ELSEVIER

Chemometrics and Intelligent Laboratory Systems 38 (1997) 173–181

Chemometrics and intelligent laboratory systems

Two-factor degeneracies and a stabilization of PARAFAC

William S. Rayens [a,*], Benjamin C. Mitchell [b]

[a] *Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, USA*
[b] *ClinTrials Research, P.O. Box 13991, Research Triangle Park, NC 27709, USA*

Received 4 April 1996; revised 25 July 1996; accepted 8 March 1997

Abstract

Mitchell and Burdick (B.C. Mitchell, D.S. Burdick, An empirical comparison of resolution methods for three-way arrays, Chemom. Intell. Lab. Syst. 20 (1993) 149–161; B.C. Mitchell, D.S. Burdick, Slowly converging PARAFAC sequences: Swamps and two-factor degeneracies, J. Chemom. 8 (1994) 155–168.) uncovered an intriguing correspondence between the existence of PARAFAC swamps and the presence of two-factor degeneracies (2FDs). This observation, coupled with the recognition that post-swamp resolutions were generally better than pre-swamp resolutions allowed the user a method of detecting when a swamp had likely been encountered and, hence, when it was safe to assume that PARAFAC had converged to a suitable resolution. Still, this correspondence alone did not suggest how one might reduce the number of PARAFAC iterations required to move through a swamp. In this paper, it is noted that serious 2FDs must produce an identifiable ill-conditioning in the least squares estimation step in PARAFAC. Moreover, a serious 2FD is only one way this ill-conditioning may occur and, hence, it is more correct to say that swamps correspond to this ill-conditioning in general, rather than to the presence of 2FDs in particular. In an attempt to reduce the number of iterations that PARAFAC spends in a swamp, a particular method of stabilization is employed and results are presented which suggest that the number iterations can often be greatly reduced.

Keywords: PARAFAC; Two-factor degeneracies (2FDs)

1. Introduction

There are two main approaches to the trilinear analysis of three-way data arrays. One approach is to resolve the array by utilizing an eigenanalysis based procedure (EBP) [1–4], which typically works well when the signal-to-noise ratio is high. The other main approach utilizes alternating least squares in an iterative manner, exploiting the conditional linearity of the trilinear model. Its iterative nature means that starting values are required, but the algorithm is guaranteed to improve the least squares fit of the model to the data at each iteration. This approach is the one commonly used by psychometricians working in three mode factor analysis [5–7]. Its prototype is the PARAFAC algorithm developed and popularized by Harshman [5,6] and introduced into the chemometrics literature by Appellof and Davidson [8].

Unfortunately, PARAFAC will sometimes encounter what has been called a *two-factor degeneracy* (2FD) [9], in which two of the estimated factors

* Corresponding author.

PII S0169-7439(97)00033-6

exhibit high uncorrected correlation coefficients (UCCs) in all three modes, but in such a manner that the product of their respective UCCs is very near -1. Typically, in this situation PARAFAC has estimated one factor to have a prominent negative peak that is the mirror image of a similarly prominent positive peak in the other factor and an unfortunate cancellation is taking place, suggesting that there are fewer constituents present than there actually are [9]. For emphasis, this situation will be referred to as a *serious 2FD*, to distinguish from a more *moderate 2FD*, in which the product of the UCCs is negative, but not so close to -1. Regardless, since the true profiles being estimated are always nonnegative, it is obviously not possible for a resolution containing any 2FD to be highly accurate. An additional problem, as Mitchell and Burdick [10,11] discovered is that PARAFAC occasionally becomes immersed in *swamps*, regions in which the sequence slows to a crawl for many iterations, suggesting convergence, before emerging to move in much larger steps, often to a superior resolution. Further, Mitchell and Burdick offered evidence suggesting that swamps and serious 2FDs tended to occur simultaneously. Taken together, these problems insist that caution is warranted when declaring that PARAFAC has converged, since one could be left in the middle of a swamp, with a strong 2FD and inferior resolutions. In response, Mitchell and Burdick introduced a test for 2FDs which allowed the user to have some indication of whether a swamp was at hand. If so, one could choose either to tighten the convergence criterion for PARAFAC, allowing it to continue until hopefully the 2FD, and the swamp, disappear, or to immediately discontinue the current run and restart with another sequence, trusting that a new swamp would not be encountered.

