

# Regularized Newton Method for efficient approximation of tensors reps in canonical tensor format Espig and Hackbusch

## 1 introduction

There are 2 approximation problems in the canonical tensor format discussed in this paper. The first problem is for a given vector representation with a fixed rank  $r$  want to minimize the difference between the tensor and its CP approximation.

The second problem the roles of  $r$  and  $\epsilon(r)$  are reversed. For a given vector and epsilon looking for minimal  $r$  and  $\hat{x}$  (Ask ed if he can help clarify this) I believe this is the problem I am looking to solve. In paper summarize existing approaching for solving the first problem in finite dimensional tensor space, using ALS algorithm. The minimization problem was also solved by a Gauss-Newton method and a Newton Method.

## 2 Pre-Hilbert tensor product spaces and the canonical tensor format

This paper focuses on algebraic tensors space  $\mathcal{T} := \bigotimes_{\mu=1}^d V_{\mu}$  of arbitrary real-valued pre-hilbert spaces  $(V_1, \langle \cdot, \cdot \rangle_1), \dots, (V_d, \langle \cdot, \cdot \rangle_d)$  for  $d \geq 3$

Definition 2.1 we call a tensor  $\nu \in \mathcal{T}$  an elementary tensor if there exists a set of  $\nu_{\mu}$  such that

$$\nu = \bigotimes_{\mu=1}^d \nu_{\mu}$$

call the  $d$ -tuple of vectors a representation system of  $\nu$ .

Since the set of  $V$  and their inner products are pre-hilbert spaces the tensor product space has induced scalar product. In the induced scalar product of  $\mathcal{T}$  is defined on elementary tensors  $\nu$  and  $w$ , defined as def 2.1 by

$$\langle \nu, w \rangle := \prod_{\mu=1}^d \langle \nu_{\mu}, w_{\mu} \rangle_{\mu}$$

This definition maps  $\mathcal{T} \times \mathcal{T}$  to the real space and is a bilinear form. The norm space also has a norm associated with the induced scalar product.

mapping  $\mathcal{T}$  is called the canonical homomorphism.

Be definition of the tensor space, every tensor in the tensor space can be written as a sum of elementary tensors. Hence the question of minimal number of producing elementary tensors arises. This question leads to the definition of the tensor rank of a tensor.

Definition 2.2 (t-terms, tensor rank, Canonical Tensor Format, Representation system) the tensor rank of  $\nu$  in  $\mathcal{T}$  is

$$\text{rank}_{\mathcal{T}}(\nu) := \min\{r \in \mathbb{N}_0 : \nu \in \mathcal{T}_r\}$$

Canonical tensor format in  $\mathcal{T}$  for variable  $r$  is

$$\nu := \sum_{i=1}^r \bigotimes_{\mu=1}^d \nu_{i\mu}$$

Call the sum of elementary tensor a tensor representation in canonical tensor format with  $r$  terms. The systems of vectors is a representation system of  $\nu$  with representation rank  $r$ .

For a finite dimension each representation has a uniquely defined coefficient  $a$  in the real numbers.

Definition 2.4 coefficient tensor.  $a \in S$  the coefficient tensor of  $v$  wrt the basis. To store the coefficient tensor  $a$  of  $v$  one needs  $\prod_{\mu=1}^d t_\mu$  memory entries. ie the memory req grow exponentially with  $d$ . but for  $v \in T_r$  one only needs  $r * \sum_{\mu=1}^d t_\mu$  entries.

### 3 The canonical tensor format with bounded terms and existence of best approximations

We can say that the existence of the best approximation is not gauranteed and moreover, because of the unboundedness of terms, the numerical treatment is not practicable.

### 4 approximation problem and objective function

In this section the two approximation problem are formulated and the objective function is defined. (For the regularized newton method the first and second derivative of the objective function are stated.)

The set  $R$  is defined in analogy to sums of elementary tensors with bounded terms

Definition Extended approximation problem

given  $\alpha \in S_R$  and  $\epsilon > 0$  find minimal  $r_\epsilon \leq R$  and  $\xi \in R_{d,r_\epsilon,t}^c$  such that

$$\|\alpha - C(\xi_\epsilon)\| \leq \epsilon$$

$$\|\alpha - C(\xi_\epsilon)\| = \min_{\xi \in R} \|\alpha - C(\xi)\|$$

Definition of the objective function

the solution of the extended problem is reduced to a finite sequence of approximation problems from the previous def. First objective function of the approximation with fixed rank  $r$ . The minimization operates on  $R$  wrt the fuction

$$1/2 \|\alpha - .\|^2 \circ C : R \rightarrow \mathbb{R}_{\geq 0}$$

find alphas definiton earlier. The complete objective equation for this problem is equation 28 and 29.

The first and second derivatives are equations 34-44.

I am not sure these equations are necessary since I already have a method for the optimized linear equation at fixed  $r$ .

### 5 solution of the approximation problem

Regularized Newton Problem

All Newton-like methods are based on approximating the objective function locally by a quadratic model and then minimizing that model approximately . The quadratic model of the objective function  $f$  at  $\xi$  in direction  $d$  is given by the taylor polynomial of second order:

$$f(\xi^k + d) \approx q_k(\xi^k) := f(\xi^k) + \langle f'(\xi^k), \xi - \xi^k \rangle + 1/2 \langle f''(\xi^k)(\xi - \xi^k), (\xi - \xi^k) \rangle$$

The successor  $\xi^{k+1}$  is the minimum of the minimization problem iff

$$\xi^{k+1} = \xi^k - d^k$$

and the hessian matrix  $f''(\xi^k)$  is positive definite, where  $d^k$  solves the Newton equation

Computational difficulties arise with the above mentioned method when the function  $f$  is strongly nonlinear. Since our objective fuction  $f$  is non-convex, the hessian is in general not positive definite. Therefore Newtons method will not converge in general. Ways to modified for unconstrained minimization to achieve global convergence.

This is not finished skipping to section 6.4

Solution of the exted approximation problem and systematic choice of initial guesses

The solution of the extended approximation problem

$$\|\alpha - C(\xi_\epsilon)\| \leq \epsilon$$

$$\|\alpha - C(\xi_\epsilon)\| = \min_{\xi \in R} \|\alpha - C(\xi)\|$$

A scheme which solves the extended approximation by the successive use of the regulation Newton method. with the concrete definition of our initial guess, it is ensured that approximation error will not increase. Some of the very useful methods in order to create and impove the inital guess.

Definition 6.7 (Fibre and Cross)

The fibre of  $i$  in direction  $\mu$  is defined as the following set

$$i^\mu := (X_{v=1}^{\mu-1} \{i_v\} X \{1, \dots, t_\mu\} X (X_{v=\mu+1}^d \{i_v\}))$$

The cross  $\kappa^i$  of  $i$  is the union of the fibres in all directions

$$\kappa^i := \bigcup_{\mu=1}^d i^\mu$$

The algorithm 2 creates an initial guess by successively computing rank-one cross approximations. The SCA algorithm produces only rough approximations of a given tensor. A new scheme to improve this approximation introduced based on rank-one approximations. Alternate over the terms of a given approximation and improve the approximation quality by defining the corresponding residual. There are well known methods for the rank one approximation of the tensor, ALS is used here. The algorithm how low numerical complexity compared to the regularized newton method thus it is a good choice for improvement of the initial guess for the regularized newton method. Algorithm 4 is used to solve the exted approximation.

This method still increments  $r$  by 1  
I am stopping reading here at algorithm 4.