leaving out levels to three-way PARAFAC models. The simulated data sets NFAK described in the Section 3.4 were used.

They divided each of the 10 by 12 by 5 arrays into 5 groups and fit models with $F_o = 1, 2, 3$, deleting one group at a time. For arrays generated from a three- factor model, they also fit $F_o = 4$.

The extension of Eastment and Krzanowski's technique to three-way PARAFAC models involves deleting a level of one of the ways, say, level i of α, and estimating the parameters of the other two ways, say, β and γ. Thus, $\mu[i, j, k]$ can be estimated without using the actual value $\boldsymbol{Y}[i, j, k]$ by estimating α without $\boldsymbol{Y}[, j,]$ or $\boldsymbol{Y}[, , k]$, estimating β without $\boldsymbol{Y}[i, ,]$ or $\boldsymbol{Y}[, , k]$, and estimating γ without $\boldsymbol{Y}[i, ,]$ or $\boldsymbol{Y}[, j,]$. Leurgans and Ross have experimented with leaving out each level of one of the simulated arrays, and have seen a need to improve initialization methods to hasten calculations. To reduce computational requirements, they have also explored leaving out several levels of each way at a time: All $\boldsymbol{Y}[i, ,]$'s with i even, all $\boldsymbol{Y}[i, ,]$'s with i odd, all $\boldsymbol{Y}[, j,]$'s with j even, all $\boldsymbol{Y}[, j,]$'s with j odd, all $\boldsymbol{Y}[, , k]$'s with k even, and all $\boldsymbol{Y}[, , k]$'s with k odd. This even-odd deletion was performed on one simulated data set, NFAK2A.

Leurgans and Ross use alternating-least-squares algorithms whose initial values are based on the decomposition described by Leurgans, Ross and Abel (1990). The algorithms are allowed to cycle until convergence is achieved or after 10^4 iterations, whichever comes first. Some of the calculations use one-step approximations to the solutions, in which only one iteration is completed after the global fit. They present some preliminary results on simulated and real data sets.

4.3 Case Deletion in Cross-Validation of Multilinear Models

Cross-validation is clearly one area in which finding approximations to case deletion estimates is relevant. The general cross-validation approach of leaving out a single observation, and the technique for multilinear models of leaving out levels suggest case deletion as a natural way to cross-validate multilinear models. Such a method would indeed use as much of the original data as possible in predicting each Y[i,j,k].

As Mosteller and Tukey (1977) pointed out, however, deleting one observation at a time may prove computationally difficult. In Section 4.2, we saw that Leurgans and Ross (1991a) found it necessary to improve initialization methods and experimented with leaving out several levels at a time to reduce computational requirements. Bunch and Nielsen's (1978) algorithm which Eastment and Krzanowski (1982) used to facilitate their calculations for two-way models does not have a generalization for three-way models. Thus leaving out levels for trilinear models would still involve intensive computation. Leurgans and Ross also looked into the one-step approximation as a means to avoid much computational labor in leaving out groups and levels.

All the more difficulty can be expected if an actual leave-out-one-observation approach were used. However, if good approximations to the actual case deletion estimates can be obtained, and if these approximations require considerably less computational labor, they would be extremely useful for cross-validation purposes.

We shall look at approximations to case deletion estimates more closely in the

next two chapters. In Chapter VI, an approximation to the estimate of the r^{th} predicted or cross-validatory residual is given. Due to the special structure of multilinear models, the calculation of this quantity, and thus of the predicted residual sum of squares (PRESS), is facilitated.

CHAPTER V

Jackknifing in Nonlinear Models: A Review of Literature

5.1 Introduction

We noted in Chapter IV that jackknifing is similar to cross-validation in that it involves leaving out one observation at a time. (Although the two methods, used separately, yield different estimates (See Efron and Gong (1983) for illustrations), one may be used in conjunction with the other.) Thus, jackknifing is another area where case-deletion approximation methods may prove useful.

We discuss here various jackknife estimators for the nonlinear regression model, with the posssible use of one or more of these methods in the multilinear setting in view. In Chapter VI, three of these methods will be used to find parameter estimates for the NFAK1A data. All the estimators discussed in this chapter are modifications of the ordinary or standard jackknife (J) defined as follows:

Recall the nonlinear regression model (2.5),

$$Y_r = \eta(x_r, \boldsymbol{\theta}) + \epsilon_r, \quad r = 1, 2, ..., N,$$

where η is a nonlinear function of the unknown parameter (vector) $\boldsymbol{\theta}$, ϵ_r are independent, $\mathrm{Var}(\epsilon_r)=\sigma^2$, unknown.

Let $\hat{\boldsymbol{\theta}}_{(r)}$ be the least squares estimator of $\boldsymbol{\theta}$ when the r^{th} data point is removed. Define

$$\boldsymbol{P}_r = N\hat{\boldsymbol{\theta}} - (N-1)\hat{\boldsymbol{\theta}}_{(r)}$$

to be the r^{th} pseudovalue;

$$\hat{\boldsymbol{\theta}}_J = 1/N \sum \boldsymbol{P}_r$$

to be the (ordinary) jackknife estimator of $\boldsymbol{\theta}$;

$$S_J = \frac{1}{N(N-1)} \sum (\boldsymbol{P}_r - \hat{\boldsymbol{\theta}}_J)(\boldsymbol{P}_r - \hat{\boldsymbol{\theta}}_J)^T$$

to be the jackknife estimator of $\mathrm{Var}(\hat{\boldsymbol{\theta}}_J)$. In practice, S_J has also been used to estimate $\mathrm{Var}(\hat{\boldsymbol{\theta}})$.

The jackknife was originally suggested by Quenouille (1956) to reduce bias and is also useful for variance estimation.

5.2 The Linear Jackknife (LP)

Fox, Hinkley and Larntz (1980) point out that the ordinary jackknife requires $n+1$ fits in the nonlinear regression case: one for the initial fit $\hat{\boldsymbol{\theta}}$, and n fits for the $\hat{\boldsymbol{\theta}}_{(r)}$'s. They propose a method requiring only one nonlinear fit, based on a linear Taylor series approximation to $\hat{\boldsymbol{\theta}}_{(r)}$:

$$\hat{\boldsymbol{\theta}}_{(r)} \doteq \hat{\boldsymbol{\theta}} - \frac{(\hat{\boldsymbol{V}}^T\hat{\boldsymbol{V}})^{-1}\hat{\boldsymbol{V}}_r^T e_r}{1-h_r} \tag{5.1}$$

$$LP_r = \hat{\boldsymbol{\theta}} + \frac{N(\hat{\boldsymbol{V}}^T\hat{\boldsymbol{V}})^{-1}\hat{\boldsymbol{V}}_r^T e_r}{1-h_r}$$

$$\hat{\boldsymbol{\theta}}_{LP} = 1/N \sum LP_r \tag{5.2}$$

$$\hat{S}_{LP} = \frac{1}{N(N-1)} \sum (LP_r - \hat{\boldsymbol{\theta}}_{LP})(LP_r - \hat{\boldsymbol{\theta}}_{LP})^T \tag{5.3}$$

where

$\hat{V}$ is the matrix of partial derivatives of η evaluated at $\hat{\theta}$,

$\hat{V}_r$ is the r^{th} row of $\hat{V}$,

e_r is the r^{th} residual, and

h_r is the r^{th} diagonal element of the hat matrix $\hat{H} = \hat{V}(\hat{V}^T\hat{V})^{-1}\hat{V}^T$.

Note that the linear jackknife, LP, is analogous to the linear model setting where V=X, the hat matrix is $X(X^TX)^{-1}X^T$, and

$$\hat{\beta}_{(r)} = \hat{\beta} - \frac{(X^TX)^{-1}X_r^Te_r}{1-h_r}$$

In Chapter VI, it will be seen that the linear jackknife is equivalent to using the linear approximation to case deletion estimates given by Ross (1987).

5.3 The Weighted Linear Jackknife (LQ)

Miller (1974) showed that the usual bias reduction properties of the standard jackknife J do not hold for nonlinear functions of linear regression parameters. Hinkley (1977) suggested an alternative weighting scheme to improve bias reduction and variance estimation of the standard jacknife estimator J in linear regression. Fox, Hinkley and Larntz extended Hinkley's work to the nonlinear model by defining the weighted linear jackknife (LQ) as follows:

Let

$$\hat{\theta}_{(r)} \doteq \hat{\theta} - (\hat{V}^T\hat{V})^{-1}\hat{V}_r^Te_r,$$

$$LQ_r = \hat{\theta} + N(\hat{V}^T\hat{V})^{-1}\hat{V}_r^Te_r,$$

and $\hat{\boldsymbol{\theta}}_{LQ}, \hat{S}_{LQ}$ be defined by replacing LP by LQ in (5.2) and (5.3). Note that $\hat{\boldsymbol{\theta}}_{(r)}$ in LQ is just $\hat{\boldsymbol{\theta}}_{(r)}$ in LP with the second term multiplied by $(1-h_r)$, which, in effect, gives less weight to high leverage points.

Weighted linear jackknife estimates are calculated for the NFAK1A data in Chapter VI.

5.4 The Modified Linear Jackknife (MLP)

Although according to Fox, Hinkley and Larntz, $\hat{\boldsymbol{\theta}}_{LP}$ is not inferior to $\hat{\boldsymbol{\theta}}_J$, Simonoff and Tsai (1986) noted that $\hat{\boldsymbol{\theta}}_{LP}$ does not directly reflect the nonlinearity of the model. They remedy this by introducing a term involving the second partial derivative:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{(r)} &\doteq \hat{\boldsymbol{\theta}} - \frac{(\hat{\boldsymbol{V}}^T\hat{\boldsymbol{V}} - [\boldsymbol{e}_{(r)}^T][\hat{\hat{\boldsymbol{\eta}}}_{(r)}])^{-1}\hat{\boldsymbol{V}}_r^T e_r}{1-h_r^*} \\
&= \hat{\boldsymbol{\theta}} - \frac{\boldsymbol{T}_r^{-1}\hat{\boldsymbol{V}}_r^T e_r}{1-h_r^*} \\
MLP_r &= \hat{\boldsymbol{\theta}} + \frac{N\boldsymbol{T}_r^{-1}\hat{\boldsymbol{V}}_r^T e_r}{1-h_r^*}
\end{aligned}$$

and $\hat{\boldsymbol{\theta}}_{MLP}$, $\hat{S}_{MLP}$ are defined by replacing LP by MLP in (5.2) and (5.3). Here,

$\boldsymbol{e}_{(r)}$ is the $(N-1)\times 1$ vector with r^{th} component removed from the residuaal vector $\boldsymbol{e}$;

$\hat{\hat{\boldsymbol{\eta}}}_{(r)}$ is the $(N-1)\times P\times P$ array with r^{th} component removed from the $N\times P\times P$ array $\hat{\hat{\boldsymbol{\eta}}}$ of second partial derivatives;

the brackets operator [ ][ ] indicates vector-array multiplication as defined in Appendix B;

$T_r = \hat{V}^T\hat{V} - [e_{(r)}^T][\hat{\hat{\eta}}_{(r)}]$;

$h_r^* = \hat{V}_r^T T_r^{-1} \hat{V}_r^T$.

Note that if the term $[e_{(r)}][\hat{\hat{\eta}}_{(r)}]$ is ignored, $\hat{\theta}_{(r)}$ in the modified linear jackknife MLP is the same as in LP; that is, instead of using $\hat{V}^T\hat{V}$, MLP uses T_r.

5.5 The Reweighted Linear Jackknife (RLQ)

Simonoff and Tsai also propose an alternative weighting scheme to LQ such that the weight of the r^{th} observation is inversely proportional to its leverage. That is, reweight LQ by multiplying the second term of LQ_r by $(1 - h_r)$. Thus, we have

$$RLQ_r = \hat{\theta} + N(\hat{V}^T\hat{V})^{-1}\hat{V}_r^T e_r(1 - h_r).$$

In previous studies, $\hat{S}_{RLQ}$, as usually defined, consistently underestimated the true variance in previous simulation studies, so the authors replaced it by:

a. $\hat{S}_{RLQ} = \hat{\sigma}^2(\hat{V}^T\hat{V})^{-1}\sum \hat{V}_r(1 - h_r)^2\hat{V}_r^T(\hat{V}^T\hat{V})^{-1}$

b. $\hat{S}_{RLQL}$, the least squares estimator of variance

c. $\hat{S}_{MLP}$

all coupled with $\hat{\theta}_{RLQ} = 1/N \sum RLQ_r$, yielding the methods RLQ, RLQL, and RLQM, respectively.

In Chapter VI, reweighted linear jackknife estimates are calculated for the NFAK1A data. These are compared with the linear and weighted linear estimates for the same data.

5.6 The Weighted Modified Linear Jackknife (MLQ)

Simonoff and Tsai noted that the modification made by the MLP method to the LP method can also be applied to the weighted linear jackknife LQ; that is, in LQ, replace $\hat{V}^T\hat{V}$ by T_r, yielding

$$MLQ_r = \hat{\theta} + NT_r^{-1}\hat{V}_r^T e_r.$$

The estimators $\hat{\theta}_{MLQ}, \hat{S}_{MLQ}$ are defined by replacing LP by MLQ in (5.2) and (5.3).

5.7 Comparison of the Estimators

Simonoff and Tsai compared least squares (LS), J, LP, LQ, MLP, MLQ, and RLQ with respect to agreement of parameter and variance estimates in examples using different nonlinear models. They also compared LS, J, LP, LQ, MLP, MLQ, RLQ, RLQL, RLQM and direct likelihood (DL) with respect to bias, root mean squared error, coverage probabilities and volume of confidence regions, using simulation. The results can be summarized as follows:

For well-behaved data, there is little difference among the methods, but the presence of outliers, leverage points or curvature effects can heavily affect all the procedures. The RLQM method provides an estimator relatively robust to outliers, leverage points and curvature effects; its confidence regions are quite good compared to the others.

Fox, Hinkley and Larntz compared J, LP and LQ using the examples given by Duncan (1978). They found that the estimators in J and LP were in close agreement,

and LQ matched J and LP except at the extremes. In terms of coverage probabilities for confidence regions, LP and LQ were better than J, and were easier to use than J when variance estimates were needed.

CHAPTER VI

Leave-Out-One Approach

In the previous two chapters, we have seen two settings in which the idea of leaving out one observation at a time proves useful. The need for some form of approximation that avoids physically deleting a case when computing parameter estimates has been emphasized. We now turn to two responses to this need. These approximations have been suggested in the literature for obtaining influence measures based on case deletion. We have, in fact, already come across one of them in Chapter V. The "novelty" here stems from the application of these approximate leave-out estimates to multilinear models. The material from Chapter III facilitates this application.