The Mitchell and Burdick solution was an excellent ‘reactive’ solution to the swamp dilemma, in the sense that if a swamp was detected, then one might be well advised to start the PARAFAC sequence again, in an effort to avoid concluding wrongly that the algorithm had converged to an acceptable resolution. This paper takes a more ‘proactive’ approach, in the sense that a tangible construct within PARAFAC, instability associated with the least squares estimation, is found to be necessarily associated with the existence of serious 2FDs and, therefore, associated with the existence of swamps. Actually, this numerical instability will be present if any two of the abovementioned UCCs is near 1 in absolute value; hence the existence of a 2FD is just a special case of a more general problem. It is shown that this instability can be suitably manipulated, through the use of a relevant method of stabilization, so that the number of iterations spent in a swamp are often greatly reduced, in turn reducing the number of PARAFAC iterations required for post-swamp convergence.

The remainder of this paper is organized as follows. First the necessary definitions and notations are briefly reviewed and the synthetic data used in the examples are described. Second, the correspondence between the presence of swamps and instabilities within PARAFAC is made precise and illustrated. This is followed with a description and implementation of a relevant stabilization method and a discussion of the results.

2. Definitions and data sets

According to the trilinear PARAFAC model, a three-way data array $A = (a_{ijk})$ as given by

$$a_{ijk} = \sum_{r=1}^{R} x_{ir} y_{jr} z_{kr} + n_{ijk}$$

for $i = 1, 2, \ldots, I$, $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$ where $\mathbf{X}_{I\times R} = (x_{ir}) = [\mathbf{X}_1 \ldots \mathbf{X}_R]$, $\mathbf{Y}_{J\times R} = (y_{jr}) = [\mathbf{Y}_1 \ldots \mathbf{Y}_R]$ and $\mathbf{Z}_{K\times R} = (z_{kr}) = [\mathbf{Z}_1 \ldots \mathbf{Z}_R]$ are called factor matrices, each column of which is called a *profile*; and N is an array of random discrepancies. The data used in this paper were generated according to this model and are the same data that Mitchell and Burdick [10,11] used in their initial study. The reader is referred to that study for more details on those data. Simply note, that 64 different signal arrays were used in the simulation, all having $I = 40$, $J = 40$, $K = 4$ and $R = 4$. The mode $\mathbf{X}$, mode $\mathbf{Y}$ and mode $\mathbf{Z}$ profiles were carefully chosen to mimic profiles of actual chemical compounds, which were used, in turn, to generate sets of four artificial compounds. In every case the four profiles in a set were linearly independent, a condition sufficient to ensure that the array rank is 4 for each signal. The noise was taken to be proportional to the signal, rather than additive white noise for which least squares fitting procedures are optimal. Since noise encountered in practice is probably best modeled as some combination of proportional noise with additive white noise, generating data based on pure white noise could produce misleadingly optimistic results. Hence, a conservative approach was taken to avoid being misled in this way.

It should be noted that the original study by Mitchell and Burdick used as many as four starting values for each data array, with each being generated by an EBP, as discussed in Leurgans et al. [3]. To simplify the presentation in this paper, only random starts were utilized, eight for each of the 64 signals arrays. For convenience, a typical PARAFAC run will be referred to by the convention **ID_no**, where **ID** identifies the signal array and **no** identifies the random start. This notation does not provide immediate information about the signals to the reader, but is a convenient way of indexing the 512 different PARAFAC runs. As noted in Mitchell and Burdick, some of the IDs generated runs all of which were free of 2FDs, some IDs produced both runs with 2FDs and runs without; and a few generated runs all of which suffered from a serious 2FD, some continuing until termination. To investigate the relationship between 2FDs and the convergence properties of the PARAFAC algorithm, the test introduced in Mitchell and Burdick was performed for every pair of factors in the resolution and produced, for each PARAFAC iteration, an assessment of the existence and severity of a 2FD. This quantity will be referred to as ξ_n, where n is the corresponding iteration number; recall ξ_n is the smallest of all the possible products of the UCCs computed pairwise from the profiles. For comparison, let ζ_n denote the largest of all such products on the nth iteration.

Finally, to assess the amount of change that results from a PARAFAC iteration as well the degree of closeness between two triple product factorizations, the quantities Δ_n and $\Upsilon_{(n,n-1)}$ are used, just as Mitchell and Burdick [10,11] suggested. If $\Upsilon_{(n,n-1)}$ is close to 1, then the two factorizations are judged similar. Likewise, a small value of $\Delta_n \equiv \log_{10}[1-\Upsilon_{(n,n-1)}]$ indicates that the current PARAFAC iteration has produced little change because successive resolutions are nearly identical. It is logical therefore to terminate the PARAFAC iterations when Δ_n becomes less than some preset value Δ_0.

