Recent developments in candecomp/parafac algorithms
Faber, Bro, Hopke

1 introduction

PARAFAC is used for curve resolution and also for quantitative analysis under the name second order calibration. Second order calibration is often performed using the algorithm generalized rank annihilation method (GRAM) for fitting th parafac model. new algorithms for fitting parafac model appeared in literature. Extensive comparison is made leading to suggestions as to which algorithms to use in particular analysis.

Attention will be given to misunderstandings that are used as a basis for some algorithms. Focus in the comparison is on the quality of the solutions in terms of estimated parameters, in terms of qualitative calibration results and in terms of robustness against over-factoring in the model.

2 Theory

2.1 CANDECOMP/PARAFAC trilinear model

Parafac model developed in 1970 by Carroll and Chang under the name candecomp canonical decomposition and by harshman under the name parafac parallel factor analysis. Very populat in chemometrics

Only least squares fitting of the parafac discussed as well as some alternative less well-defined estimation principles. It is also possible to find the parameters of the model in other ways such as wighted least squares principle. **18,19**

written in terms of frontal slices. Tthe parafac model in terms of fronal slices can be written as

$$R_k = X D_k Y^T + E_k, for k = 1, ..., K$$

Where X (I x F) is the first mode loading matrix with typical elements $x_i f$, Y is the second mode loading matrix Z is the third loading matrix D is a diagonal matrix which contains the kth row of Z on its diagonal. E holds the residual variation of R.

Another notation harranges the 3 way data into a matrix called unoflding matricization. Rep of kolda bader. This rep is not illuminating at first sight but enables specifying the model in ordinary matrix algebra and provides a good starting point for describing certain algorithms compactly.

2.2 Decomposition methods

describe and compare algorithms in next sections. Only ALS handles higher-order arrays, constrained models and weighted least squares loss functions.

2.2.1 ALS

Principle: Separate the optimization problem into conditional subproblems and solve them in least squares sense though established numerical methods. Consists of 3 least squares steps each providing a better estimate of one set of loadings and the overall algorithm, will therefore improve the least squares fit of the model of data. If alg converges to global minimum, least squares model found. Problem of assesing whether this global min is found is interesting but not treated here.

Check source 29 to include line-search to speed up convergence

2.2.2 DIrect trilinear decomposition

DTLD based on solving the GRAM eig-val problem. Gram applicable when only 2 slices in one mode. In gram the loadings in 2 of the modes are found directly from 2 slices of the array. Depending on how the subspaces of the problem are found GRAM can provide a least squares solution to a Tucker model. The implication of the Gram method does not have well defined properties wrt fit. The adequacy of the approx depends on how well the tucker model is in accorance with the parafac model.
In DTLD, the gram method is extended to data with more than 2 slices by generating 2 pseudoslices as differently weighted averages of all the slices. From the 2 p-slices the loadings in 2 modes are found using gram. the loadings in the last mode found directly.the choice of which slabs to use to make p-slabs not always obvious, check papers for details.

2.2.3 Alternating trilinear decomposition

ATLD based on formulation of parafac. Slicewise rep expressed wrt any of the modes. Updates rows of the loading matrix at a time using slices. Similar to ALS.Use of The difference between ATLD and ALS is because the updates is not a least squares update of the rows of a factor matrix. The outcome of the ATLD update is a matrix of size FxF where F is the rank. Iff the matrix is diagonal then updating the kth row of Z with diagonal part will provide a least squares update. However in general the matrix will not be diagonal. Updating all the factor matrices in a sim fation means that the overall algorithm has no well-definied optimization goal and will not nec converge and may end in different solutions.

2.2.4 SWATLD

Similar to ATLD none of the 3 different updates acutally optimize the least squares parafac loss function nor do they optimize the same function. the algorithm does not have any well defined convergence properties.

2.2.5 Pseudo alternating least squares algorithm

PALS based on updating scheme inspired by the update used in ATLD. alternates between loss functions none which coincide with least squares parafac. An overall objective not optimized. Does not belong in this class.

2.2.6 ALternating coupled vectors resolution

ACOVER differs from above algorithms components are not estimated simultaneously but sequintially. It is not clear why the algorithm can be expected to yield a reasonable solution to the parafac problem. In the algorithm provided the orthogonality criterion does not hold in pratice for the estimated parameters

2.2.7 Alternating slice-wise diagonalization

Method similar to ATLD but introduces an initial compression of the data using SVD. the algorithm suffers theoretical shortcomings of above algorithms. penalty term is set the influence of parameter was found to be little in practice.

2.2.8 Alternating coupled matrices resolution

similar to ACOVER in reference 40.

2.2.9 summary

Algorithms above can be described in several ways:
1. least squares als vs ad hoc loss function. ALS fits parafac model in LS sense and gram fits a certain associated model. None of the remaining methods fit the model in least squares sense. The algorithms have many steps that min diff loss functions. Convergence and statistical props not defined.

2. Identifiability conditions. Only als capable of fitting all identified parafac models, other algorithms based on pseudo-inverses of the component matrices, can only fit a subset of the feasible parafac models.
3. Statistical properties. Little known about stat of different algorithms. ALS lower bound for high signal to noise ratios.
4. Extensions of models. ALS is useful in situations where constraints need to be incorporated where weighted loss function is needed where missing values are handled and higher order models required
Look to sources 3,51, 52
None altern can handle these situations directly.

2.3   Evaluation criteria

look at mean squared error of component matrices upon normalizing columns, speed cpu time, lack of fit of data in percentages, sensitivity to overfactoring and predicted ability root mean squared error of calibration.

3   results and discussion

ALS yields more stable results than the alternatives. ALS clearly superior while SWATLD and ASD are the best alternatives. ALS is the slowest DTLD is fastest then ATLD and ASD. With overfactoring the rankings were approximately the same. ALS and PALS are extremely slow here. ALS gives best answer