Section 6.1 introduces the case deletion model for nonlinear regression, from which $\hat{\boldsymbol{\theta}}_{(r)}$, the estimate of the parameter $\boldsymbol{\theta}$ when the r^{th} case is deleted can be obtained. Linear and quadratic approximations to $\hat{\boldsymbol{\theta}}_{(r)}$ are given. The linear approximation is derived by Ross (1987) using the definitions and notation for nonlinear models given in Section 2.3.1. This approximation turns out to be the same as the expression for $\hat{\boldsymbol{\theta}}_{(r)}$ in Chapter V. The quadratic approximation is derived by Ross using Clarke's (1980) formula and assuming planarity of the expectation surface.

In Section 6.2, the approximations are applied to simulated data for a one-factor trilinear PARAFAC model. Since the approximation formulas involve the derivative

matrix $\boldsymbol{V}$, the array of second derivatives $\ddot{\boldsymbol{\eta}}$, and the relative curvature array $\boldsymbol{B}_a$, they can now be readily applied because of the results derived in Chapter III. The estimates obtained by using the two approximations are compared with the estimates obtained by actually leaving out one observation at a time. Some of the jackknife estimators in Chapter V are also evaluated for this data set. These are compared with the standard jackknife, which is obtained by actual case deletion.

6.1 Case Deletion in Nonlinear Regression

Ross (1987) examines case deletion in nonlinear regression models in the context of assessing influence. He notes that, unlike in linear regression, case deletion is not easy to do in nonlinear regression, since the entire nonlinear regression must be repeated every time an observation is deleted. Computation is simplified by using the linear approximation suggested by Fox, Hinkley and Larntz (1980), or by constructing a quadratic approximation adapted from Clarke (1980). Ross explores the relationship of the geometry of case deletion to these approximations.

Define the case deletion model

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{d}_r\zeta + \boldsymbol{\epsilon} \tag{6.1}$$

where $\boldsymbol{Y}$, $\boldsymbol{\eta}(\boldsymbol{\theta})$, and $\boldsymbol{\epsilon}$ are the same as in the nonlinear regression model (2.6), $\boldsymbol{d}_r$ is an $N \times 1$ vector having 1 in the r^{th} position and zeros elsewhere. Fitting (6.1) with $P+1$ parameters $(\boldsymbol{\theta}, \zeta)$ yields the least squares estimate $\hat{\boldsymbol{\theta}}_r$ for $\boldsymbol{\theta}$ when the r^{th} case is deleted. The estimate of ζ is the r^{th} predicted residual.

Denote the expected value of $\boldsymbol{Y}$ by

$$\boldsymbol{R}(\boldsymbol{\theta}, \zeta) = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{d}_r \zeta.$$

Then, using the notation of Section 2.3.1, the first and second partial derivatives of $\boldsymbol{R}(\boldsymbol{\theta}, \zeta)$ are:

$$\begin{aligned}
\boldsymbol{R}_p &= \dot{\boldsymbol{\eta}}_p = \boldsymbol{V}_p,\ 1 \le p \le P \\
\boldsymbol{R}_\zeta &= \boldsymbol{d}_r \\
\boldsymbol{R}_{pq} &= \ddot{\boldsymbol{\eta}}_{pq},\ 1 \le p, q \le P \\
\boldsymbol{R}_{p\zeta} &= \boldsymbol{R}_{\zeta\zeta} = \boldsymbol{0},\ 1 \le p \le P.
\end{aligned}$$

In the notation parallel to that of section 2.3.1, let:

$M_{(r)}$ be the expectation surface defined by $\boldsymbol{R}(\boldsymbol{\theta}, \zeta)$;

$\overset{N \times P+1}{\boldsymbol{V}_{(r)}} = [\boldsymbol{V}\ \boldsymbol{d}_r]$ be the matrix of first partial derivatives;

$\boldsymbol{G}_{(r)} = \boldsymbol{V}_{(r)}^T \boldsymbol{V}_{(r)}$.

$\boldsymbol{H} = \boldsymbol{V}\boldsymbol{G}^{-1}\boldsymbol{V}^T$.

Ross gives the following decomposition of $\boldsymbol{R}_{pq}$, which is analogous to (2.7):

$$\boldsymbol{R}_{pq} = \Gamma^s_{(r)pq} \boldsymbol{R}_s + \Lambda_{pq} \boldsymbol{d}_r + b^a_{pq} \boldsymbol{\phi}_a, \tag{6.2}$$

where $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_m$ is an orthonormal basis for the acceleration space of $M_{(r)}$ at $\boldsymbol{R}(\boldsymbol{\theta}, \zeta)$. The $P+1 \times P+1$ matrix of coefficients of the second fundamental form (see Section 2.3.1) of $M_{(r)}$ corresponding to the normal direction $\boldsymbol{\phi}_a$ and the $P \times P$ matrix of coefficients of the second fundamental form on M corresponding to the

normal direction ϕ_a are related by

$$\boldsymbol{B}_{(r)a} = \begin{bmatrix} \boldsymbol{B}_a & 0 \\ 0 & 0 \end{bmatrix}.$$

6.1.1 Linear Approximation

Let $\boldsymbol{C}_r = \boldsymbol{G}^{-1}\boldsymbol{V}^T\boldsymbol{d}_r$ and h_r be the r^{th} diagonal element of $\boldsymbol{H}$ (called the *leverage* of the r^{th} case) evaluated at $\hat{\boldsymbol{\eta}}$.. The first-order approximations to the estimates of $\hat{\boldsymbol{\theta}}_{(r)}$ and $\hat{\zeta}$ are obtained by using the same linear approximation used by Fox, Hinkley and Larntz(1980), that is, by replacing $\boldsymbol{R}(\boldsymbol{\theta}, \zeta)$ with a first-order Taylor series approximation about $(\hat{\boldsymbol{\theta}}, 0)$. This gives the estimates

$$\hat{\boldsymbol{\theta}}_{(r)} \doteq \hat{\boldsymbol{\theta}} + \boldsymbol{u}_{(r)} \tag{6.3}$$

$$\hat{\zeta} \doteq \frac{e_r}{1 - h_r} \tag{6.4}$$

$$\text{where } \boldsymbol{u}_{(r)} = -\frac{e_r}{1 - h_r}\boldsymbol{C}_r \tag{6.5}$$

Note that Equation (6.3) is just $\hat{\boldsymbol{\theta}}_{(r)}$ of Chapter V.

6.1.2 Quadratic Approximation

A quadratic approximation to $\hat{\boldsymbol{\theta}_{(r)}}$ and $\hat{\zeta}$ can be constructed by adapting a higher-order formula due to Clarke(1980) to the case deletion model. Ross(1987) states that "the difference between a connection coefficient Γ^s_{pq} for M and the corresponding coefficient $\Gamma^s_{(r)pq}$ for $M_{(r)}$ is a quantity that depends on the nonplanarity of M as a submodel of $M_{(r)}$ but not on the nonuniformity of parameterization. Thus for a model M which is sufficiently planar, the nonzero connection coefficients of $M_{(r)}$ will

be approximately equal to those of M." Under the assumption of planarity, Ross derives the quadratic approximation as:

$$\hat{\boldsymbol{\theta}}^p_{(r)} - \hat{\boldsymbol{\theta}}^p \doteq u^p_{(r)} - \frac{1}{2}\Gamma^p_{qs}u^q_{(r)}u^s_{(r)} \tag{6.6}$$

$$\hat{\zeta} \doteq \frac{e}{1-h_r}. \tag{6.7}$$

The second term in Equation 6.6 uses the Einstein summation convention described in Section 2.3.1. The superscripts in that equation denote particular rows or elements of the matrix or vector.

6.2 Case Deletion for Multilinear Models

Since multilinear models are a special case of nonlinear models, the linear and quadratic approximations to the parameter estimates when the r^{th} case is deleted for a nonlinear model can be used in the multilinear setting. The various jackknife estimates for nonlinear regression can also be applied to multilinear models. The conditional linearity of multilinear models suggests that the computational advantages of these approximations would even be more pronounced in this setting. Indeed, we have seen in Chapter III that the special structure of these models leads to nice forms for the tangent and acceleration vectors. Since the matrix $\boldsymbol{V}$ is essential for the computation of the jackknife estimates, and the accelerations are needed for the quadratic approximation, the structure of multilinear models facilitates obtaining these estimates.

6.2.1 A Numerical Example

The leave-out-one approach for nonlinear regression was applied to a one-factor trilinear model using the NFAK1A data described in Section 3.4. The least squares estimates using the full data were computed using a FORTRAN program developed by R.T.Ross (See Leurgans, Ross and Abel, 1990.) Actual leave-out-one estimates were also obtained by deleting one observation at a time and re-computing the estimates. Computations were done on the CRAY at the Ohio Supercomputer Center and on the DEC 5500 at the Mathematical Sciences Computing Laboratory. Case deletion estimates were obtained using the linear approximation LP, and the quadratic approximation described by Ross (1987) with the planarity assumption. The linear, weighted linear, and reweighted linear jackknife estimates given in Chapter V were also calculated. The results are summarized in the tables and figures below. The parameterization used was that of setting the last elements of α and β to 1.

Table 3 gives summary statistics for 600 case deletion estimates obtained by actually leaving out one observation at a time. The first two columns of numbers are the known, true parameter values, and the least squares estimates for the full data. The subscript LS of $\boldsymbol{\theta}$ denotes that the true parameter values were normalized with the same constants used to normalize the least squares estimates. The full least squares estimates are generally close to the true values, except maybe for $\boldsymbol{\alpha}[3]$, which differs by 0.124, and $\boldsymbol{\gamma}[4]$, with a difference of 24.77. This can be seen in Figure 1, which shows the global estimates on the y-axis plotted against the true parameter

values on the x-axis, for each of the three ways. The plotting symbol is "L" (for least squares). The dotted line is the 45-degree line, where the true values equal the estimates.

The individual values of the leave-out-one estimates have means and medians that are almost exactly equal to the global estimates. Each of the 600 actual deletion estimates are plotted (plotting symbol is "@") in Figure 2 against the true value of the α's, β's and γ's. It is not surprising that the points cluster tightly about the 45-degree line, (except for $\gamma[4]$), since they should be very close to the global least squares estimates, and the global estimates are, for the most part, near the true $\boldsymbol{\theta}$.

Table 4 gives summary statistics for 600 case deletion estimates using the linear approximation (6.3). The true parameter values and the least squares estimates for the full data are given again for comparison purposes. As in the actual deletions, the means and medians of the the leave-out-one linear approximation estimates are almost exactly equal to their full least squares counterparts. The ranges of the linear deletion estimates are slightly larger than the ranges of the actual deletion estimates, and the standard errors in Table 4 are somewhat larger than those in Table 3, especially for the γ's. In Figure 3, each of the 600 linear estimates are plotted (plotting symbol is "@") against the true θ's. Observe that the points cluster around the 45-degree line, but show a wider spread than Figure 2 for the γ's. The estimates for $\gamma[4]$ all exceed the true value, just as in Figures 1 and 2.

Table 5 presents summary statistics on the absolute difference between the actual case deletion estimates and those obtained using the linear approximation. In the

first two columns, the signed difference "actual minus linear" are given in parentheses. The differences appear negligible, except for the parameters of the third way. Considering the magnitude of the γ's relative to the α's and β's, it may be useful to describe the differences as fractions of the full least squares estimates. For the α's, the largest absolute difference between the actual and linear leave-out estimates is 0.072 (for $\alpha[8]$); this difference is 0.014 times as big as the full LSE. Among the β's, the maximum absolute difference between the actual and leave-out estimates is 0.006 (for $\beta[11]$), which corresponds to a relative difference of 0.0049. The maximum absolute difference among the γ's is 21.462, or 0.0135 relative to the full LSE's. Figure 4 is a plot of the 600 linear estimates versus the actual deletion estimates. Observe that the points (plotting symbol is "@") lie more or less along the 45-degree line.

In Table 6 the case deletion estimates using the quadratic approximation are summarized. Huge differences from the global estimates and from the true parameter values can be seen in the $\hat{\theta}_{(r)}$'s, particularly for estimates of some of the α's. The medians are equal to the global LSE's (up to three decimal places), but as expected, the means are pulled considerably away by the extreme values. Further investigation showed that the minima occur at the same value of r, observation number 593, or $\boldsymbol{Y}[10, 11, 3]$. A look at the quantiles of $\hat{\alpha}_{(r)}$ suggests that there are other wild values, but most of the estimates are "well-behaved." For instance, the first five percentiles are: -46.87, -8.34, -2.22, -0.64, and -0.14.

Examining the approximation (6.6), we see that the difference between the linear

and quadratic approximations is a sum of products of connection coefficients and rows of $\boldsymbol{u}_{(r)}$. Since we saw in Chapter 2 that the terms in the parameter-effects curvature array were not that large (although larger than the intrinsic curvatures), the extreme values must be coming from summing products of rows of $\boldsymbol{u}_{(r)}$. Indeed, a moderately large value of $\boldsymbol{u}_{(r)}$ would more greatly affect the quadratic estimates than the linear ones. Thus, the quadratic leave-out estimates may behave fairly well in most cases but can yield terrible estimates for observations with high leverages.

The leverages for the NFAK1A data range from 3.2×10^{-7} to 0.2747; the quartiles are: 0.0004, 0.0220, and 0.0671. The leverage of observation number 593 is 0.0815, which is the 83.6^{th} quantile, and is almost twice as large as the mean leverage of $P \div N = 25 \div 600 = 0.0417$.

Table 7 presents a comparison of the true parameter values, $\boldsymbol{\theta}$, the least squares estimates using the full data (LS), the ordinary jackknife estimates (J), the linear (LP), weighted linear (LQ), and reweighted linear (RLQ) jackknife estimates discussed in Chapter V. The subscript LS indicates that the vector was normalized with the same constants used to normalize the least squares estimates. The LP, LQ, and RLQ estimates differ very little among themselves and are not too far from the standard jackknife estimates, nor from $\boldsymbol{\theta}$. The standard estimates generally do better than the approximations LP, LQ and RLQ, especially for the γ's, although for seven of the parameters, the approximations are closer to $\boldsymbol{\theta}$. In Figure 5, the least squares (LSE), standard jackknife (J), linear jackknife (LP), weighted linear jackknife (LQ), and reweighted linear jackknife (RLQ) estimates are plotted against

the true $\boldsymbol{\theta}$ using "L," "J," "P," "Q," and "R," respectively, as the plotting symbols. The symbols appear almost indistinguishable because they overlap a lot. The characters that appear like a "U" near the 45-degree line for γ is actually the overalp of "L" and "J," the least squares and ordinary jackknife estimates. Note that the figure agrees with our findings from Table 7.