3. PARAFAC instability and 2FDs

A typical implementation of PARAFAC begins by choosing starting values for the **X** and **Y** factors, say $\mathbf{X}_0$ and $\mathbf{Y}_0$ and then using these, the data, and ordinary least squares to estimate **Z** say with $\mathbf{Z}_1$. $\mathbf{Z}_1$ in turn, might then be used with $\mathbf{X}_0$ to produce $\mathbf{Y}_1$ which is used with $\mathbf{Z}_1$ to produce $\mathbf{X}_1$. The resulting triple, $(\mathbf{X}_1, \mathbf{Y}_1, \mathbf{Z}_1)$, signifies the end of the first PARAFAC iteration. These conditional iterations are continued until convergence. It is instructive to look more closely at the least squares estimation. Without loss of generality, it is assumed that estimates of **X** and **Y** are being used to produce an updated estimate of **Z**. The least squares problem is posed by first 'unfolding' the data a_{ijk} into an $IJ \times K$ matrix

$$\mathbf{A} = \begin{pmatrix} a_{111} & a_{112} & - & a_{11K} \\ a_{121} & a_{122} & - & a_{12K} \\ - & - & - & - \\ a_{1J1} & a_{1J2} & - & a_{1JK} \\ a_{211} & a_{212} & - & a_{21K} \\ a_{221} & a_{222} & - & a_{22K} \\ - & - & - & - \\ a_{2J1} & a_{2J2} & - & a_{2JK} \\ - & - & - & - \\ - & - & - & - \\ a_{I11} & a_{I12} & - & a_{I1K} \\ a_{I21} & a_{I22} & - & a_{I2K} \\ - & - & - & - \\ a_{IJ1} & a_{IJ2} & - & a_{IJK} \end{pmatrix}$$

and then forming $\mathbf{A} = \mathbf{M}_{IJ\times R}\mathbf{Z}^{t}_{R\times K}$ where **M** is a full rank matrix formed with estimates of the **X** and **Y** profiles, structured so that the model definition is satisfied. The updated estimate of the **Z** factor is determined (symbolically) by $\mathbf{Z}^{t} = (\mathbf{M}^{t}\mathbf{M})^{-1}\mathbf{M}^{t}\mathbf{A}$ Denote $\mathbf{M}^{t}\mathbf{M} \equiv \mathbf{T}$; it is not hard to show that **T** has the form:

$$\mathbf{T}(x, y) = \left(\mathbf{X}_i^{t}\mathbf{X}_j\mathbf{Y}_i^{t}\mathbf{Y}_j\right)_{R\times R} \quad i, j = 1, \ldots, R.$$

If, however, there exists a perfect uncorrected correlation between a pair of X's and the corresponding pair of Y's, then $\mathbf{T}$ will have deficient rank. That is, if $X_i = aX_j$ and $Y_i = bY_j$ then

$$\begin{aligned}&\left(X_i^t X_1 Y_i^t Y_1,\ X_i^t X_2 Y_i^t Y_2,\ \dots\ X_i^t X_R Y_i^t Y_R\right)\\ &= \left(abX_j^t X_1 Y_j^t Y_1,\ \dots\ abX_j^t X_R Y_j^t Y_R\right)\\ &= ab\left(X_j^t X_1 Y_j^t Y_1,\ \dots\ X_j^t X_R Y_j^t Y_R\right)\\ &= ab \times \text{the } j\text{th row of } \mathbf{T}\end{aligned}$$

These simple observations can be collected into the following results. So that the notation might stay simple, the results are stated for only one of the three estimation steps, but obviously hold for the other two as well.

Theorem 3.1. *Let $T(x, y)$ be defined as above and assume there exists an i and j, so that $\mathbf{X}_i = a\mathbf{X}_j$ and $\mathbf{Y}_i = b\mathbf{Y}_j$. Then the rank of $\mathbf{T}$ is at most $R - 1$.*

Hence, in the presence of a perfect 2FD, the matrix $\mathbf{T}$ will have deficient rank.

Of course, no one expects that the uncorrected correlations within any two pairs of profiles will be exactly 1 or -1. However, the intuition suggested by this result is very clear: as profiles become more dependent, as measured by the UCCs, then $\mathbf{T}$ will become more unstable. In particular, one would expect the condition number of $\mathbf{T}$, typically defined as the ratio of the largest and smallest eigenvalues, to be very large in the presence of either a very small ξ_n (serious 2FD) or a very large ζ_n.

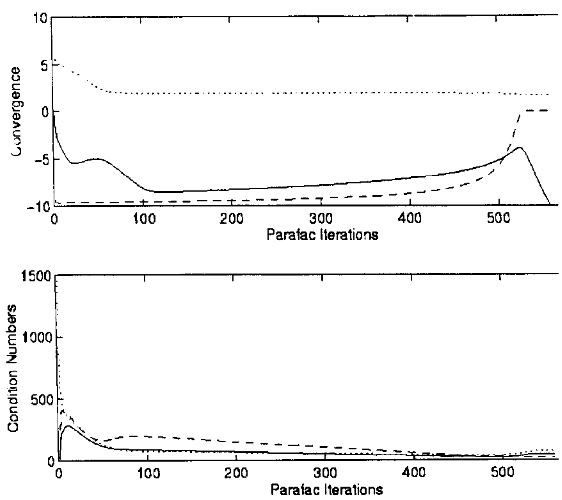

Fig. 1. Original PARAFAC for 341_6.

