

# Tensor decomposition in post-hartree fock methods I. two electron integrals and MP2

Benedikt, Auer, Espig and Hackbusch doi 10.1063/1.3514201

## 1 introduction

Advantage of using decomposed tensors is mainly due to the fact of reduced data handling and decreasing complexity of math operations dealing with decomposed high dimensional tensors. A lot of effort has been devoted to overcome the steep scaling by reducing the number of wavefunction parameters. One ansatz is to use tensor decomposition techniques in order to compress the amount of data.

Cholesky decomposition (CD) can be used to decompose 4-D arrays like the 2-electron integrals into expansions in two dimensional quantities **22-33**

Density fitting/resolution of the identity techniques **38-41** used for approximating 2-electron integrals.

These approaches allow to reduce the comp effort typically by 1 order of magnitude and have been used in DFT for a long time.

For explicit correlated methods **42-47** help overcome problems associated with arising additional multicenter integrals.

decomp tech are applied to energy decomp like Laplace transformed MP2 the complexity of such methods can be reduced to linear scaling.

A further possibility to reduce parameters in post HF calc, is the construction of optimal valence space rep in which the number of orbitals that have to be correlated is sig decreased.

In higher order CC methods SVD techniques have been studied **69,70**

This approach is aimed at reducing the effort to calculate the two-electron integrals but to reduce the scaling of post HF methods. multi-D tensors are rep using an expansion of 1-D quantities. instead of constructing a low rank approx of a high-D tensor a priori, the detour via a trivial decomposition with high rank and a subsequent rank reduction is taken. In the optimal case the rank reduction should yield parameters that have been compressed to their most compact form.

## 2 theory

tensor product approximation and rank reduction

Focus on the CP format because it is advantageous for post HF methods and yields a representing format of minimal dimensionality. Memory requirements are  $4 \cdot n \cdot R$  the complexity of algebraic operations can also be decreased to only linear scaling wrt to the dimension. Disadvantage to CP is that the low rank rep is an iterative scheme that requires the computation and partly storage of all integrals.

In principle the best way to obtain a representation in the CP format is to derive it a priori from the equations that define the high dimensional tensors such that these do not have to be calculated explicitly. If this is not possible one can find a trivial decomposition of this tensor with very high rank that can be reduced with a reduction algorithm.

### rank reduction example

for a 4 dimensional quantity, four vectors are required, three unit vectors holding the first three indices in vector format and the last representing vector containing all values for this special multi-index. In this trivial decomposition the initial rank for the four dimensional tensor is always  $n^3$  To obtain a low rank approximation, this initial quantity can be reduced in rank within a given accuracy  $\epsilon$

$$J(\tilde{A}) = \|A - \tilde{A}\| \geq \epsilon$$

where  $\tilde{A}$  is the new approximated tensor with the reduced rank  $R$  title. This is done by a minimization procedure where the error  $\epsilon$  is the  $l^2$ -norm, to which the low rank approximation will be converged, is given as input to the reduction algorithm.

For solving this minimization problem there are different choices like ALS, modified newton method or accelerated gradient.

#### look to source 77 and 78

In this work gradient method used because it has better convergence and a complexity comparable to the ALS method.

The most critical part of the AG algorithm is the computation of the exact line search parameter  $\alpha_k$  contained in  $R_{\geq 0}$

Given a direction  $\mathbf{D}^k$  a solution of the one-dimensional nonlinear equation

$$p(\alpha_k) = \langle J'(\tilde{A}^k + \alpha \mathbf{D}^k), \mathbf{D}^k \rangle = 0$$

has to be found. this line search is avoided in Armijo-type inexact line search is applied [source 92]

Generally for CP tensor of order d, the function p is a polynomial of degree at most 2d-1. Hence a third order derivative free procedure (3-pg) for finding zeros of a function is applied [source 93]. The 3-pg method is globally linear convergent for a function f in the set  $C^2[X, Y]$  where x,y in the set of real number with  $f(x)f(y) \neq 0$ . the order of convergence is defined by the real root of the polynomial  $t^3 - t^2 - t - 1 \approx 1.8393$  The 3-pg method is equivalent to the newton method for polynomials of degree 3. The complexity of the computation of the gradient is  $O(d R \ln(R + R))$  where d is the dimension of the tensor. This is the most complex part of the method so the cost to perform the minimization is the number of max iterations times the gradient scaling.

The current rank-reduction scheme is organized as follows: 1st a pivoting routine used to find a large entry in the original tensor on a special cross over all dimensions, this is used as a new rank in the representing vectors for the approximation. The new rank is then optimized to improve the A-residual value and is added to the set ranks in  $\tilde{A}$ . The iterative procedure is then used to improve all ranks in the vectors to further lower the difference between  $\tilde{A}$  and A. If the difference is lower than the given threshold parameter e it is stopped and the final number of ranks is obtained, else the cycle starts again and adds another rank to the approximated  $\tilde{A}$ . This way the rank grows one by one during the iterations until the given accuracy is reached.

The scheme is not optimized for the problem of approximating two-electron integrals or wavefunction parameters. There is much room for improvement from a viewpoint of computational efficiency.