Table 8 gives the standard errors for the various jackknife estimates. We see that the ordinary jackknife estimates have larger standard errors than the approximations for most of the parameters except, notably, for the γ's.

We can see a trade-off between bias and variance reduction in this example. The jackknife estimates using the case deletion approximations have values farther from $\boldsymbol{\theta}$ but have smaller standard errors for most of the parameters. The ordinary jackknife estimates have slightly larger standard errors but yield values closer to the true parameter.

Table 3: True, Least-Squares, and Actual Case Deletion Estimates

Theta	True Value θ_{LS}	Full LSE	Leave-out-one Estimates $\hat{\theta}_{(r)}$ Using Actual Deletion				
			Min	Max	Median	Mean	SD
$\alpha[1]$	0.000	-0.015	-0.022	-0.008	-0.015	-0.015	0.002
$\alpha[2]$	0.000	0.004	-0.003	0.010	0.004	0.004	0.002
$\alpha[3]$	0.000	0.124	0.117	0.138	0.124	0.124	0.002
$\alpha[4]$	0.018	0.031	0.024	0.039	0.031	0.031	0.002
$\alpha[5]$	0.293	0.334	0.327	0.343	0.334	0.334	0.002
$\alpha[6]$	1.753	1.692	1.682	1.699	1.692	1.692	0.002
$\alpha[7]$	4.410	4.378	4.370	4.385	4.378	4.378	0.002
$\alpha[8]$	5.168	5.205	5.200	5.211	5.205	5.205	0.001
$\alpha[9]$	3.091	3.108	3.101	3.116	3.108	3.108	0.002
$\alpha[10]$	1.023	1.000	0.992	1.007	1.000	1.000	0.002
$\beta[1]$	0.000	-0.011	-0.013	-0.009	-0.011	-0.011	0.000
$\beta[2]$	0.000	-0.003	-0.005	-0.001	-0.003	-0.003	0.000
$\beta[3]$	0.000	0.012	0.011	0.015	0.012	0.012	0.000
$\beta[4]$	0.000	-0.014	-0.017	-0.012	-0.014	-0.014	0.000
$\beta[5]$	0.001	0.010	0.008	0.012	0.010	0.010	0.000
$\beta[6]$	0.009	0.006	0.004	0.008	0.006	0.006	0.000
$\beta[7]$	0.061	0.053	0.051	0.055	0.053	0.053	0.000
$\beta[8]$	0.247	0.232	0.229	0.234	0.232	0.232	0.000
$\beta[9]$	0.630	0.636	0.634	0.638	0.636	0.636	0.000
$\beta[10]$	1.060	1.063	1.061	1.066	1.063	1.063	0.000
$\beta[11]$	1.221	1.215	1.213	1.218	1.215	1.215	0.000
$\beta[12]$	0.996	1.000	0.998	1.002	1.000	1.000	0.000
$\gamma[1]$	1584.986	1589.683	1584.400	1595.480	1589.664	1589.684	1.223
$\gamma[2]$	1553.923	1561.232	1556.647	1566.264	1561.205	1561.235	1.143
$\gamma[3]$	1509.538	1510.737	1506.839	1514.435	1510.738	1510.737	1.169
$\gamma[4]$	1440.879	1465.645	1461.822	1470.880	1465.600	1465.645	1.148
$\gamma[5]$	1320.854	1315.337	1309.790	1319.701	1315.331	1315.337	1.189

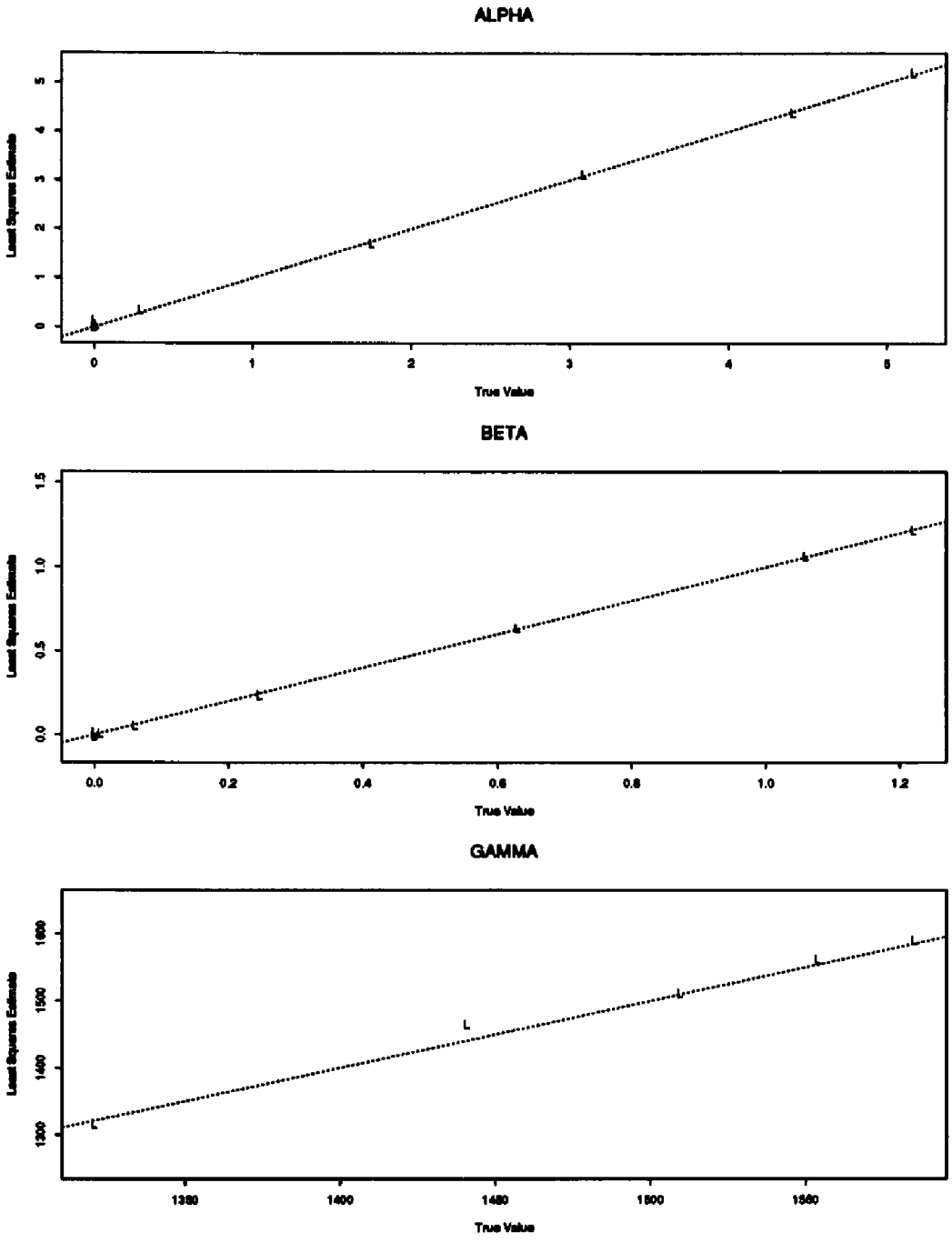

Figure 1: Full Least Squares Estimates vs. True Parameter Values

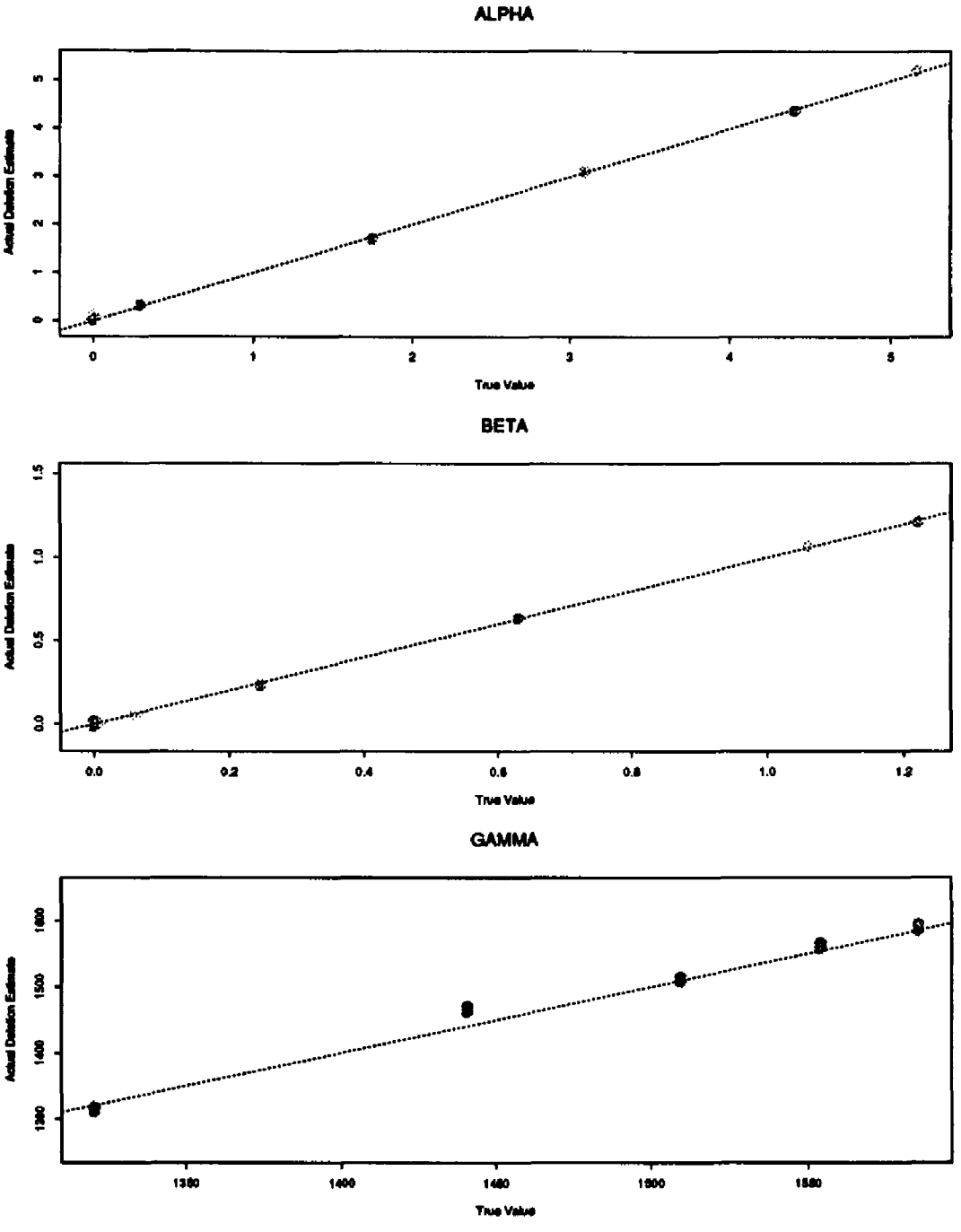

Figure 2: Actual Deletion Estimates vs. True Parameter Values

Table 4: True, Least-Squares, and Linear Case Deletion Estimates

Theta	True Value θ_{LS}	Full LSE	Leave-out-one Estimates $\hat{\theta}_{(r)}$ Using Linear Approximation				
			Min	Max	Median	Mean	SD
$\alpha[1]$	0.000	-0.015	-0.023	-0.006	-0.015	-0.015	0.001
$\alpha[2]$	0.000	0.004	-0.007	0.014	0.004	0.004	0.001
$\alpha[3]$	0.000	0.124	0.117	0.133	0.124	0.124	0.001
$\alpha[4]$	0.018	0.031	0.020	0.040	0.031	0.031	0.001
$\alpha[5]$	0.293	0.334	0.326	0.344	0.334	0.334	0.001
$\alpha[6]$	1.753	1.692	1.669	1.712	1.692	1.692	0.002
$\alpha[7]$	4.410	4.378	4.317	4.428	4.378	4.378	0.006
$\alpha[8]$	5.168	5.205	5.133	5.265	5.205	5.206	0.007
$\alpha[9]$	3.091	3.108	3.065	3.144	3.108	3.109	0.004
$\beta[1]$	0.000	-0.011	-0.013	-0.009	-0.011	-0.011	0.000
$\beta[2]$	0.000	-0.003	-0.005	-0.001	-0.003	-0.003	0.000
$\beta[3]$	0.000	0.012	0.010	0.015	0.012	0.012	0.000
$\beta[4]$	0.000	-0.014	-0.017	-0.012	-0.014	-0.014	0.000
$\beta[5]$	0.001	0.010	0.008	0.013	0.010	0.010	0.000
$\beta[6]$	0.009	0.006	0.004	0.009	0.006	0.006	0.000
$\beta[7]$	0.061	0.053	0.050	0.058	0.053	0.053	0.000
$\beta[8]$	0.247	0.232	0.229	0.234	0.232	0.232	0.000
$\beta[9]$	0.630	0.636	0.633	0.639	0.636	0.636	0.000
$\beta[10]$	1.060	1.063	1.058	1.067	1.063	1.063	0.000
$\beta[11]$	1.221	1.215	1.210	1.219	1.215	1.215	0.001
$\gamma[1]$	1584.986	1589.683	1571.201	1610.934	1589.683	1589.656	2.159
$\gamma[2]$	1553.923	1561.232	1542.514	1582.103	1561.232	1561.200	2.166
$\gamma[3]$	1509.538	1510.737	1493.948	1532.632	1510.737	1510.701	2.059
$\gamma[4]$	1440.879	1465.645	1449.357	1485.238	1465.645	1465.594	1.986
$\gamma[5]$	1320.854	1315.337	1300.719	1332.920	1315.337	1315.312	1.837