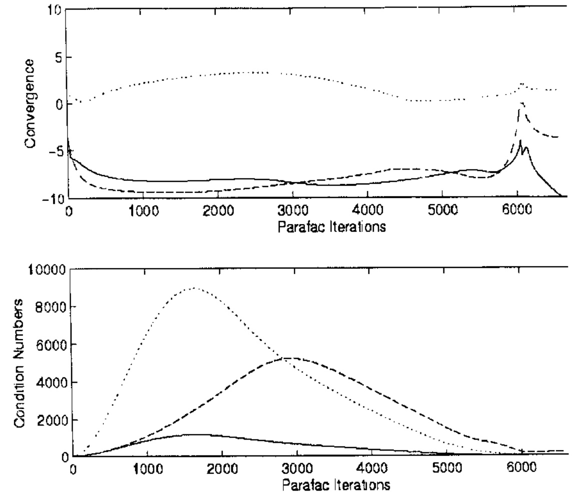

Fig. 2. Original PARAFAC for 231_3.

The synthetic data seem to bear out this intuition. For example, 341_6 and 231_3 were particularly difficult runs, as seen in Figs. 1 and 2.

The top plot in each figure records the convergence criterion $\log_{10}[1 - \Upsilon_{(n,n-1)}]$ (solid line), $10 \times \xi_n$ (dashed line) and $10 \times \zeta_n$ (dotted line), while the bottom plot records the condition numbers of $\mathbf{T}(y, z)$ (solid line), $\mathbf{T}(x, z)$ (dotted line) and $\mathbf{T}(x, y)$ (dashed line). It is evident that the 2FD for 341_6 is quite serious and this is reflected in the condition numbers, which follow a pattern that is completely consistent with ξ_n and ζ_n. Indeed, near the beginning, one of the estimation steps produced condition numbers in the range of 1500. Although it is difficult to see on the plot, all three of the estimation steps consistently yielded condition numbers greater than 100 until after the 100th iteration.

The behavior of 231_3 is even more extreme, with condition numbers higher than 8000. Once again, the condition numbers seemed to be at their worst in the middle of the associated swamp. On the other hand, 113_6 (Fig. 3) did not produce a prolonged swamp, but did encounter a serious 2FD early on. As expected, the condition numbers are moderately high in this region, but become very stable when ξ_n and ζ_n move back toward 0. Such patterns were noticed throughout the synthetic runs, suggesting that numerical instability in the estimation step is associated with the existence of swamps and may be affecting the number of iterations PARAFAC spends therein.

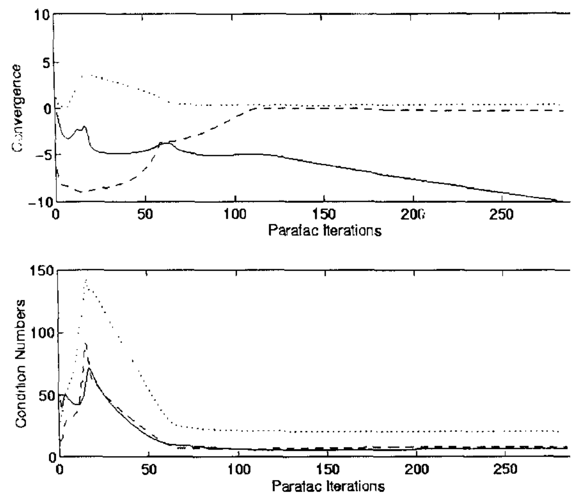

Fig. 3. Original PARAFAC for 113_6.

Before introducing the stabilization procedure, it should be emphasized that no claim is being made that poor conditioning will correspond to the mere presence of a 2FD. Nor is it reasonable to expect that stabilization can fully remove a 2FD. Rather, Theorem 1 simply points out that either large negative values of ξ_n (serious 2FDs) or large positive values of ζ_n or both, will necessarily produce instabilities in the estimation steps.

4. Ridge regression and condition numbers

It is well known that one method for stabilizing the variance, or dispersion, of the estimated parameters in a least squares regression problem is ridge regression. Unfortunately, there is not always as much guidance in estimating the ridge parameters as one would like. That is, if one chooses a certain ridge estimator based on an improvement in numerical stability, it is not always clear what statistical properties this estimator possesses. This problem was addressed by Casella [12] and to a lesser extent by Belsley et al. [13]. Casella's ideas are reviewed in this section, with the goal of replacing $\mathbf{T}^{-1}\mathbf{XA}$ with a more stable estimator that retains some identifiably good statistical properties.

