Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Optimization Problems in Contracted Tensor Networks

(revised version: April 2012)

by

Mike Espig, Wolfgang Hackbusch, Stefan Handschuh, and Reinhold Schneider

Preprint no.: 66 2011

Optimization Problems in Contracted Tensor Networks

Mike Espig [*] Wolfgang Hackbusch [*] Stefan Handschuh [*] Reinhold Schneider [†]

April 18, 2012

Abstract

We discuss the calculus of variations in tensor representations with a special focus on tensor networks and apply it to functionals of practical interest. The survey provides all necessary ingredients for applying minimization methods in a general setting. The important cases of target functionals which are linear and quadratic with respect to the tensor product are discussed, and combinations of these functionals are presented in detail. As an example, we consider the representation rank compression in tensor networks. For the numerical treatment, we use the nonlinear block Gauss-Seidel method. We demonstrate the rate of convergence in numerical tests.

Keywords: tensor format, tensor representation, tensor network, variational calculus in tensor networks.

1 Introduction

Different tensor formats are of large recent interest and there has been a competition between them in terms of storage and computational efficiency. In this paper, we want to introduce a general approach that covers all these rivaling formats.

We are going to address some general optimization problems such as best approximation, solution of linear systems and minimization of the Rayleigh quotient in high dimensions. Here all tensors are represented in tensor networks. For the numerical treatment we are use the nonlinear block Gauss-Seidel method. Let us start with the description of our problem setting.

Let $(V_\mu, \langle , \rangle_{V_\mu})$ be a real pre-Hilbert spaces and $\mathcal{V} := \bigotimes_{\mu=1}^{d} V_\mu$ equipped with the induced inner product and norm.

Notation 1.1. *Let X be a vector space and $f : X \to \mathbb{R}$. We will use the short notation $\mathfrak{M}(f, X)$ for the set of minimizers of the induced minimization problem, i.e.*

$$\mathfrak{M}(f, X) := \{x \in X : f(x) = \inf f(X)\}. \tag{1}$$

Problem 1.2. *Given a functional $F : \mathcal{V} \to \mathbb{R}$ and a set $\mathcal{M} \subset \mathcal{V}$, we are searching for a minimizer of the constrained optimization problem where the original set $\mathcal{M}$ is confined to tensors which we can represent in a parametrized way, i.e. we are searching for*

$$u \in \mathfrak{M}(F, \mathcal{M} \cap \mathcal{U}), \tag{2}$$

[*]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[†]Technical University Berlin, Germany

where $\mathcal{U} \subset \mathcal{V}$ is the image of a multilinear map $U : P \to \mathcal{V}$. The multilinear map U is called a tensor format from a parameter space P into the tensor product space $\mathcal{V}$, see Definition 2.3 for an explicit description.

We will see that a contracted tensor network is a special tensor representation, see Definition 2.6 for more details. Let us mention a few basic examples which are important in several practical applications in high dimensions.

(i) The approximation of $v \in \mathcal{V}$ in a specific tensor representation, i.e. $F(u) = \|u - v\|^2$, $u \in \mathcal{U}$.

(ii) The solution of equations $Au = b$ or $g(u) = 0$ where $A, g : \mathcal{V} \to \mathcal{V}'$. Here we have $F(u) = \|Au - b\|^2_{\mathcal{V}'}$ resp. $\|g(u)\|_{\mathcal{V}'}$.

(iii) If $A : \mathcal{V} \to \mathcal{V}'$ is bounded, symmetric and coercive with respect to $||.||_{\mathcal{V}}$ and $b \in \mathcal{V}'$ given, we may instead of the first functional in (ii) focus on $F(u) := \frac{1}{2}\langle Au, u\rangle - \langle b, u\rangle$.

(iv) Computation of the lowest eigenvalue of a symmetric operator $A : \mathcal{V} \to \mathcal{V}'$ by minimizing the Rayleigh quotient: $F(u) := \langle Au, u\rangle / \langle u, u\rangle$ over $\mathcal{M} = \mathcal{V}\backslash\{0\}$. This problem is equivalent to the minimization problem
$$\text{find } u \in \mathfrak{M}\big(F,\ \{u \in \mathcal{U} :\ ||u|| = 1\}\big).$$

In the first three examples we have $\mathcal{M} \cap \mathcal{U} = \mathcal{U}$, while in the last example we have an additional constraint, namely $\mathcal{M} = \{W \in \mathcal{V} : \langle W, W\rangle = 1\}$.