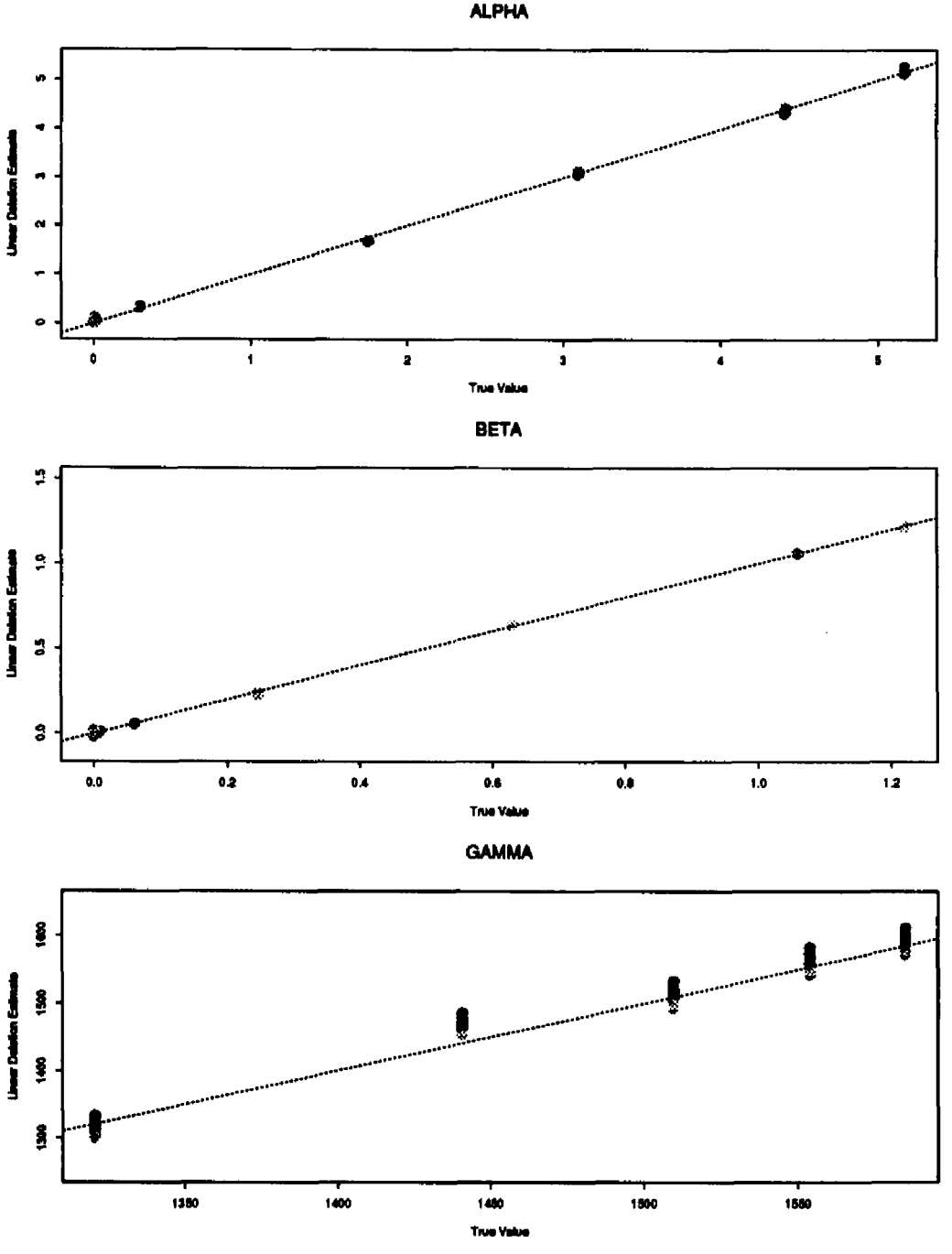

Figure 3: Linear Approximation Deletion Estimates vs. True Parameter Values

Table 5: Absolute Difference of Actual and Linear Approximation Estimates

Theta	Minimum	Maximum	Median	Mean	SD
$\alpha[1]$	0.000 (-0.016)	0.016 (+0.014)	0.000	0.001	0.002
$\alpha[2]$	0.000 (-0.017)	0.017 (+0.017)	0.000	0.001	0.002
$\alpha[3]$	0.000 (-0.008)	0.016 (+0.016)	0.000	0.001	0.002
$\alpha[4]$	0.000 (-0.014)	0.018 (+0.018)	0.000	0.001	0.002
$\alpha[5]$	0.000 (-0.015)	0.015 (+0.014)	0.000	0.001	0.002
$\alpha[6]$	0.000 (-0.020)	0.023 (+0.023)	0.000	0.001	0.003
$\alpha[7]$	0.000 (-0.051)	0.060 (+0.060)	0.000	0.002	0.006
$\alpha[8]$	0.000 (-0.060)	0.072 (+0.072)	0.000	0.002	0.007
$\alpha[9]$	0.000 (-0.036)	0.043 (+0.043)	0.000	0.002	0.004
$\beta[1]$	0.000 (-0.004)	0.004 (+0.003)	0.000	0.000	0.001
$\beta[2]$	0.000 (-0.003)	0.004 (+0.004)	0.000	0.000	0.001
$\beta[3]$	0.000 (-0.003)	0.005 (+0.005)	0.000	0.000	0.001
$\beta[4]$	0.000 (-0.004)	0.004 (+0.004)	0.000	0.000	0.000
$\beta[5]$	0.000 (-0.004)	0.004 (+0.004)	0.000	0.000	0.000
$\beta[6]$	0.000 (-0.004)	0.004 (+0.004)	0.000	0.000	0.000
$\beta[7]$	0.000 (-0.004)	0.004 (+0.003)	0.000	0.000	0.000
$\beta[8]$	0.000 (-0.004)	0.004 (+0.004)	0.000	0.000	0.000
$\beta[9]$	0.000 (-0.003)	0.003 (+0.003)	0.000	0.000	0.000
$\beta[10]$	0.000 (-0.004)	0.005 (+0.005)	0.000	0.000	0.001
$\beta[11]$	0.000 (-0.005)	0.006 (+0.006)	0.000	0.000	0.001
$\gamma[1]$	0.008 (-21.462)	21.462 (+17.316)	1.035	1.345	1.941
$\gamma[2]$	0.008 (-21.078)	21.078 (+20.951)	1.012	1.368	2.082
$\gamma[3]$	0.007 (-20.033)	20.033 (+16.718)	0.985	1.351	1.931
$\gamma[4]$	0.007 (-19.788)	19.788 (+16.219)	0.955	1.321	1.904
$\gamma[5]$	0.010 (-17.758)	17.758 (+14.556)	0.862	1.256	1.835

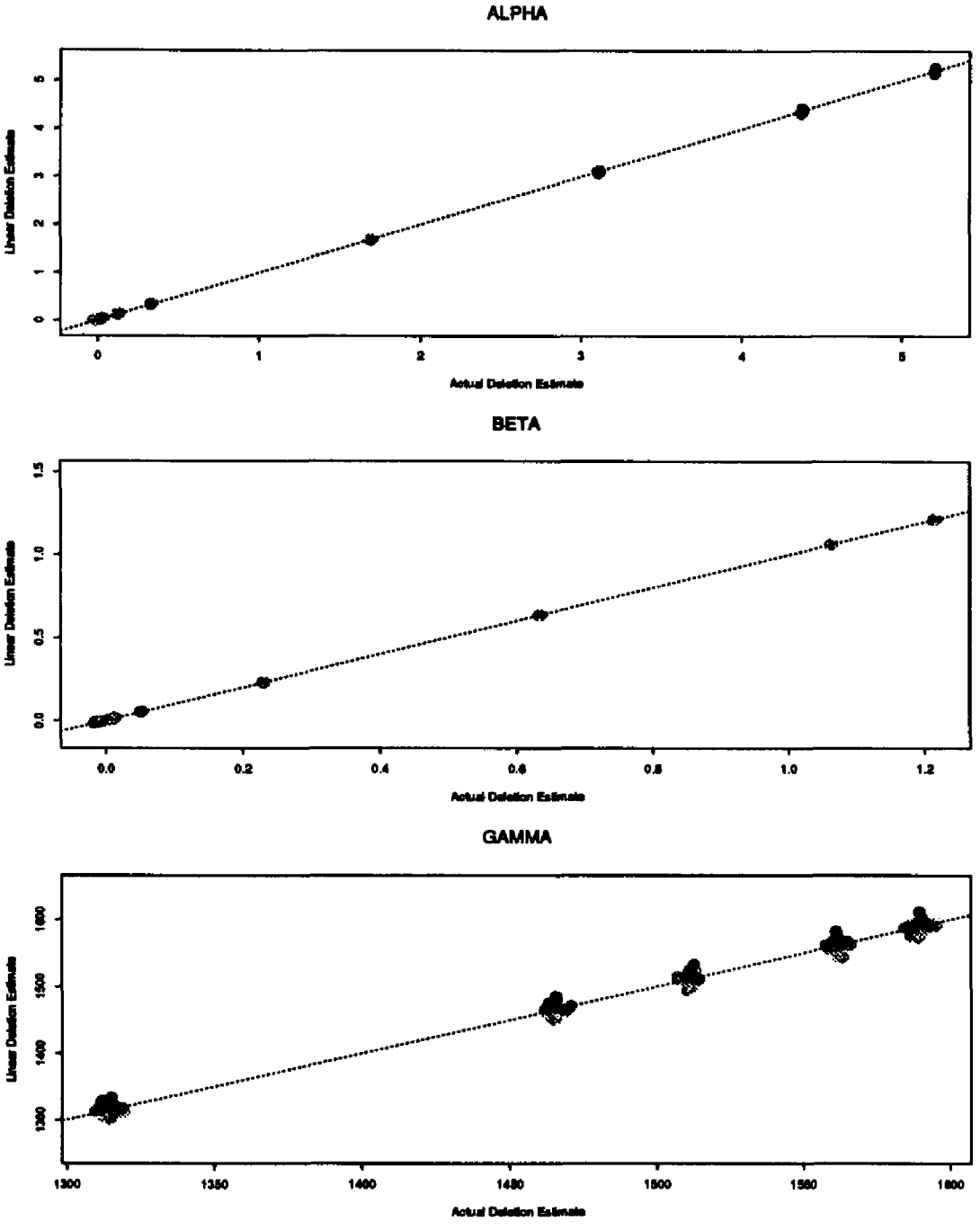

Figure 4: Linear Approximation vs. Actual Deletion Estimates

Table 6: Comparison of True, Least Squares, and Jackknife Estimates

Theta	θ_{LS}	Full LSE	J_{LS}	LP	LQ	RLQ
$\alpha[1]$	0.000	-0.015	-0.015	-0.011	-0.011	-0.011
$\alpha[2]$	0.000	0.004	0.005	0.016	0.015	0.014
$\alpha[3]$	0.000	0.124	0.121	0.128	0.129	0.130
$\alpha[4]$	0.018	0.031	0.032	0.011	0.011	0.011
$\alpha[5]$	0.293	0.334	0.334	0.316	0.317	0.318
$\alpha[6]$	1.753	1.692	1.694	1.637	1.640	1.643
$\alpha[7]$	4.410	4.378	4.379	4.260	4.263	4.266
$\alpha[8]$	5.168	5.205	5.205	5.055	5.062	5.069
$\alpha[9]$	3.091	3.108	3.108	3.003	3.009	3.015
$\alpha[10]$	1.023	1.000	1.001	1.000	1.000	1.000
$\beta[1]$	0.000	-0.011	-0.011	-0.013	-0.013	-0.013
$\beta[2]$	0.000	-0.003	-0.003	-0.003	-0.003	-0.003
$\beta[3]$	0.000	0.012	0.012	0.011	0.011	0.011
$\beta[4]$	0.000	-0.014	-0.013	-0.005	-0.006	-0.006
$\beta[5]$	0.001	0.010	0.010	0.008	0.008	0.009
$\beta[6]$	0.009	0.006	0.006	0.004	0.004	0.004
$\beta[7]$	0.061	0.053	0.053	0.044	0.045	0.045
$\beta[8]$	0.247	0.232	0.233	0.251	0.249	0.247
$\beta[9]$	0.630	0.636	0.635	0.642	0.641	0.640
$\beta[10]$	1.060	1.063	1.063	1.074	1.072	1.070
$\beta[11]$	1.221	1.215	1.216	1.254	1.246	1.240
$\beta[12]$	0.996	1.000	0.999	1.000	1.000	1.000
$\gamma[1]$	1584.986	1589.683	1588.966	1605.455	1608.613	1610.859
$\gamma[2]$	1553.923	1561.232	1559.278	1580.720	1583.546	1585.479
$\gamma[3]$	1509.538	1510.737	1510.828	1532.512	1534.493	1535.722
$\gamma[4]$	1440.879	1465.645	1465.739	1496.080	1495.273	1494.384
$\gamma[5]$	1320.854	1315.337	1314.962	1330.070	1332.550	1334.162