Consider the regression problem

$$\mathbf{Y}_{N\times 1} = \mathbf{X}_{N\times R}\,\beta_{R\times 1} + \epsilon_{N\times 1}$$

and consider replacing the usual least squares estimate of β, given by $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$ with the more general estimate $\hat{\beta}_{\mathbf{H}} = \mathbf{H}^{-1}\mathbf{X}^t\mathbf{Y}$. If the condition number of $\hat{\beta}_{\mathbf{H}}$ is defined as

$$k(\hat{\beta}_{\mathbf{H}}) = \|\mathbf{H}\|\|\mathbf{H}^{-1}\|, \quad \text{where}$$

$$\|\mathbf{H}\| \equiv \sup_{a^t a = 1} (a^t\mathbf{H}^t\mathbf{H}a)^{1/2}$$

it is easy to check that $\kappa(\hat{\beta}_{\mathrm{OLS}} = \lambda_1/\lambda_R$, where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_R$ are the eigenvalues of $\mathbf{X}^t\mathbf{X}$. However, if $\mathbf{H}$ is taken to be the *ordinary ridge estimator* $\mathbf{X}^t\mathbf{X} + k\mathbf{I}$ for $k > 0$, it is also easy to see that $k(\hat{\beta}_{\mathrm{Ridge}}) = (\lambda_1 + k)/(\lambda_R + k) \leq \lambda_1/\lambda_R$ and the effect of the stabilization is clear. Likewise, if the generalized ridge estimator is defined as

$$\hat{\beta}_{\mathrm{GRidge}} = \mathbf{P}(\mathbf{D}_\lambda + \mathbf{K})^{-1}\mathbf{P}^t$$

where

$$\mathbf{K} = \mathrm{diag}(k_1, k_2, \ldots, k_R), \qquad \mathbf{P}^t\mathbf{T}\mathbf{P}$$

$$= \mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_R) \quad \text{and} \quad \mathbf{P}^t\mathbf{P} = \mathbf{I}$$

then it follows that

$$\kappa\left(\hat{\beta}_{\mathrm{GRidge}}\right) = \max_i\{\lambda_i + k_i\} / \min_i\{\lambda_i + k_i\}$$

and the goal is to choose the k's so that $\kappa(\hat{\beta}_{\mathrm{Ridge}}) \leq \kappa(\hat{\beta}_{\mathrm{OLS}})$.

This choice of ridge parameters to improve conditioning is not a difficult task in and of itself, but as mentioned above, one would like to achieve this improvement while at the same time retaining desirable statistical properties. Casella focused on the property of minimaxity and derived conditions under which a minimax estimator could be chosen to improve conditioning. More precisely, since $\hat{\beta}_{\mathrm{OLS}}$ is already minimax with respect to squared error loss (and the usual dispersion assumptions on ϵ). Casella was motivated to devise a scenario under which either minimaxity could be retained and conditioning would improve or overall risk (mean squared error) could be improved without worsening conditioning. The first scenario is perhaps more natural in the presence of moderate-to-serious instabilities, while the second may be appropriate for mild-to-moderate instabilities. Even though it is well known that the choice of minimaxity as a statistical criterion has its limitations and that the use of ridge-type estimators produces some ac-

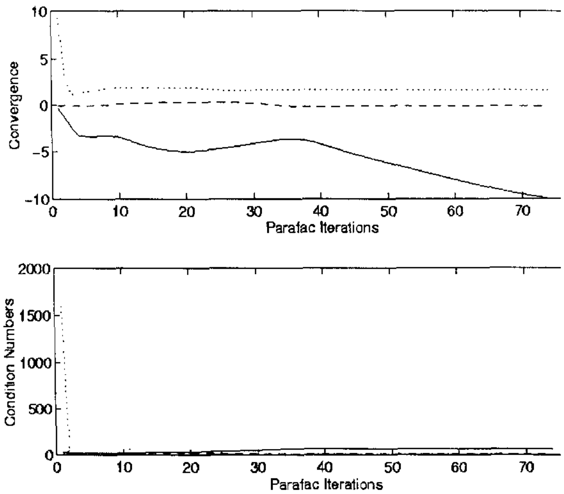

Fig. 4. Stabilized PARAFAC for 341_6.

companying bias, Casella's ideas were adopted in an effort to improve the conditioning of PARAFAC and shorten the number of iterations required for convergence.