### 3 Decomposition of the 2-electron integrals and integral transformations

For a rep in the CP tensor format, the 2-e integrals can be cast into a decomposed form by trivial decomposition, so that the initial rank R scales with  $O(N^3)$  where N is the basis functions.

$$\langle \mu\nu | \sigma\rho \rangle = \sum_{r=1}^R \chi_r^{(\mu)} \otimes \chi_r^{(\nu)} \otimes \chi_r^{(\sigma)} \otimes \chi_r^{(\rho)}$$

After casting the integrals into CP format the AO-MO transformation can be rewritten as the separate transformation of the four representing vectors,

$$v_{cd}^{ab} = \sum_{\mu\nu\sigma\rho} C_\mu^a C_\nu^b C_\sigma^c C_\rho^d \sum_{r=1}^R \chi_r^{(\mu)} \otimes \chi_r^{(\nu)} \otimes \chi_r^{(\sigma)} \otimes \chi_r^{(\rho)}$$

One can separate each piece of the second sum as a contraction over indices from the first sum.

Thus the integral transformation is carried out by simple matrix-vector multiplications of the MO coefficient matrices C with the corresponding representing vectors of the decomposed two-electron integrals in the AO basis.

the scaling of the tranformation is then reduced to  $O(N^5)$  to  $O(N^2R)$ . One important fact is that in principle the rank does not change during the transformation. This means that the compression in the CP tensor format is independent of the basis chosen to represent the two electron integrals. If a low rank

approx is found from canonical orbitals it should have the same rank as for localized or natural orbitals. The low rank approx exists both in the MO and AO basis.

An alternative could be to first transform the AO integrals after trivial decomposition to the MO basis and only reduce the rank of the MO integrals or perform the trivial decomposition and subsequent reduction only for the MO integrals after they have been obtained from any QC software package.

It might be advantageous to derive schemes in which the starting point for the decomposition and rank reduction are representations of the two electron integrals as obtained from schemes RI/DF or Cholesky decomposition.

## 4 MP2 algorithm based on decomposition integrals and denominator

For MP2 technique and decompositions to obtain the initial t2 amplitudes look to paper.

## 5 Computational details

The reduction algorithm from this paper scales with initial rank times final rank and the number of iterations. So for large initial ranks the rank reduction is very time consuming step. Due to the trivial decomposition that is currently used to cast the 2-electron integrals into CP tensor format, the initial rank always scales as  $O(N^3)$ . However the canonical format allows for an efficient parallel procedure by treating large quantities in a different way. The large initial tensor in CP format is split into slices including a fixed number of ranks, the rank reduction is then applied to individual slices, so the reduced slices are obtained, which can be merged again.

To obtain a low rank approximation this procedure can be repeated until the rank does not change any more or a full reduction can be carried out for the merged tensor representation. This procedure can be distributed to multiple processes and also lends itself to a distributed integral direct algorithm. This would also eliminate the need for storing the 2-electron integrals as done in the current pilot implementation.

## 6 results

Three kinds of reductions were tested. A sliced rank reduction described above, a single sliced reduction followed by a full rank reduction and a full rank reduction. The first method gave the least amount of reduction in rank. The one slice and reduction had very large compression even in with high accuracy requirements ( $\|T - Ta\| < 1e-6$ ). The larger the molecules the better the compression. The full reduction has very similar results to the second reduction. the final ranks deviate at most .5 %.

Best balance of reduction and time is the slice reduction followed by full.

Reductions using sliced rank method scale independent of threshold value and the scaling cannot be reduced drastically by a sliced rank reduction. Though the prefactor can be reduced by an order of magnitude.

For slicing methods, large slices generally lead to better reduction of rank and can reduce basis set size scaling.

rank reduction algorithms current state is achieved with slice sizes between 1000 and 2000.

Specific values of integrals in the tensor and reduced tensor are compared, two equivalent integrals had smaller difference than  $1e-7$ , 2 orders of mag smaller than the error in the l2 norm introduced by the low rank approximation.

### AO-MO transform with decomposed integrals

for alkyl chains it can be concluded that the MO integrals can be decreased from  $N^4$  to  $N^2$ . The scaling of the number of parameters for representing the two electron integrals is similar to RI/DF and CD.

**MP2 algorithm with decomposed integrals** Scaling of the reduced t2 amplitude ranks are understood. t2 amps can be reduced to similar or better ranks than the MO integrals. for the least accuracy the scaling with system and basis is almost linear Low scaling with basis set methods often do not show advantageous scaling with basis set size.

MP2 found 2 ways, 1 using reduced AO integrals; 2 t2 amps and MO ints are obtained from conventional calc then tranformed into the CP format and reduced to the given accuracy. The second method gives better accuracy relative to the full MP2 result. The error can be decreased by more than 3 orders of mag by changing the threshold parameter.

## 7 Discussion and Outlook

Using the canonical tensor product format with a low rank approximation one can reduce the storage and computational effort of post -HF ab initio methods drastically. As the rank of the decomposed does not change during AO-MO tranformation it is independent of basis. It is possible to compress the rank of the MO integrals so that the scaling of the rank R' with system size can be reduced further.

Once amps and ints are available in CP format MP2 energy can be calculated using scalar products of corresponding vectors. The resulting energy yields mHatrtee accuracy in the MP2 energy. higher threshold values can produce microhartree accuracy.