

# Improving speed of multiway algorithms: compression

## Bro Andersson

### 1 introduction

Annoying aspect of estimating multimodels using ALS is time consumption of algorithms. Way to increase speed is to compress the data array initially then subsequently estimate the model from compressed data. Natural since multiway model is a compression of original data into fewer parameters, implying that the systematic variation in the data is expressible in less than the original number of data points. Hence the model to be estimated should be estimatable from another condensed representation. Since multiway model can be considered a multilinear decomposition preserving the systematic variation in the data, it is useful to use a multilinear decomposition for compression as well.

After estimating parameters in compressed space these then can be transformed to the original space and hopefully provide a good approximate solution to the solution that would be found if estimating the model directly. In second paper will refer to the model used to compress the data as the compression model and the model operating on the compressed array as the analytical model.

Methods before have been introduced called CANDELINC canonical decomposition with linear constraints. In this model only orthonormal bases are allowed but any non-orthonormal basis can be orthogonalized prior to compression w/o loss of information

Stressed by Kiers and Harshman there is no need for special algorithms in the CANDELINC approach. On regresses the data onto the basis, uses any existing multiway algorithms on the compressed array and decompress the result by premultiplying the solution with the projection bases. This only holds for unconstrained models with nonweighted least squares optimization criterion.

In this paper Tucker3 suggested for finding the compression bases as the algorithm is fast and has the property of providing optimal bases in the least squares sense.

The new methods makes multiway modeling faster and more memory efficient. Constraints and weighting schemes are very important in the modeling of compressed arrays.

### 2 Theory

Rank of systematic variation is meant the minimum rank of an appropriate basis for the space spanned by the systematic variation in a particular mode ie the rank if no noise is present.

For the first mode the rank of and a basis for the variable space can be determined from analysing the  $I \times JK$  unfolded matrix  $X$  obtained by concatenating the  $K$  layers of the size  $I \times J$  of  $X$  one after another. Many methods for determining the proper rank, judging the residuals, using cross-validation or methods similar to Malinowski's indicator function. For compression it is not essential to find the exact rank but rather define the rank so large that the correct rank is less than the defined rank. The candelinc model is developed for estimating the multiway models under linear constraints. Theory for CANDELINC model says if parafac model given by  $A$ ,  $B$ , and  $C$  is sought subject to the above constraints then it is only necessary to estimate the smaller linear projection matrices. These matrices can be found by estimating a parafac model on an array of  $Y$  of size  $R_a \times R_b \times R_c$  found by regressing  $X$  onto the orthonormal bases  $U$ ,  $V$  and  $Z$ . If orthonormal bases are bases for the systematic variation then the model from  $Y$  will give the sought solution.

Crucial point in this method is to find good bases for the respective modes. If these are good one would expect the analytical model estimated from the compressed space to be equal to the model estimated from the raw data.

A possible way to find the bases would be from the singular vectors from a svd of the array unfolded for each direction.

A better way is to say that  $U$ ,  $V$  and  $Z$  should be given a least squares estimate of the array  $Y$ . This leads to a set of bases which preserve most of the original variation in the compressed array.  $Y$  corresponds to

the core array in Tucker3 model. A fast tucker3 model is key to a successful compression method. algorithm developed in part 1 [20].

Its important to find a model that doesn't underestimate wrt dimensions. The way to choose the appropriate number of components in the compresison model depends on the data type. Results presented here and another ref indicate that using the same number of factors in the tucker3 model as in the analytical model will work satisfactory in many cases though not all.

## 2.1 Modifications of the compression approach

In situations with one very high dimensional mode one can compress only in high dimensional mode using a basis of dimension  $d$ , the dimension of the smallest mode.

If loading matrices of model are required to be nonnegative this is a problem, as the bounded least squares problem of the uncompressed problem turns into a more general and complicated inequality constrained least squares problem in compressed space. Currently no method able to handle this situation but is being worked on.

## 3 Results

Most important finding is that the analytical model obtained from compressed data is almost always identical to the one obtained from the raw data. The tucker3 based compression is always better than the tucker1 based compression, SVD.

Tucker3 compressed PARAFAC modeling is faster than uncompressed modeling. Especially if model is slightly over-parameterized the gain is large.

Additionally modeling based on tucker3 compression is faster than Tucker1-based compression. Tucker1 estimation of bases is performed on quite large arrays which takes longer.

Tucker3-based compression and calculation is 5 to 80 times cheaper than estimating the model from raw data in terms of flops and 3 to 40 times cheaper wrt speed. Much of the computation is used estimating the Tucker3model. So a faster Tucker3 would mean faster compression. One idea is to use few iterations for tucker3.

## 4 conclusion