The reader is referred to Casella [12] for a full discussion of the class of estimators studied. In particular, it is important to note that it is not always possible to simultaneously produce a condition improving minimax estimator (CIME). In this situation, any minimax estimator in the class studied could worsen conditioning and any estimator that improves conditioning is guaranteed to fail to be minimax. Further, even if a CIME does exist, the improvement in conditioning may not be enough to keep PARAFAC from continually becoming more and more ill-conditioned as the algorithm continues. In light of these contingencies, it was decided that the following common-sense method would be implemented:

— If the condition number of **T** is below some (low) preset tolerance, then an estimator will be produced from Casella's results that reduces overall mean squared error. If a CIME is possible, then this estimator will not worsen the condition number. Otherwise, it could.

— If the condition number of **T** is above some (high) preset tolerance, then a crude stabilization will be employed for that step on that iteration of PARAFAC, only. In the examples which follow, the smallest eigenvalue of **T** was replaced by the second smallest. There is no 'theory' to this, of course, other than the recognition that something radical needs to be done to reduce the standard error of estimation when the condition numbers start to get beyond the control of the formal procedure.

— if the condition number of **T** is between these two tolerances then the focus will be on reducing condition number according to Casella's results, a method less radical than suggested by the second bullet. Of course, if a CIME is possible then minimaxity is retained as well.

This algorithm was programmed in MATLAB (with low tolerance = 500 and high tolerance = 1000) and used in place of the ordinary least squares step in PARAFAC. The results were encouraging and are reported in Section 5.

5. Results of the stabilization

As mentioned above, both, 231_3 and 341_6 were difficult runs. Figs. 4 and 5 show the results obtained when PARAFAC is stabilized.

Notice that the number of iterations spent in the swamp has been greatly reduced. For 231_3 the reduction is from over 6500 to under 300, while for 341_6 the reduction is from around 500 to under 100. It is also evident that the corresponding condition numbers were held in check throughout. Fig. 6 con-

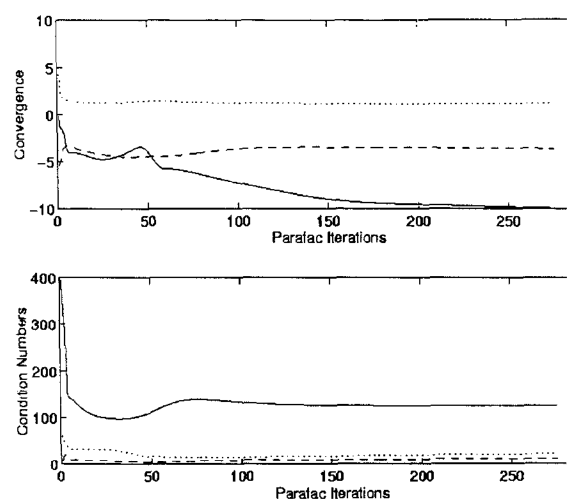

Fig. 5. Stabilized PARAFAC for 231_3.

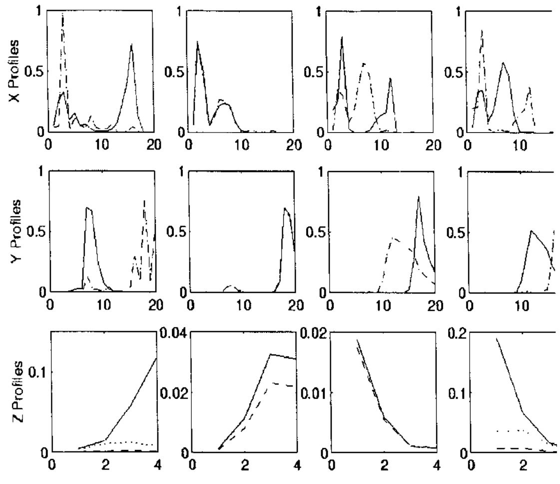

Fig. 6. Profiles for 341_6.

tains the resulting profiles (solid line is the signal, dashed line is original result, and dotted line is the stabilized result).

There are no appreciable differences, overall, in the estimated profiles, although this was not the case for all the synthetic runs. Sometimes the original result would appear to be a little better in some modes, while the stabilized result would perhaps be better in others. One would expect to inherit a certain amount of bias with the stabilized procedure and this was evident in some of the profiles. Finally, Fig. 7 contains the results for 113_6, which had a temporary problem early on with a serious 2FD. Notice that the swamp was diminished along with the corresponding condition numbers.