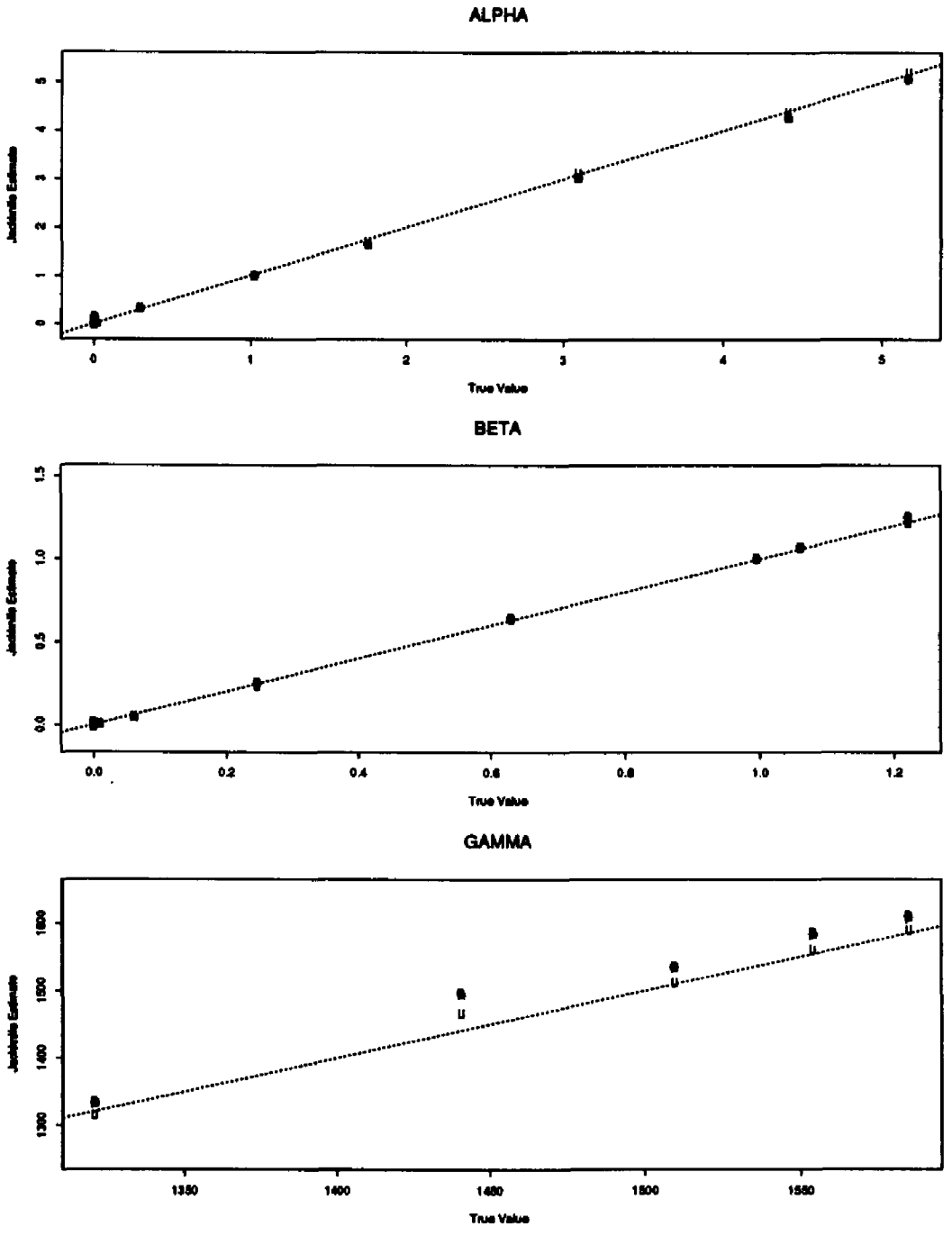

Figure 5: Various Jackknife Estimates vs. True Parameter Values

Table 7: Standard Errors of Jackknife Estimates

Parameter	J_{LS}	LP	LQ	RLQ
$\alpha[1]$	0.0432	0.0290	0.0273	0.0258
$\alpha[2]$	0.0448	0.0286	0.0271	0.0256
$\alpha[3]$	0.0442	0.0276	0.0259	0.0244
$\alpha[4]$	0.0457	0.0257	0.0242	0.0228
$\alpha[5]$	0.0439	0.0295	0.0278	0.0262
$\alpha[6]$	0.0439	0.0591	0.0551	0.0515
$\alpha[7]$	0.0411	0.1428	0.1328	0.1237
$\alpha[8]$	0.0343	0.1692	0.1572	0.1464
$\alpha[9]$	0.0426	0.1033	0.0962	0.0896
$\alpha[10]$	0.0451	0.0000	0.0000	0.0000
$\beta[1]$	0.0108	0.0070	0.0065	0.0061
$\beta[2]$	0.0106	0.0075	0.0070	0.0065
$\beta[3]$	0.0111	0.0076	0.0071	0.0066
$\beta[4]$	0.0113	0.0074	0.0069	0.0065
$\beta[5]$	0.0104	0.0072	0.0067	0.0063
$\beta[6]$	0.0107	0.0070	0.0066	0.0062
$\beta[7]$	0.0106	0.0082	0.0076	0.0070
$\beta[8]$	0.0109	0.0077	0.0071	0.0066
$\beta[9]$	0.0104	0.0089	0.0077	0.0068
$\beta[10]$	0.0098	0.0108	0.0089	0.0074
$\beta[11]$	0.0091	0.0128	0.0103	0.0084
$\beta[12]$	0.0099	0.0000	0.0000	0.0000
$\gamma[1]$	29.9025	52.8910	48.9352	45.4321
$\gamma[2]$	27.9393	53.0591	48.8509	45.1873
$\gamma[3]$	28.5752	50.4390	46.7533	43.4583
$\gamma[4]$	28.0747	48.6561	45.0435	41.8351
$\gamma[5]$	29.0803	45.0062	41.5059	38.4405

We close this section with a discussion of the computational requirements for obtaining both the actual and approximate leave-out-one estimates. Obtaining the actual case deletion estimates was computationally intensive and required running on the CRAY at the Ohio Supercomputer Center. Running the FORTRAN program on the DEC at the Mathematical Sciences Computing Laboratory would have to be done as a batch job because of the longer processing time, and could possibly be hampered by disk space problems for a large data set. For the NFAK1A data, getting the actual deletion estimates involved fitting the model 601 times. In contrast, using the linear and quadratic approximations required only one global fit.

Once the velocities were calculated, the linear leave-out-one estimates were obtained quite easily, compared to the labor involved in the actual deletions. The programs were done in S (See Becker et al.,1988) and computations performed on the DEC at the Mathematical Sciences Computing Laboratory.

The quadratic estimates took much more time to compute than the linear ones. Evaluating the right hand side of Equation (6.6) took particularly long to compute, and had to be run as an S batch job.

A possible difficulty with the calculations using the approximations should be noted, however. If some of the parameter estimates for the full least squares fit are zero, $\hat{\boldsymbol{V}}^T\hat{\boldsymbol{V}}$ may become singular, and the aprroximations could then not be calculated. This situation could be remedied by deleting the zero rows of the full estimates and calculating a modified derivative matrix, or by replacing the estimates which are zero by tiny values and proceeding with the computations.

CHAPTER VII

Conclusion

7.1 Summary of Results

In Chapter III, closed-form expressions for the first- and second-order partial derivatives of the expectation function of a trilinear PARAFAC model were presented. Linear dependencies among the tangent and acceleration vectors were investigated. For the one-factor model, the acceleration vectors were found to be linearly independent, but some dependencies were shown to exist among the columns of blocks of $\boldsymbol{V}$ and blocks of $\ddot{\boldsymbol{\eta}}$. The number of dependencies was determined and used to compute relative curvature arrays for a numerical example.

In Chapters V and VI, approximations to the case deletion estimates $\hat{\boldsymbol{\theta}}_{(r)}$ and various jackknife estimates of $\boldsymbol{\theta}$ were considered. These were then applied to multilinear models, taking into account the special structure of $\boldsymbol{V}$ and $\ddot{\boldsymbol{\eta}}$ discussed in Chapter III.

For a numerical example using simulated data for a one-factor model, the linear approximation to the case deletion estimates was found to perform fairly well, compared with actual case deletion estimates. The largest absolute differences between the actual leave-out estimates and their linear approximation counterparts occurred

for the γ's. These five largest values were in the 20's; when compared to the differences between the global estimates and the true parameter values, however, the five resulting relative differences ranged from 3 to about 5. In view of the considerable computational requirements for actual case deletions, the approximate case deletion estimates show promise.

7.2 How Results Can be Used

The linear leave-out estimates can be used to compute various jackknife estimates of $\boldsymbol{\theta}$, as described in Chapter V and illustrated in the numerical example of Chapter VI. They are also useful for cross-validation purposes, as discussed in Chapter IV. In particular, the linear approximation formulas can be used to find estimates for models with increasing values of F, as well as to calculate the predicted residual sum of squares, in order to determine the number of factors present.

Since the leverages h_r are a by-product of the approximate leave-out calculations, the assessment of influence is another area where the deletion estimates are relevant. The construction of influence measures for nonlinear regression models was, in fact, the motivation for the derivation of results cited in Chapter VI. Influential points for multilinear models could be studied using such measures.

In general, the approximate deletion estimates are potentially relevant in any analysis which involves leaving out one or more observations. Situations where the observations are not purposely deleted, but are actually missing, could provide another realm of possible applications. In dealing with experimental designs with structurally missing observations, for instance, one can surmise that the adaptation

to multilinear models of the linear approximation estimates can similarly be made or extended to such designs.

7.3 Areas for Further Study

The numerical example used in this dissertation was for a one-factor trilinear PARAFAC model used to fit simulated data. It would be interesting to see how well the linear approximation works for real data. Estimating the parameters of an $F - factor$ model, with $F > 1$, to either real or simulated data would also be of interest. Simulation studies could be done to cross-validate $F - factor$ models using the approximate deletion estimates.

Linear dependencies among the velocities and accelerations were investigated mostly for the one-factor case. More work could be done to determine if, by setting certain constraints, some theoretical results could be obtained for the general F-factor model.

Curvature measures have not, to my knowledge, been used in multilinear analysis. The measures of nonlinearity introduced in Chapter III are just the tip of the iceberg, so to speak. Calculating these and other measures for additional data sets would give us a feel for the curvature behavior of multilinear models. Other curvature formulas for nonlinear models might also be simplified or derived analytically when applied to multilinear models.

The nice structure of the velocities and accelerations for PARAFAC models suggests that this advantage could be explored for other types of multilinear models, for instance, the Tucker models. Briefly, Tucker3 models can be seen as generaliza-

tions of the PARAFAC models with F_1 factors for the first way, F_2 factors for the second way, and F_3 factors for the third way. Tucker2 models can be interpreted as allowing general interactions between the F_1 factors associated with the first way and the F_2 factors associated with the second way, the third way not itself being associated with a single factor but reflecting the importance of the interactions. How the leave-out-one approximations will perform when applied to these models could be investigated. These models also have applications in spectroscopy, so their study is of theoretical as well as of practical interest.

Appendix A

S Notation

The S language was used to calculate the deletion estimates using approximations and to prepare the figures in this document. Aside from its computational and graphical use, S also provides convenient matrix and array notation. The S notation was therefore used extensively in this document. The relevant notation is explained below. For more details on the S language, see Becker, et.al. (1988).

A.1 S Matrix Notation

Let $\boldsymbol{X}$ be an $N \times P$ matrix. The $(n,p)^{th}$ element of $\boldsymbol{X}$ is denoted by $\boldsymbol{X}[n,p]$. The n^{th} row of $\boldsymbol{X}$ is denoted $\boldsymbol{X}[n,]$ and the p^{th} column is denoted $\boldsymbol{X}[,p]$.

The $N-1 \times P$ matrix consisiting of all rows except the n^{th} row and of all columns of $\boldsymbol{X}$ is written as $\boldsymbol{X}[-n,]$. Similarly, the $N \times P-1$ matrix consisting of all rows and of all columns except the p^{th} column of $\boldsymbol{X}$ is written as $\boldsymbol{X}[,-p]$. The $N-1 \times P-1$ matrix consisting of all rows except the n^{th} and of all columns except the p^{th} is denoted $\boldsymbol{X}[-n,-p]$.

The $R \times P$ sub-matrix consisting of the rows $\boldsymbol{X}[r_1,]$, $\boldsymbol{X}[r_2,]$, ..., $\boldsymbol{X}[r_R,]$ is written as $\boldsymbol{X}[(r_1,r_2,\ldots,r_R),]$. If r_1, r_2, ..., r_R are consecutive integers, we write $\boldsymbol{X}[r_1 : r_R,]$. Similarly, the $N \times C$ sub-matrix consisting of the columns $\boldsymbol{X}[,c_1]$,

$\boldsymbol{X}[,c_2]$, ..., $\boldsymbol{X}[,c_C]$ is written as $\boldsymbol{X}[,(c_1,c_2,\ldots,c_C)]$. $\boldsymbol{X}[,c_1 : c_C]$. If c_1, c_2, ..., c_C are consecutive integers, we write $\boldsymbol{X}[,c_1 : c_C]$. The $R \times C$ sub-matrix consisting of the entries in rows r_1 through r_R of $\boldsymbol{X}$ and in columns c_1 through c_C of $\boldsymbol{X}$ is denoted by $\boldsymbol{X}[r_1 : r_R, c_1 : c_C]$.

A.2 S Array Notation

The S notation for arrays follows basically the same pattern as for matrices, except for the addition of one or more indices to represent the additional face or faces. Let $\boldsymbol{Y}$ be an $N \times P \times Q$ array. The $(n,p,q)^{th}$ element of $\boldsymbol{Y}$ is denoted by $\boldsymbol{Y}[n,p,q]$. The n^{th} level of the first face of $\boldsymbol{Y}$ is written as $\boldsymbol{Y}[n,,]$; the p^{th} level of the second face of $\boldsymbol{Y}$ is written as $\boldsymbol{Y}[,p,]$; and the q^{th} level of the third face of $\boldsymbol{Y}$ is written as $\boldsymbol{Y}[,,q]$.

An array formed by leaving out a level of one or more faces of $\boldsymbol{Y}$ is denoted by placing a minus sign before the appropriate index or indices. For instance, the $N-1 \times P-1 \times Q$ array consisting of all levels except the n^{th} level of the first face, all except the p^{th} level of the second face, and all levels of the third face of $\boldsymbol{Y}$ is written as $\boldsymbol{Y}[-n,-p,]$.

The S notation for sub-matrices formed by taking blocks of rows and columns, as described above, also extends directly to a notation for sub-arrays formed by taking blocks of levels of each face.

Appendix B

Matrix Notation

B.1 Kronecker Product

Let $\boldsymbol{A} = (a_{mn})$ be an $M \times N$ matrix, and $\boldsymbol{B} = (b_{pq})$ be a $P \times Q$ matrix. The Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is an $MP \times NQ$ matrix defined by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1N}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2N}\boldsymbol{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1}\boldsymbol{B} & a_{M2}\boldsymbol{B} & \dots & a_{MN}\boldsymbol{B} \end{bmatrix}. \tag{B.1}$$

Kronecker products occur in expressions for the derivative matrix $\boldsymbol{V}$ and for the array of acceleration vectors $\ddot{\boldsymbol{\eta}}$.

B.2 Khatri-Rao Circle Product

Let $\boldsymbol{A}$ be an $I \times F$ matrix with column vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_F$, and let $\boldsymbol{B}$ be a $J \times F$ matrix with column vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_F$. Then the Khatri-Rao product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is an $IJ \times F$ matrix

$$\boldsymbol{A} \odot \boldsymbol{B} = \left[\boldsymbol{a}_1 \otimes \boldsymbol{b}_1 \mid \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 \mid \dots \mid \boldsymbol{a}_F \otimes \boldsymbol{b}_F \right]. \tag{B.2}$$

(Khatri and Rao, 1968).

The Khatri-Rao circle product is used in expressing the derivative matrix $\boldsymbol{V}$ in closed form for models with F factors.

