

# Novel trilinear decomposition algorithm for second order linear calibration

Chen, Wu, Jiang, Yu

## 1 introduction

Algorithms which do not require an accurate estimation of the number of factors important (not factor matrices this means rank). With some collinearity PARAFAC is slow to converge and large data arrays causes iterations to be computationally expensive.

**Proposed remedies are in sources 16-23**

Alsberg and Kvalheim described a method to accelerate the convergence of PARAFAC by compressing high dimensional array which equivalent to CANDELIN approach (17-18-19)

Bro and anderson reduce size of array using a fast Tucher3 algorithm.

All methods reduce size of matrices to speed up iterations.

Also come methods to reduce number of iterations required.

Michell and Burdick 22 suggest avoiding swamping due to degeneracy by stopping a run as soon as it enters a swamp area and randomly starting a new run.

Paatero introduced a penalty term with decreasing impact to accelerate the optimizing procedure

**Kiers -3 step algorithm which greatly reduce the number of iterations required**

All these methods suffer the same problem, requiring accurate estimation of rank.

Method Self-weighted alternating trilinear decomposition (SWATLD)

minimizes 3 objective functions with intrinsic relationship rather than the objective function of PARAFAC.

Method can resist the influence of excess factors and reduce number of iterations.

## 2 Nomenclature

$$[X_{..k} = A_{I \times F} \text{diag}(c_k) B_{J \times F}^T + E_{..k} (k = 1, 2, \dots, K) \dots$$

the kth frontal slice, similar for the other 2 dimensions (or N dimension) and the E are the corresponding slices of the threeway residue array. the diagonal matrices have elements equal to the kth jth and ith row of factor matrices.  $\text{cond}(A)$  2-norm condition number of loading matrix A.

## 3 model and algorithm

**Tri-linear decomposition model first proposed by harshman 3 and Caroll and Chang 4**

Model can be expressed in matrix form as equation 1 listed above.

Also proposed a ALS parafac method in sources to solve problem by successively assuming factors in 2 modes known and estimating the last mode. Follows objective function

$$S(A, B, c_1, \dots, c_K) = \sum_{k=1}^K \|X_{..k} - A \text{diag}(c_k) B^T\|_F^2$$

optimal unbiased estimation makes sense, it suffers from low convergence rate and requires accurate estimation of number of factors. Slow convergence can happen from multicollinearity of data. This means considerable differences in the estimated parameters induce only a small diff in terms of loss function value. A way to speed the optimizing procedure is to change the response behavior of the loss function (by some penalty term). It is an art choosing the penalty parameter.

Another approach is to optimize some other objective functions with favorable response surface instead of parafac. The obj func should have intrinsic relationship with the trilinear model and their solutions should be equivalent to the actual underlying factor matrices.

To have an algorithm with high convergence rate use alternatively optimization the 3 objective functions rather than utilized in parafac

reason for alternatively minimize is they have strong intrinsic relationship and solutions equal to actual loading matrix of error free trilinear model.

It can be expected that alternatively min the 3 functions may avoid possible swamp areas.

SWATLD unique in the way it tries to avoid possible swamp areas by alternatively min 3 functions.

SWATLD may not be a monotonically decreasing one.

## 4 experimental

only tested against parafac, other methods use compression techniques.

## 5 results and discussion

### 5.0.1 The convergence property and accelerating capacity of SWATLD

400 arrays of randomly simulated data were run 10 times by both SWATLD and PARAFAC with random initialization. All 4000 runs of SWATLD converged to a satisfactory result within fewer iterations than PARAFAC. The SWATLD converged very quickly also for randomly simulated data array with severe multicollinearity check table 2. when noise is very low SWATLD takes 37 times less iterations and 35 times less in terms of computation time

All results demonstrate that the unique optimizing scheme of SWATLD can speed the convergence rate through alternatively minimizing three different objective functions with intrinsic relationships to avoid possible swamp areas.

### 5.0.2 property of SWATLD being insensitive to the excess factors used in calculation

For  $N \neq F$  the results of Parafac are linear combinations of the actual underlying factors instead of the actual underlying factors themselves. Increasing  $N$  seems to hardly influence on the results of SWATLD. The excess columns represent noise which can be easily discriminated from the desired columns. SWATLD does not require an accurate estimation of the number of factor in mixtures.

### 5.0.3 Influence of noise on results qualities

low noise perfect consistence with real loading matrices. when noise increased to .01 results still very good. further noise deteriorates quality of results. noise as high as .02 still gives okay results, slightly worse than parafac.

### 5.0.4 the results of HPLC-DAD data array

when number of columns grows in parafac it takes many iterations and does not converge. The SWATLD converged within several seconds longest was 16.7s. insensitive to excess factors in calculations. The accuracy is also very good using SWATLD. All of the parafac results are worse than the swatld results

## 6 conclusions

optimizing for trilinear decomposition proposed contrived for second-order linear calibration. SWATLD properties of fast convergence and insensitive to excess factors.