The case of interest for our work is summarized in the following abstractly formulated Problem 1.3.

Problem 1.3. *For a given function $F : \mathcal{V} \to \mathbb{R}$ and a tensor format $U : P \to \mathcal{V}$ we consider the following problem:*
$$\textit{find } \mathbf{u} \in \mathfrak{M}(J, M),\ J := F \circ U : P \to \mathcal{V} \to \mathbb{R} \textit{ and } M \subseteq P. \tag{3}$$
We call the function $J : P \to \mathbb{R}$ objective function*.*

2 Mathematical description of tensor formats and tensor networks

A tensor format is described by the parameter space and a multilinear map into the tensor space of higher order. The parameter space consist of two different types of parameters: the parameters of vector space meaning and interior parameters. We will describe this in more details below. Let in the following $\mathcal{V} = \bigotimes_{\mu=1}^{d} V_\mu$ be the tensor product of vector spaces $V_1, \ldots, V_d$.

Notation 2.1. *Let $A \in \{\mathbb{R},\ V_1, \ldots, V_d\}$, $\ell \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, and $\mathbb{N}^\ell := \bigtimes_{\nu=1}^{\ell} \mathbb{N}\ (1 \leq \ell)$. The set of maps with finite support from $\mathbb{N}^\ell$ into A is defined by*
$$\mathcal{M}_0(\mathbb{N}^\ell, A) := \begin{cases} A, & \ell = 0, \\ \big\{u : \mathbb{N}^\ell \to A \,|\, \#\,\mathrm{supp}(u) \in \mathbb{N}\big\}\,, & \ell \geq 1. \end{cases} \tag{4}$$
The natural number ℓ is called the degree *of $u \in \mathcal{M}_0(\mathbb{N}^\ell, A)$. # denotes the cardinality of a set.*

Let us start with an example for pointing out our further intentions.

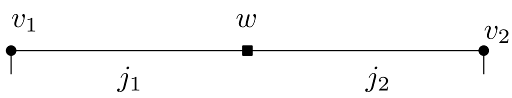


Figure 1: The tensor network graph of the tensor network from Example 2.2.

Example 2.2. *A tensor network is described by its tensor network graph $G = (N, E)$. An example of a tensor network graph is plotted in Figure 1. The set of nodes N contains two different types of nodes, i.e. we have $N = \{v_1, v_2\} \cup \{w\}$. The set for vertices of vector space meaning $\{v_1, v_2\}$ and the set of nodes for the coefficients $\{w\}$, where in Figure 1 the symbol ⫯ stands for nodes of vector space meaning and the symbol ■ denotes vertices for the coefficients. We have two edges $E = \{\{v_1, w\}, \{w, v_2\}\}$ in our example. The tensor network format introduced by the tensor network graph is the following multilinear map:*

$$\begin{aligned} U_G \quad &: \quad \mathcal{M}_0(\mathbb{N}^2, \mathbb{R}) \times \mathcal{M}_0(\mathbb{N}, V_1) \times \mathcal{M}_0(\mathbb{N}, V_2) \to V_1 \otimes V_2 \\ (w, v_1, v_2) \mapsto U_G(w, v_1, v_2) \quad &:= \quad \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \underbrace{w(j_1, j_2)}_{\in \mathbb{R}} \underbrace{v_1(j_1)}_{\in V_1} \otimes \underbrace{v_2(j_2)}_{\in V_2}, \end{aligned}$$

where for a better understanding the edges are identified by there corresponding summation indices, i.e. $j_1 \simeq \{v_1, w\}$ and $j_2 \simeq \{w, v_2\}$. For given so called representation ranks $\underline{r} = (r_1, r_2) \in \mathbb{N}^2$, the tensor network representation $U_{G,\underline{r}}$ introduced by the tensor network format U_G is the restriction of U_G onto $\mathcal{M}_0(\mathbb{N}_{\leq r_1} \times \mathbb{N}_{\leq r_2}, \mathbb{R}) \times \mathcal{M}_0(\mathbb{N}_{\leq r_1}, V_1) \times \mathcal{M}_0(\mathbb{N}_{\leq r_2}, V_2)$, i.e.

$$U_{G,\underline{r}}(w, v_1, v_2) := \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} w(j_1, j_2) v_1(j_1) \otimes v_2(j_2),$$

Notice that the representation rank $\underline{r}$ refers to the support of the representation system (w, v_1, v_2) and not to the represented tensor.