B.3 Vectorization

Let $\boldsymbol{X} = (x_{ij})$ be an $I \times J$ matrix. The vectorization of $\boldsymbol{X}$, denoted by $vec(\boldsymbol{X})$, arranges the elements of $\boldsymbol{X}$ into a vector of length IJ by varying the row index first before varying the column index. In effect, $vec(\boldsymbol{X})$ stacks the columns of $\boldsymbol{X}$ on top of each other. It follows that the vectorization of the transpose of $\boldsymbol{X}$, $vec(\boldsymbol{X}^T)$, is a vector of length IJ obtained by varying the column index first before the row index. Thus

$$vec(\boldsymbol{X}) = \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{I1} \\ X_{12} \\ X_{22} \\ \vdots \\ X_{I2} \\ \vdots \\ \vdots \\ X_{1J} \\ X_{2J} \\ \vdots \\ X_{IJ} \end{bmatrix} \text{ and } vec(\boldsymbol{X}^T) = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1J} \\ X_{21} \\ X_{22} \\ \vdots \\ X_{2J} \\ \vdots \\ \vdots \\ X_{I1} \\ X_{I2} \\ \vdots \\ X_{IJ} \end{bmatrix}.$$

Let $\boldsymbol{Y}$ be an $I \times J \times K$ array. The vectorization of $\boldsymbol{Y}$, denoted by $vec(\boldsymbol{Y})$, arranges the elements of $\boldsymbol{Y}$ into a vector of length IJK such that the first index varies most rapidly and the third index varies most slowly. To denote the arrangement of the elements of $\boldsymbol{Y}$ where the last index varies most rapidly and the first index varies most slowly, the notation $vec(\boldsymbol{Y}^T)$ shall be used, in keeping with the notation for matrices. Thus,

$$vec(\boldsymbol{Y}) = \begin{bmatrix} Y_{111} \\ \vdots \\ Y_{I11} \\ \vdots \\ \vdots \\ Y_{1J1} \\ \vdots \\ Y_{IJ1} \\ \vdots \\ \vdots \\ Y_{11K} \\ \vdots \\ Y_{I1K} \\ \vdots \\ \vdots \\ Y_{1JK} \\ \vdots \\ Y_{IJK} \end{bmatrix} \quad \text{and} \quad vec(\boldsymbol{Y}^T) = \begin{bmatrix} Y_{111} \\ \vdots \\ Y_{11K} \\ \vdots \\ \vdots \\ Y_{1J1} \\ \vdots \\ Y_{1JK} \\ \vdots \\ \vdots \\ Y_{I11} \\ \vdots \\ Y_{I1K} \\ \vdots \\ \vdots \\ Y_{IJ1} \\ \vdots \\ Y_{IJK} \end{bmatrix}.$$

Vectorization of the matrix or array of expected values $\boldsymbol{\mu}$ or, equivalently, of $\boldsymbol{\mu}^T$, facilitates obtaining closed-form expressions for the velocity and acceleration vectors. Whereas in the literature on multilinear models, $vec(\boldsymbol{\mu})$ is used most often, if not always, $vec(\boldsymbol{\mu}^T)$ was used in this document in keeping with the conventional notation of factorial experiments where the last index is varied most rapidly.

B.4 Matrix-Array Multiplication

Let $\boldsymbol{B}$ be an $N_1 \times N_2$ matrix, and $\boldsymbol{C}$ be an $N_2 \times N_3 \times N_4$ array. The product

$$\boldsymbol{A} = [\boldsymbol{B}][\boldsymbol{C}],$$

denoted by the square bracket multiplication [ ][ ], is an $N_1 \times N_3 \times N_4$ array whose element in the n^{th} face, p^{th} row, and q^{th} column is

$$\boldsymbol{A}[n,p,q] = \sum_{i=1}^{N_2} \boldsymbol{B}[n,i]\boldsymbol{C}[i,p,q].$$

(See Bates and Watts, 1988.)

Matrix-array multiplication is performed to obtain the curvature arrays.

Appendix C

Proofs of Lemmas

This section presents the proofs of the lemmas stated in Section 3.3. These lemmas aim to determine whether any linear dependencies exist among the velocity and acceleration vectors. Once the number of such dependencies is known, one can find the combined dimension of the tangent and acceleration spaces, $P + P'$, which is necessary in the calculation of the curvature arrays. Although the value of $P+P'$ can be obtained numerically given some data, the nice structure of the deriavtive matrix $\boldsymbol{V}$ and the array of second derivatives $\ddot{\boldsymbol{\eta}}$ for multilinear models leads one to think that there must be some closed-form solution. For the one-factor trilinear PARAFAC model, we have found that linear independence or dependence can, indeed be proven analytically.

Section ?? gives the proof for linear independence of the acceleration vectors. In Section ??, pairwise independence of the tangent and accelration vectors is investigated. Mutual independence of these vectors is the subject of Section ??. All the proofs use notation from Sections 3.1.1 and 3.1.2. The alternative expressions for the columns of $\boldsymbol{V}$ and $\ddot{\boldsymbol{\eta}}$ presented in those sections now come in handy, since some of them serve a proof's purpose better than others.

C.1 Proof of Lemma III.1

To determine if any linear dependencies exist among the columns of $\ddot{\eta}$, we wish to know if there are nonzero coefficients q_{ijf}, r_{jkf}, and s_{ikf} that satisfy

$$\sum_{ijf} q_{ijf}\ddot{\eta}[,\mathcal{I}_f[i],\mathcal{J}_f[j]] + \sum_{jkf} r_{jkf}\ddot{\eta}[,\mathcal{J}_f[j],\mathcal{K}_f[k]] + \sum_{ikf} s_{ikf}\ddot{\eta}[,\mathcal{I}_f[i],\mathcal{K}_f[k]] = \overset{IJK\times 1}{\mathbf{0}}, \tag{C.1}$$

or

$$\sum_{i,j,f} q_{ijf}\begin{bmatrix} \overset{(i-1)JK\times 1}{\mathbf{0}} \\ \overset{(j-1)K\times 1}{\mathbf{0}} \\ \boldsymbol{\gamma}_f \\ \overset{(J-j)K\times 1}{\mathbf{0}} \\ \overset{(I-i)JK\times 1}{\mathbf{0}} \end{bmatrix} + \sum_{j,k,f} r_{jkf}\boldsymbol{\alpha}_f \otimes \begin{bmatrix} \overset{(j-1)K\times 1}{\mathbf{0}} \\ \mathbf{e}_k \\ \overset{(J-j)K\times 1}{\mathbf{0}} \end{bmatrix} + \sum_{i,k,f} s_{ikf}\begin{bmatrix} \overset{(i-1)JK\times 1}{\mathbf{0}} \\ \boldsymbol{\beta}_f \otimes \mathbf{e}_k \\ \overset{(I-i)JK\times 1}{\mathbf{0}} \end{bmatrix} = \overset{IJK\times 1}{\mathbf{0}},$$

where the summation index i ranges from 1 to $I-1$, the index j goes from 1 to $J-1$, and the index k from 1 to K.

The proof consists of taking mutually exclusive and exhaustive blocks of rows in the above equation, then solving for the coefficients, if any. Blocks of $J \cdot K = JK$ rows ($JK \times 1$ vectors) are considered. We start with the I^{th} or last block of JK rows, then look at the i'^{th} block, where i' is any of the other values of i.

Consider the I^{th} block of JK rows:

$$\sum_{i,j,f} q_{ijf}\mathbf{0} + \sum_{j,k,f} r_{j,k,f}\boldsymbol{\alpha}_f[I]\begin{bmatrix} \overset{(j-1)K\times 1}{\mathbf{0}} \\ \mathbf{e}_k \\ \overset{(J-j)K\times 1}{\mathbf{0}} \end{bmatrix} + \sum_{i,k,f} s_{ikf}\mathbf{0} = \overset{JK\times 1}{\mathbf{0}}$$

Since $\boldsymbol{\alpha}_f[I] = 1$ for $f = 1,\ldots,F$, we have

$$\sum_{f=1}^{F}\sum_{k=1}^{K} r_{jkf}\mathbf{e}_k = \overset{K\times 1}{\mathbf{0}} \text{ for } j = 1,\ldots,J-1$$

$$\text{or} \quad \sum_{f=1}^{F} r_{jkf} \;=\; 0 \;\text{ for } j = 1,\ldots,J-1,\; k = 1,\ldots,K. \tag{C.2}$$

For $i = i'$, the i^{th} block of JK rows, for $i' = 1,\ldots,I-1$ is

$$\begin{aligned}
\overset{JK\times 1}{\mathbf{0}} \;=\; \sum_{f=1}^{F} \Bigg\{ & q_{i'1f} \begin{bmatrix} \boldsymbol{\gamma}_f \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \ldots + q_{i'(J-1)f} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \boldsymbol{\gamma}_f \\ \mathbf{0} \end{bmatrix} + \sum_{k=1}^{K} r_{1kf}\boldsymbol{\alpha}_f[i'] \begin{bmatrix} \boldsymbol{e}_k \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \\
& + \sum_{k=1}^{K} r_{(J-1)kf}\boldsymbol{\alpha}_f[i'] \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \boldsymbol{e}_k \\ \mathbf{0} \end{bmatrix} + \sum_{k=1}^{K} s_{i'kf}\boldsymbol{\beta}_f \otimes \boldsymbol{e}_k \Bigg\}
\end{aligned} \tag{C.3}$$

We shall further divide the i^{th} block of JK rows into mutually exclusive and exhaustive sub-blocks of K rows ($K \times 1$ vectors) each :

The J^{th} or last block of K rows in Equation C.3 is

$$\begin{aligned}
\sum_{f=1}^{F}\sum_{k=1}^{K} s_{i'kf}\boldsymbol{e}_k \;&=\; \overset{K\times 1}{\mathbf{0}} \text{ since } \boldsymbol{\beta}_f[J] = 1 \text{for } f = 1,\ldots,F, \\
\text{or} \quad \sum_{f=1}^{F} s_{i'kf} \;&=\; 0 \;\text{ for }\; i' = 1,\ldots,I-1 \; k = 1,\ldots,K.
\end{aligned} \tag{C.4}$$

For $j = 1,\ldots,J-1$, the j^{th} block of K rows in Equation C.3 is

$$\begin{aligned}
\overset{K\times 1}{\mathbf{0}} \;&=\; \sum_{f=1}^{F} \Big\{ q_{i'jf}\boldsymbol{\gamma}_f + \sum_{k=1}^{K} r_{jkf}\boldsymbol{\alpha}_f[i']\boldsymbol{e}_k + \sum_{k=1}^{K} s_{i'kf}\boldsymbol{\beta}_f[j]\boldsymbol{e}_k \Big\} \\
\text{or } 0 \;&=\; \sum_{f=1}^{F} \Big\{ q_{i'jf}\boldsymbol{\gamma}_f[k] + r_{jkf}\boldsymbol{\alpha}_f[i'] + s_{i'kf}\boldsymbol{\beta}_f[j] \Big\} \\
\text{for } i' \;&=\; 1,\ldots,I-1,\; j = 1,\ldots,J-1,\; k = 1,\ldots,K.
\end{aligned} \tag{C.5}$$

From the above equality, Equations C.2 and C.4 we have Lemma III.1. In particular, for $F = 1$, $q_{ijf} = r_{jkf} = s_{ikf} = 0$.

C.2 Proof of Lemma III.2

When $F = 1$, the tangent vectors are given by Equations 3.6, 3.7, and 3.8 as

$$\boldsymbol{V}_A[,i] = \begin{bmatrix} {\scriptstyle (i-1)JK\times 1} \\ \mathbf{0} \\ \boldsymbol{\beta}\otimes\boldsymbol{\gamma} \\ {\scriptstyle (I-i)JK\times 1} \\ \mathbf{0} \end{bmatrix}, \; i = 1,\ldots,I-1, \tag{C.6}$$

$$\boldsymbol{V}_B[,j] = \boldsymbol{\alpha}\otimes\begin{bmatrix} {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{\gamma} \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \end{bmatrix}, \; j = 1,\ldots,J-1, \tag{C.7}$$

$$\text{and } \boldsymbol{V}_C[,k] = \boldsymbol{\alpha}\otimes\boldsymbol{\beta}\otimes\boldsymbol{e}_k, \; k = 1,\ldots K. \tag{C.8}$$

and the acceleration vectors are given by Equations 3.16, 3.18, and 3.17 as

$$\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]] = \begin{bmatrix} {\scriptstyle (i-1)JK\times 1} \\ \mathbf{0} \\ {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{\gamma} \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \\ {\scriptstyle (I-i)JK\times 1} \\ \mathbf{0} \end{bmatrix}, \tag{C.9}$$

$$\ddot{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]] = \boldsymbol{\alpha}\otimes\begin{bmatrix} {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{e}_k \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \end{bmatrix}, \tag{C.10}$$

$$\text{and } \ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]] = \begin{bmatrix} {\scriptstyle (i-1)JK\times 1} \\ \mathbf{0} \\ \boldsymbol{\beta}\otimes\boldsymbol{e}_k \\ {\scriptstyle (I-i)JK\times 1} \\ \mathbf{0} \end{bmatrix} \tag{C.11}$$

for $i = 1,\ldots,I-1$, $j = 1,\ldots,J-1$, $k = 1,\ldots,K$.

The proof proceeds by taking corresponding (nonzero) blocks of rows of the tangent and velocity vectors and determining if one can be expressed as a linear combination of the other.

1. Independence of $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$

   We wish to determine if there exist coefficients c_j, not all 0, such that

   $$\boldsymbol{V}_A[,i] = \sum_{j=1}^{J-1} c_j \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]\,.$$

   Consider Equations C.6 and C.9. For a fixed i, we need only to look at the i^{th} block of JK rows because all other rows are 0 :

   $$\boldsymbol{\beta} \otimes \boldsymbol{\gamma} = \sum_{j=1}^{J-1} c_j \begin{bmatrix} \overset{(j-1)K\times 1}{\mathbf{0}} \\ \boldsymbol{\gamma} \\ \overset{(J-j)K\times 1}{\mathbf{0}} \end{bmatrix}$$

   Since $\boldsymbol{\beta}[J] = 1$, the last K rows of the above equation yield $\boldsymbol{\gamma} = \sum_{j=1}^{J-1} c_j \mathbf{0}$. There are no nonzero cofficients c_j that will satisfy this unless all the elements of $\boldsymbol{\gamma}$ are 0. Thus, $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$ are linearly independent.

2. Independence of $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$

   Do there exist coefficients $a_k \neq 0$ such that

   $$\boldsymbol{V}_A[,i] = \sum_{k=1}^{K} a_k \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]\ ?$$

   Consider Equations C.6 and C.11. For a fixed i, we again need only look at the i^{th} block of JK rows :

   $$\boldsymbol{\beta} \otimes \boldsymbol{\gamma} = \sum_{k=1}^{K} a_k \boldsymbol{\beta} \otimes \boldsymbol{e}_k$$

   The above equality holds if we take $a_k = \boldsymbol{\gamma}[k],\ k = 1,\ldots,K$. Therefore $\boldsymbol{V}_A[,i]$ is a linear combination of $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$.