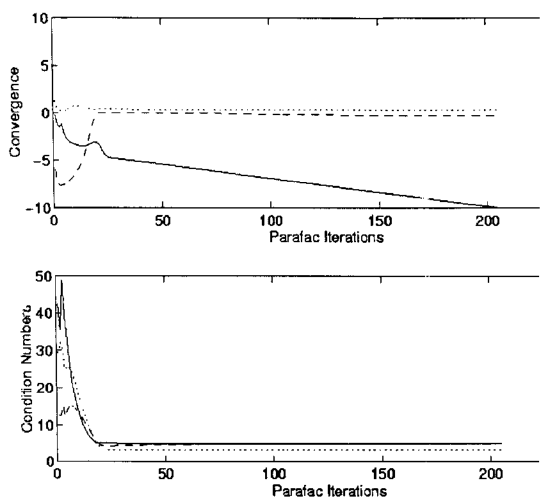

Fig. 7. Stabilized PARAFAC for 113_6.

6. Discussion

As a general rule, the stabilization routine was excellent at reducing the number of PARAFAC runs required for convergence. There are many other ways to stabilize the OLS step in a regression problem and there is nothing particularly special about the routine that was adopted herein. However, Casella's methodology is nicely related to the idea of numerical conditioning and, hence, attractive for the problem at hand. As one would expect, the accumulation of bias is part of the price that has to be paid for using a procedure such as ridge regression. This bias was evident in some mostly minor peak shifts in the resulting resolutions. Perhaps the most important contribution this paper makes is the deceptively simple recognition that PARAFAC will be prone to instabilities in meaningful situations, notably in the presence of severe 2FDs. Hence, it seems imperative that some stabilization method be employed. The results in this paper also suggest that PARAFAC could potentially suffer serious problems when the true signals exhibit high positive UCCs. That is, even if, especially if, the original routine manages to find a path that successfully iterates toward the true signals, this path would likely contain unforeseen obstacles generated by the instabilities in the various estimation steps along the way.

Regardless of whether the true profiles exhibit high correlations or low correlations, one cannot be assured that any particular starting value will produce a path that avoids swamps and iterates quickly to the truth. Clearly one would not want to require the routine to produce orthogonal estimated signals at each iteration, unless there was some reason to believe the true signals were orthogonal. Likewise, starting with an orthogonal array may guarantee one of a 'good start' from a stability point of view, but may produce a path that is eventually repleat with instabilities and if the true signals are significantly correlated, it stands to reason that one may started 'too far' from the truth. So unless the path taken by such a routine can be more fully understood, one is left to react sensibly to obstacles as they arise.

Finally, one point should be made about scale. It is generally accepted that 'scale does not matter' when PARAFAC is being run. What is meant by this, of course, is that the resulting resolutions are unique only up to an arbitrary scale factor. It should not be assumed that the scale is not an issue in how long it might take to achieve convergence. That is, the squared lengths of the profiles will certainly affect the condition numbers of the **T** matrices, *even if the attending profiles are orthogonal within their respective factors*. For example, consider the profiles

$$X_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$Y_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad Y_2 = \begin{pmatrix} \omega \\ 0 \end{pmatrix}$$

It is trivial to show that the condition number of the corresponding **T** matrix is ω^2 which clearly goes to infinity as the length of $\mathbf{Y}_2$ goes to infinity.

This point surfaced in the synthetic data several times, for example in 324_2 (Fig. 8). Even though the condition numbers are not outrageously high, they are inconsistent with ξ_n and ζ_n, both too moderate to suggest any problem with conditioning. However, it is instructive to monitor the squared lengths of the estimated profiles over the PARAFAC iterations (see Fig. 9), where it is clear that one profile (**X**) is dominating the other two (one too small to be visible on

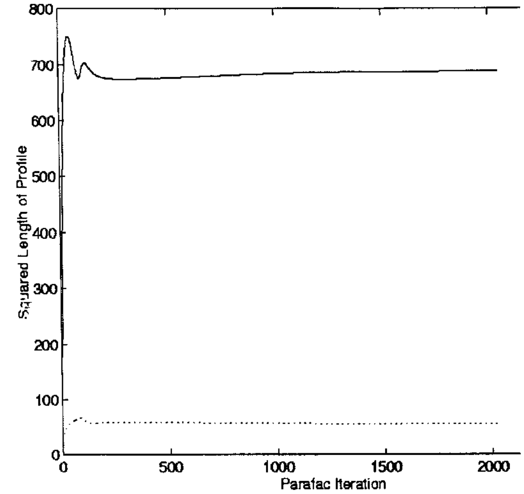

Fig. 9. Squared length of profiles for 324_2.

the plot). In fact the maximum squared length of an **X** profile over all the iterations was 7.51, compared to a maximum of 65 for the **Y** profiles and 0.0031 for the **Z** profiles. There is very little doubt that this problem with length is contributing to the moderate problem with stability. This problem could presumably be controlled by normalizing the factor matrices at each iteration. For illustration, this was implemented along with the mild stabilization discussed above and the results for 324_2 are displayed in Fig. 10. It is clear that the stabilization problem disappeared and the number of required iterations were significantly reduced.