3. Independence of $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]$

   Do there exist coefficients $b_{jk} \neq 0$ such that

$$\boldsymbol{V}_A[,i] = \sum_{j,k} b_{jk}\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]\ ?$$

   Consider Equations C.6 and C.10. For a fixed i, the i^{th} block of JK rows is

$$\boldsymbol{\beta} \otimes \boldsymbol{\gamma} = \sum_{j,k} b_{jk}\boldsymbol{\alpha}[i] \begin{bmatrix} {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{e}_k \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \end{bmatrix}.$$

   Since $\boldsymbol{\beta}[J] = 1$, the last K rows of the above equation yields $\boldsymbol{\gamma} = \sum_{j,k} b_{jk}\boldsymbol{\alpha}[i]\mathbf{0}$. There are no nonzero cofficients b_{jk} that will satisfy this unless all the elements of $\boldsymbol{\gamma}$ are 0. Thus, $\boldsymbol{V}_A[,i]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]$ are linearly independent.

4. Independence of $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$

   We wish to determine if there exist coefficients c_i, not all 0, such that

$$\boldsymbol{V}_B[,j] = \sum_{i=1}^{I-1} c_i\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]].$$

   In Equations C.7 and C.9, consider the last (I^{th}) block of JK rows :

$$\boldsymbol{\alpha}[I] \begin{bmatrix} {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{\gamma} \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \end{bmatrix} = \sum_{i=1}^{I-1} c_i\mathbf{0}$$

   The above equality holds only if $\boldsymbol{\alpha}[I]\boldsymbol{\gamma} = 1 \cdot \boldsymbol{\gamma} = \mathbf{0}$. So $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$ are linearly independent.

5. Independence of $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$

Do there exist coefficients $b_{ik} \neq 0$ such that

$$\boldsymbol{V}_B[,j] = \sum_{i,k} b_{ik} \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]] \;?$$

In Equations C.7 and C.11, consider the last (I^{th}) block of JK rows:

$$1 \cdot \begin{bmatrix} {\scriptstyle (j-1)K\times 1} \\ \mathbf{0} \\ \boldsymbol{\gamma} \\ {\scriptstyle (J-j)K\times 1} \\ \mathbf{0} \end{bmatrix} = \sum_{i,k} b_{ik}\mathbf{0}$$

Again, there are no constants b_{ik} which can satisfy the above equation unless $\boldsymbol{\gamma} = \mathbf{0}$. Thus, $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$ are linearly independent.

6. Independence of $\boldsymbol{V}_B[,j]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]$

Do there exist coefficients $a_k \neq 0$ such that

$$\boldsymbol{V}_B[,j] = \sum_{k=1}^{K} a_k \ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]] \;?$$

Consider Equations C.7 and C.10. For a given j, we need only to consider the j^{th} sub-block of K rows within each of the I blocks of JK rows, since all other elements are 0 :

$$\boldsymbol{\alpha}[i]\boldsymbol{\gamma} = \sum_{k=1}^{K} a_k \boldsymbol{\alpha}[i]\boldsymbol{e}_k = \boldsymbol{\alpha}[i] \sum_{k=1}^{K} a_k \boldsymbol{e}_k$$

The above equation is satisfied when $a_k = \boldsymbol{\gamma}[k],\ k = 1,\ldots,K$. Therefore $\boldsymbol{V}_B[,j]$ is a linear combination of $\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]$.

7. Independence of $\boldsymbol{V}_C[,k]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$

We wish to determine if there exist coefficients a_{ij}, not all 0, such that

$$\boldsymbol{V}_C[,k] = \sum_{i,j} a_{ij} \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]\,.$$

In Equations C.8 and C.9, consider the I^{th} block of JK rows:

$$\boldsymbol{\alpha}[I]\boldsymbol{\beta} \otimes \boldsymbol{e}_k = 1 \cdot \boldsymbol{\beta} \otimes \boldsymbol{e}_k = \sum_{i,j} a_{ij}\mathbf{0} \, .$$

Since there are no coefficients a_{ij} that will satisfy this equation, unless $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{V}_C[,k]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{J}[j]]$ are linearly independent.

8. Independence of $\boldsymbol{V}_C[,k]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$

   Do there exist coefficients $b_i \neq 0$ such that

$$\boldsymbol{V}_C[,k] = \sum_{i=1}^{I-1} b_i \ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]] \; ?$$

   Consider Equations C.8 and C.11. Similar to the above, the last block of JK rows is

$$\boldsymbol{\beta} \otimes \boldsymbol{e}_k = \sum_{i=1}^{I-1} b_i \mathbf{0} \, ,$$

   which holds only if all the β's are 0. So $\boldsymbol{V}_C[,k]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{I}[i],\mathcal{K}[k]]$ are linearly independent.

9. Independence of $\boldsymbol{V}_C[,k]$ and $\ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]]$

   Do there exist coefficients $c_j \neq 0$ such that

$$\boldsymbol{V}_C[,k] = \sum_{j=1}^{J-1} c_j \ddot{\boldsymbol{\eta}}\,[,\mathcal{J}[j],\mathcal{K}[k]] ?$$

   In Equations C.8 and C.10, consider the J^{th} sub-block of K rows in each of the I blocks of JK rows :

$$\begin{aligned} \boldsymbol{\alpha}[i]\boldsymbol{\beta}[J]\boldsymbol{e}_k &= \sum_{j=1}^{J-1} c_j \boldsymbol{\alpha}[i]\mathbf{0} \\ \text{or } \boldsymbol{\alpha}[i] \cdot 1 \cdot \boldsymbol{e}_k &= \boldsymbol{\alpha}[i] \sum_{j=1} J - 1 c_j \mathbf{0} \, . \end{aligned}$$

Since there are no coefficients c_j which will satisfy the above equations unless each $\alpha[i]$ is 0, $V_C[,k]$ and $\tilde{\eta}[,\mathcal{J}[j],\mathcal{K}[k]]$ are linearly independent.

C.3 Proof of Lemma III.3

To determine if the velocities and accelerations are mutually independent for a one-factor trilinear model, we wish to know if there exist nonzero coefficients a_i, b_j, c_k, q_{ij}, r_{jk}, and s_{ik} such that

$$\begin{aligned} \overset{IJK\times 1}{\mathbf{0}} = \sum_i a_i \boldsymbol{V}[,\mathcal{I}[i]] \;+\; \sum_j b_j \boldsymbol{V}[,\mathcal{J}[j]] + \sum_k c_k \boldsymbol{V}[,\mathcal{K}[k]] + \\ \sum_{i,j} q_{ij}\tilde{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]] \;+\; \sum_{j,k} r_{jk}\tilde{\boldsymbol{\eta}}[,\mathcal{J}[j],\mathcal{K}[k]] + \sum_{i,k} s_{ik}\tilde{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{K}[k]]. \quad \text{(C.12)} \end{aligned}$$

We start by evaluating each of the linear combinations in the above equation :

1. $\sum_{i=1}^{I-1} a_i \boldsymbol{V}[,\mathcal{I}[i]]$

   From Equation 3.13, we know $\boldsymbol{V}[,\mathcal{I}[i]]$ is nonzero only on $\boldsymbol{\xi}[i,,]$, where $\boldsymbol{V}[,\mathcal{I}[i]]$ equals $\boldsymbol{\beta}\otimes\boldsymbol{\gamma}$, independent of i. Note that the term $\boldsymbol{\xi}[I,,]$ is never disturbed because it is out of the sum. If we define an $I\times 1$ vector $\boldsymbol{a}_0$ to be

$$\boldsymbol{a}_0 = \begin{bmatrix} a_1 \\ \vdots \\ a_{I-1} \\ 0 \end{bmatrix},$$

   then

$$\sum_{i=1}^{I-1} a_i \boldsymbol{V}[,\mathcal{I}[i]] = \boldsymbol{a}_0 \otimes \boldsymbol{\beta} \otimes \boldsymbol{\gamma}. \quad \text{(C.13)}$$

2. $\sum_{j=1}^{J-1} b_j \boldsymbol{V}[,\mathcal{J}[j]]$

As in the above, we see from Equation 3.14 that $\boldsymbol{V}[,\mathcal{J}[j]]$ is nonzero only on $\boldsymbol{\xi}[,j,]$. If we define a $J \times 1$ vector

$$\boldsymbol{b}_0 = \begin{bmatrix} b_1 \\ \vdots \\ b_{J-1} \\ 0 \end{bmatrix},$$

then

$$\sum_{j=1}^{J-1} b_i \boldsymbol{V}[,\mathcal{J}[j]] = \boldsymbol{\alpha} \otimes \boldsymbol{b}_0 \otimes \boldsymbol{\gamma}. \tag{C.14}$$

3. $\sum_{k=1}^{K} c_k \boldsymbol{V}[,\mathcal{K}[k]]$

Similarly, defining a $K \times 1$ vector $\boldsymbol{c}$ to be

$$\boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix},$$

we see from Equation 3.15 that

$$\sum_{k=1}^{K} c_k \boldsymbol{V}_C[,k] = \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \boldsymbol{c}. \tag{C.15}$$

4. $\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} q_{ij} \ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$

Equation 3.23 shows that $\ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]]$ is equal to $\boldsymbol{\gamma}$ on $\boldsymbol{\xi}[i,j,]$, and zero elsewhere. If we define an $I \times J$ matrix $\boldsymbol{Q}$ to be

$$\boldsymbol{Q}[i',j'] = \begin{cases} \boldsymbol{q}[i',j'] & i' \leq I-1;\ J' \leq J-1 \\ 0 & i' = I or j' = J \end{cases} \quad \text{so that} \quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{bmatrix},$$

then

$$\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} q_{ij} \ddot{\boldsymbol{\eta}}[,\mathcal{I}[i],\mathcal{J}[j]] = vec(\boldsymbol{Q}^T) \otimes \boldsymbol{\gamma}. \tag{C.16}$$

5. $\sum_{j=1}^{J-1}\sum_{k=1}^{K} r_{jk}\ddot{\eta}[,\mathcal{J}[j],\mathcal{K}[k]]$

It can be seen from Equation 3.25 that $\ddot{\eta}[,\mathcal{J}[j],\mathcal{K}[k]]$ is equal to α on $\xi[,,k]$ and zero elsewhere. Defining a $J \times K$ matrix R to be

$$R[j,k] = \begin{cases} r_{jk} & j \leq J-1 \\ 0 & j = J, \end{cases}$$

we have

$$\sum_{j=1}^{J-1}\sum_{k=1}^{K} r_{jk}\ddot{\eta}[,\mathcal{J}[j],\mathcal{K}[k]] = \alpha \otimes vec(R^T) \tag{C.17}$$

6. $\sum_{i=1}^{I-1}\sum_{k=1}^{K} s_{ik}\ddot{\eta}[,\mathcal{I}[i],\mathcal{K}[k]]$

Again, it can be seen from Equation 3.24 that $\ddot{\eta}[,\mathcal{I}[i],\mathcal{K}[k]]$ is nonzero only on $\xi[i,,k]$, at which it is equal to β. If we define an $I \times K$ matrix S as

$$S = \begin{bmatrix} s \\ 1\times K \\ 0 \end{bmatrix}$$

similar to the above, then

$$\sum_{i=1}^{I-1}\sum_{k=1}^{K} s_{ik}\ddot{\eta}[,\mathcal{I}[i],\mathcal{K}[k]] = vec(\varphi_{12}(\beta \times S^T)) \tag{C.18}$$

where φ_{12} exchanges ways 1 and 2 of an array.

From Equations C.13, C.14, C.15, C.16, C.17, C.18, and C.12, the columns of V and $\ddot{\eta}$ are linearly independent if there exist nontrivial a_0, b_0, c, Q, R, and S, such that

$$\begin{aligned} \overset{IJK\times 1}{0} &= a_0 \otimes \beta \otimes \gamma \;+\; \alpha \otimes b_0 \otimes \gamma + \alpha \otimes \beta \otimes c \\ &+ vec(Q^T) \otimes \gamma \;+\; \alpha \otimes vec(R^T) + vec(\varphi_{12}(\beta \times S^T)) \end{aligned} \tag{C.19}$$

Grouping the second and fifth terms of the above equation, we get

$$\boldsymbol{\alpha} \otimes (\boldsymbol{b}_0 \otimes \boldsymbol{\gamma} + vec(\boldsymbol{R}^T)).$$

$$\text{But } \boldsymbol{b}_0 \otimes \boldsymbol{\gamma} + vec(\boldsymbol{R}^T) = vec(\boldsymbol{R}^T + \boldsymbol{\gamma} \times \boldsymbol{b}_0).$$

Observe that $\boldsymbol{R}^{T*} = -\boldsymbol{b}_0 \times \boldsymbol{\gamma}$ is always valid, because $\boldsymbol{R}^{T*}[,J] = -\boldsymbol{\gamma}\boldsymbol{b}_0[J] = 0$.

Similarly, grouping the first and sixth terms of Equation C.19, gives

$$vec(\boldsymbol{\gamma}\times\boldsymbol{\beta}\times\boldsymbol{a}_0+\varphi_{12}(\boldsymbol{\beta}\times\boldsymbol{S}^T) = vec(\varphi_{12}(\boldsymbol{\beta}\times\boldsymbol{\gamma}\times\boldsymbol{a}_0+\boldsymbol{\beta}\times\boldsymbol{S}^T) = vec(\varphi_{12}(\boldsymbol{\beta}\times(\boldsymbol{\gamma}\times\boldsymbol{a}_0+\boldsymbol{S}^T)).$$

Again, note that $\boldsymbol{S}^{T*} = -\boldsymbol{\gamma} \times \boldsymbol{a}_0$ is always valid.

Thus, there exist 2 nontrivial solutions to Equation C.19. It remains to be shown that there exist no others. Set

$$\begin{aligned} \Delta_1 &= \boldsymbol{\gamma} \times \boldsymbol{b}_0 + \boldsymbol{R}^T \\ \text{and } \Delta_2 &= \boldsymbol{\gamma} \times \boldsymbol{a}_0 + \boldsymbol{S}^T \end{aligned}$$

These are the "amounts" by which $\boldsymbol{R}^T$ and $\boldsymbol{S}^T$ fail to cause the singularities just derived.

Equation C.19 now is

$$\boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \boldsymbol{c} + vec(\boldsymbol{Q}^T) \otimes \boldsymbol{\gamma} + vec(\Delta_1 \times \boldsymbol{\alpha}) + vec(\varphi_{12}(\boldsymbol{\beta} \times \Delta_2)) = \boldsymbol{0}. \qquad \text{(C.20)}$$

We need to show that this implies $\boldsymbol{c}$, $\boldsymbol{Q}$, Δ_1, and Δ_2 are all zero. Now look at the above equation as an array equation :

$$\boldsymbol{c} \times \boldsymbol{\beta} \times \boldsymbol{\alpha} + \boldsymbol{\gamma} \times \boldsymbol{Q}^T + \Delta_1 \times \boldsymbol{\alpha} + \varphi_{12}(\boldsymbol{\beta} \times \Delta_2) = {}^{K\times J\times I}\boldsymbol{0}. \qquad \text{(C.21)}$$

Consider sub-matrices for each way :

1. Take the [,,I] subface:

$$\boldsymbol{c} \times \boldsymbol{\beta} \times \boldsymbol{\alpha}[I] + \boldsymbol{\gamma} \times \boldsymbol{Q}^T[,I] + \Delta_1 \boldsymbol{\alpha}[I] + \boldsymbol{\beta} \times \Delta_2[,I] = \mathbf{0}$$

But $\boldsymbol{\alpha}[I] = \mathbf{0}$ by convention, $\boldsymbol{Q}[,I] = \mathbf{0}$ by definition, and $\Delta_2[,I] = 0$ by definition of $\boldsymbol{S}$ and $\boldsymbol{a}_0$. So the above equation reduces to

$$\boldsymbol{c} \times \boldsymbol{\beta} + \Delta_1 = \mathbf{0} \tag{C.22}$$

Now take the [,J] column:

$$\boldsymbol{c}\boldsymbol{\beta}[J] + \Delta_1[,J] = \mathbf{0}.$$

But $\Delta_1[,J] = \mathbf{0}$ by definition of $\boldsymbol{R}$ and $\boldsymbol{b}_0$, and $\boldsymbol{\beta}[J] = 1$ by convention. Therefore,

$$\boldsymbol{c} = \mathbf{0}.$$

From Equation C.22,

$$\Delta_1 = \mathbf{0}.$$

2. Repeat the above using the [,J,] subface. Equation C.21 gives

$$\boldsymbol{c} \times \boldsymbol{\beta}[J] \times \boldsymbol{\alpha} + \boldsymbol{\gamma} \times \boldsymbol{Q}^T[J,] + \Delta_1[,J,] \times \boldsymbol{\alpha} + \boldsymbol{\beta}[J]\Delta_2 = \mathbf{0}.$$

Since $\boldsymbol{\beta}[J] = 1$, and $\boldsymbol{Q}^T[J,] = \mathbf{0}$, and $\Delta_1[,J] = \mathbf{0}$,

$$\boldsymbol{c} \times \boldsymbol{\alpha} + \Delta_2 = \mathbf{0}.$$

But $\boldsymbol{c} = \mathbf{0}$, so

$$\Delta_2 = \mathbf{0}.$$

3. It only remains to show that $\boldsymbol{Q} = \mathbf{0}$. Consider the [k,,] face. from Equation C.21, we get

$$\boldsymbol{c}[k] \times \boldsymbol{\beta} \times \boldsymbol{\alpha} + \boldsymbol{\gamma}[k]\boldsymbol{Q}^T + \Delta[k,] \times \boldsymbol{\alpha} + \boldsymbol{\beta} \times \Delta_2[k,,] = \mathbf{0}$$

Since $\boldsymbol{c}[k] = 0$, $\Delta_1[k,] = \mathbf{0}$, and $\Delta_2[k,,] = \mathbf{0}$,

$$\boldsymbol{\gamma}[k]\boldsymbol{Q}^T = \mathbf{0}.$$

So $\boldsymbol{\gamma}[k] \neq 0$ implies

$$\boldsymbol{Q}^T = \mathbf{0} \text{ or } \boldsymbol{Q} = \mathbf{0}.$$

Appendix D

S Code

D.1 General Steps

This section outlines the steps for obtaining : (1) the case deletion estimates for a one-factor trilinear model using the linear approximation; (2) the linear, weighted linear, and reweighted linear jackknife estimates; (3) the derivative matrix, leverages, and other quantities incidental to the calculations of (1) and(2). The S functions used to carry out the steps, variable names used in these functions, and relevant comments are given in parentheses following each step.

To obtain the $\hat{\theta}_r$'s using the linear approximation:

1. Obtain full least squares estimates by running R.T.Ross' FORTRAN program on the complete data set.

2. Set $\alpha[I]$ and $\beta[J]$ equal to 1 by dividing α and β by the appropriate constants; then adjust γ by multiplying by the same constants. For comparison purposes, do the same normalization on the true values of the parameters. (The normalized full least squares estimates are named *alpha.norm, beta.norm, gamma.norm*; the normalized true parameter values are named *alpha.comp, beta.comp, gamma.comp*)

3. Read in the data. (The function fun.dataray reads in the data in the usual input format to FORTRAN except that the title, names, and other characters are removed.)

4. Obtain the derivative matrix $\boldsymbol{V}$. (The function fun.v will return an $N \times P$ matrix.)

5. Compute $(\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T$. (The function fun.vtvivt will return a $P \times N$ matrix. If the "solve" command complains that $\boldsymbol{V}$ is singular, check using the "svd" command. If the singular value decomposition confirms the singularity of $\boldsymbol{V}$, the problem may be due to several parameter estimates being zero. If so, modify $\boldsymbol{V}$ by deleting the full estimates which are zero or setting them to a very small value, say 10^{-6}. Also, for large matrices, using "svd" instead of "solve" might be more efficient.)

6. Obtain the leverages h_r. (The function fun.hr will return a vector of length N.)

7. Compute

$$rho = \frac{\boldsymbol{e}}{\boldsymbol{1} - \boldsymbol{h_r}}.$$

(The function fun.rho will return a vector of length N.)

8. Calculate the $\hat{\boldsymbol{\theta}}_r$'s. (The function fun.thetadel will return the $P \times N$ matrix containing the linear approximation leave-out estimates of the P parameters when each of the N observations are deleted.)

To obtain the linear (LP), weighted linear (LQ) and reweighted linear (RLQ) jackknife estimates :

1. Do steps (1) to (6) above.

2. To find LP, compute *rho* (see step (7) above); to find LQ, compute $rholq = e$; to find RLQ, compute $rhorlq = e(1 - h_r)$. (Each of the functions fun.rho, fun.rholq, and fun.rhorlq will return a vector of length N.)

3. To find LP, compute

$$nmur = n(V^T V)^{-1} V^T \frac{e}{1 - h_r};$$

which is the second term in Equation 5.2; to find LQ, compute *nmurlq*, the second term in Equation ??; to find RLQ, compute *nmurrlq*, the second term in Equation ??. (Use *rho*, *rholq*, and *rhorlq*, respectively, as arguments to the function fun.nmur.)

4. Obtain the jackknife estimates LP, LQ and RLQ. (Use *nmur*, *nmurlq*, and *nmurrlq*, respectively, as arguments to the function fun.lp.)

D.2 S Functions

Below are the S functions referred to in the previous section. The arguments have been given names that are as obvious as possible. The discussion in the previous section should also shed light on what the arguments represent.

Most of functions given here are very short; they can easily be combined into one long function, or a main function that calls each of the shorter ones. For clarity

and ease of understanding, the simpler functions are given. Also, performing all the calculations in one swoop may cause dynamic memory problems for large data sets, so it is sometimes useful to be able to break up the program into smaller pieces.

```
fun.dataray ← function(filename = "nfak1a.data")
{ dataraw ← scan(filename, what = "")
dataraw ← as.numeric(dataraw)
I ← dataraw[1]
J ← dataraw[2]
K ← dataraw[3]
N ← I * J * K
dataray ← array(dataraw[9 + I + J + K : 8 + I + J + K + N], c(I, J,
K))
return(dataray) }

fun.kronecker ← function(amat, bmat)
{ T.amat as.matrix(amat)
T.bmat as.matrix(bmat)
T.matprod matrix(aperm((outer(T.amat,T.bmat)),c(3,1,4,2)),
ncol = ncol(T.amat) * ncol(T.bmat))
return(T.matprod) }
```

```
fun.v ← function(alpha.norm, beta.norm, gamma.norm)

{ I ← length(alpha.norm)

J ← length(beta.norm)

K ← length(gamma.norm)

P ← I + J + K - 2

v11 ← fun.kronecker(diag(I - 1), fun.kronecker(beta.norm, gamma.norm))

v1 ← rbind(v11, matrix(rep(0, J * K * (I - 1)), ncol =I - 1))

v2b ← rbind(fun.kronecker(diag(J - 1), gamma.norm),

matrix(rep(0, K * (J - 1)), ncol = J - 1))

v2 ← fun.kronecker(alpha.norm, v2b)

v3 ← fun.kronecker(fun.kronecker(alpha.norm, beta.norm), diag(K))

v ← cbind(v1, v2, v3)

return(v) }

fun.vtvivt ← function(v)

{ vtv ← t(v) % * % v

vtvivt ← solve(vtv, t(v))

return(vtvivt) }

fun.hr ← function(v, vtvivt)

{ h ← v % * % vtvivt
```

```
hr ← diag(h)

return(hr) }

fun.rho ← function(alpha.norm, beta.norm, gamma.norm, dataray, hr)

{ etaray ← outer(outer(alpha.norm, beta.norm), gamma.norm)

residray ← dataray - etaray

resid ← c(residray)

rho ← resid/(1 - hr)

return(rho) } endquote

    fun.thetadel ← function(alpha.norm, beta.norm, gamma.norm,

    vtvivt, rho)

    { I ← length(alpha.norm)

    J ← length(beta.norm)

    K ← length(gamma.norm)

    P ← (I - 1) + (J - 1) + K

    ur ← -1 * vtvivt * outer(rep(1, P), rho)

    theta.norm ← c(alpha.norm[ - I], beta.norm[ - J], gamma.norm)

    thetadel ← theta.norm + ur

    return(thetadel) }

    fun.nmur ← function(I, J, K, vtvivt, rho)
```

```
{ N ← I * J * K

P ← I + J + K - 2

nmur ← N * vtvivt * outer(rep(1, P), rho)

return(nmur) }

fun.lp ← function(alpha.norm, beta.norm, gamma.norm, nmur)

{ I ← length(alpha.norm)

J ← length(beta.norm)

K ← length(gamma.norm)

N ← I * J * K

P ← (I - 1) + (J - 1) + K

theta.norm ← c(alpha.norm[ - I], beta.norm[ - J], gamma.norm)

lpr ← theta.norm + nmur

lp ← apply(lpr, 1, mean)

varlp ← apply(lpr, 1, var)

varlp ← varlp/N

selp ← sqrt(varlp)

return(list(lp, selp)) }
```

BIBLIOGRAPHY

[1] C.J. Appelloff and E.R. Davidson. Strategies for analyzing data from video flurometric monitoring of liquid chromatographic effluents. *Anal. Chem.*, 53:2053–2056, 1981.

[2] D.M. Bates and D.G. Watts. Relative curvature measures of nonlinearity. *J. Roy. Stat. Soc.,Ser. B*, 42(1):1–25, 1980.

[3] D.M. Bates and D.G. Watts. *Nonlinear Regression Analysis and its Applications.* wiley, New York, 1988.

[4] R.A. Becker, J.M. Chambers, and A.R. Wilks. *The New S Language.* Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, 1988.

[5] J. R. Bunch and C.P. Nielsen. Updating the singular value decomposition. *Numerische Mathematik*, 31:111–129, 1978.

[6] J.Douglas Carroll and Jih jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decompositions. *Psychometrika*, 35:283–319, 1970.

[7] G.P.Y. Clarke. Moments of the least squares estimators in a nonlinear regression model. *J. Roy. Stat. Soc., Ser. B*, 42:227–237, 1980.

[8] J.J. Dongarra. *LINPACK Users' Guide.* SIAM, Philadelphia, 1979.

[9] G.T. Duncan. An empirical study of jackknife-constructed confidence regions in nonlinear regression. *Technometrics*, 20:123–129, 1978.

[10] H.T. Eastment and W. J. Krzanowski. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics*, 24:73–77, 1982.

[11] B. Efron and G. Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37:36–48, 1983.

[12] Richard Harshman, Peter Ladefoged, and Louis Goldstein. Factor analysis of tongue shapes. *J.Acoust. Soc. Am.*, 62:693, 1977.

[13] C.G. Khatri and C. Radhakrishna Rao. Solutions to some functional equations and their applications to characterizations of probability distributions. *Sankhya A*, 30:167–180, 1968.

[14] P.M. Kroonenberg. *Three-mode principal component analysis: Theory and Applications.* DSWO Press, Leiden, 1983.

[15] W.J. Krzanowski. Cross-validation in principal component analysis. *Proc. 19th International Biometrics Society*, 1986.

[16] S.E. Leurgans and R.T. Ross. Cross-validation for multilinear models: applications to biophysics. Department of Statistics and Biochemistry, Ohio State University, 1991.

[17] S.E. Leurgans and R.T. Ross. Multilinear models: Applications in spectroscopy. Technical Report 459, Department of Statistics, The Ohio State University , Columbus, Ohio, 1991.

[18] S.E. Leurgans, R.T. Ross, and R.B. Abel. A decomposition for 3-way arrays. Technical Report 448, Department of Statistics, The Ohio State University , Columbus, Ohio, 1990.

[19] R.G. Miller. An unbalanced jackknife. *Ann. Stat.*, 2:880–891, 1974.

[20] F. Mosteller and J.W. Tukey. *Data Analysis and Regression.* Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

[21] M.H. Quenouille. Notes on bias in estimation. *Biometrika*, 43:353–360, 1956.

[22] D.A. Ratkowsky. *Nonlinear Regression Modelling.* Marcel Dekker, Inc., New York, 1983.

[23] W.H. Ross. The geometry of case deletion and the assessment of influence in nonlinear regression. *Can. J. Statist.*, 15:91–103, 1987.

[24] J.S. Simonoff and Chih-Ling Tsai. Jackknife-based estimators and confidence regions in nonlinear regression. *Technometrics*, pages 103–112, 1986.

[25] J.E. Draheim S.R. Durell and E.L.Gross. *Arch. Biochem. Biophys.*, 267:217–227, 1988.

[26] M. Stone. Cross-validatory choice and assessment of statistical predictions(with discussion). *J. Roy. Stat. Soc., Ser. B*, 36:111–148, 1974.

[27] Svante Wold. Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, 20:397–405, 1978.