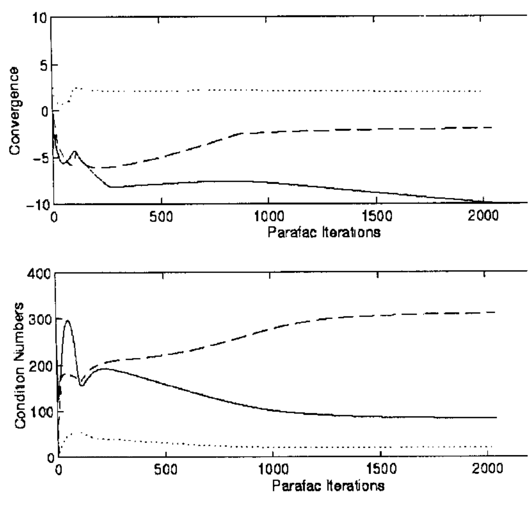

Fig. 8. Original PARAFAC for 324_2.

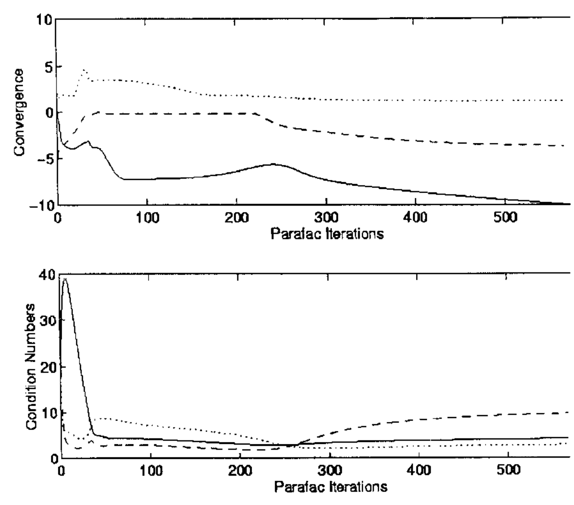

Fig. 10. Normalized PARAFAC for 324_2.

Acknowledgements

During the preparation of this manuscript Professor Rayens was supported by NSF Grant ATM 9108177.

References

[1] D.S. Burdick, X.M. Tu, L.B. McGown, D.W. Millican, Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array, J. Chemom. 4 (1990) 15–28.

[2] S.E. Leurgans, R.T. Ross, Multilinear models: Applications in spectroscopy, Stat. Sci. 3 (1992) 289–319.

[3] S.E. Leurgans, R.T. Ross, R.B. Abel, A decomposition for three-way arrays, SIAM J. Matrix Anal. Appl., to appear.

[4] E. Sanchez, B.R. Kowalski, Tensorial resolution: A direct trilinear decomposition, J. Chemom. 4 (1990) 29–45.

[5] R.A. Harshman, How can I know if it's real? A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling, in: Law, Snyder, Hattie, McDonald (Eds.), Research Methods for Multilinear Data Analysis, Praeger, New York, 1984, pp. 566–591.

[6] R.A. Harshman, Foundations of the PARAFAC Procedure: Models and Conditions for an Exploratory Multi-mode Factor Analysis, UCLA Working Papers in Phonetics, vol. 16, 1970, pp. 1–84.

[7] R.A. Harshman, M.E. Lundy, The PARAFAC model for three-way factor analysis and multidimensional scaling, in: Law, Snyder, Hattie, McDonald (Eds.), Research Methods for Multilinear Data Analysis, Praeger, New York, 1984, pp. 122–215.

[8] C.J. Appellof, E.R. Davidson, Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents, Anal. Chem. 53 (13) (1981) 2053–2056.

[9] J.B. Kruskal, R.A. Harshman, M.E. Lundy, How 3-MFA data can cause degenerate PARAFAC solutions among other relationships, in: Coppi, Bolasco (Eds.), Multiway Data Analysis, Elsevier Science Publishers B.V., Amsterdam, 1989, pp. 115–122.

[10] B.C. Mitchell, D.S. Burdick, An empirical comparison of resolution methods for three-way arrays, Chemom. Intell. Lab. Syst. 20 (1993) 149–161.

[11] B.C. Mitchell, D.S. Burdick, Slowly converging PARAFAC sequences: Swamps and two-factor degeneracies, J. Chemom. 8 (1994) 155–168.

[12] G. Casella, Condition numbers and minimax estimators, J. Am. Stat. Assoc. 80 (391) (1985) 753–758.

[13] D.A. Belsley, E. Kuh, R.E. Welsch, Regression Diagnostics, Wiley, New York, 1980.