Definition 2.3 (Parameter Space, Tensor Format)**.** *Let $d,\ L \in \mathbb{N}_0$ and furthermore $\ell_1, \ldots, \ell_d, \tilde{\ell_1}, \ldots, \tilde{\ell_L} \in \mathbb{N}_0$. The vector space $\mathcal{S}$ of parameters of vector space meaning for $\mathcal{V}$ is defined by*

$$\mathcal{S} := \bigtimes_{\mu=1}^{d} \mathcal{M}_0(\mathbb{N}^{\ell_\mu}, V_\mu). \tag{5}$$

In a similar way we define the space $\mathcal{C}$ for the interior parameter

$$\mathcal{C} := \bigtimes_{\nu=1}^{L} \mathcal{M}_0(\mathbb{N}^{\tilde{\ell}_\nu}, \mathbb{R}). \tag{6}$$

We call the Cartesian product

$$P_{d,L} = \mathcal{S} \times \mathcal{C} \tag{7}$$

a parameter space *of* order (d, L)*. A* tensor format *of order (d, L) in $\mathcal{V}$ is a multilinear map*

$$U : P_{d,L} \to \mathcal{V} \tag{8}$$

from the parameter space into the tensor space.

We will see in the following that a tensor network is a special tensor format, where the definition of a tensor network is based on the tensor network graph.

Definition 2.4 (Tensor Network Graph, Degree Map). *Let*

$$N_s = \left\{ v_\mu \in \bigcup_{\ell \in \mathbb{N}_0} \mathcal{M}_0(\mathbb{N}^\ell, V_\mu) : 1 \leq \mu \leq d \right\}$$

be a set of nodes of vector space meaning with $\# N_s = d$ *and*

$$N_c = \left\{ w_\nu \in \bigcup_{\ell \in \mathbb{N}_0} \mathcal{M}_0(\mathbb{N}^\ell, \mathbb{R}) : 1 \leq \nu \leq L \right\}$$

be a finite subset of nodes of interior parameters with $\# N_c = L$. *Further let* $N := N_s \cup N_c$ *and* $E \subset \{\{n_1, n_2\} : n_1, n_2 \in N, n_1 \neq n_2\} \subset \mathfrak{P}(N)$ *a set of edges. We call the finite graph* $G := (N, E)$ *a* tensor network graph in $\mathcal{V}$ *of* order (d, L). *The* degree map *of* G *is defined as* $g : N \to \mathbb{N}, n \mapsto \#\{e \in E : n \in e\}$, *such that* g *assigns each element of* N *the number of edges, it is connected to.*

In graph theory there are different ways to describe a graph. For our work, the most useful is the incidence map.

Definition 2.5 (Incidence Map). *Let* $G = (N, E)$ *be a tensor network graph of order* (d, L). *Since we have chosen all tensor network graphs to be finite, we can select an edge enumeration, i.e. there is a bijective map* $e : \mathbb{N}_{\leq m} \to E$, *where* $m := \#E$. *We call the map*

$$\mathcal{I} \quad : \quad N \times \bigtimes_{\ell=1}^{m} \mathbb{N} \to \bigcup_{\ell=1}^{m} \mathbb{N}^\ell \tag{9}$$

$$(n, j_1, \dots, j_m) \quad \mapsto \quad \mathcal{I}(n, \underline{j}) := (j_\ell : 1 \leq \ell \leq m, n \in e(\ell)) \,. \tag{10}$$

the incidence map *of* G, *where the order of the* j_ℓ *is being preserved.*

We will not distinguish between N and $\mathbb{N}_{\leq d+L}$ such that we identify both sets with each other, i.e. there is a bijective map $\varphi : \mathbb{N}_{\leq d+L} \to N$ such that we can uniquely identify $\mu \in \mathbb{N}_{\leq d+L}$ with $n = \varphi(\mu)$. If it is clear from context we simply write μ with the meaning of $\varphi(\mu)$, $(\mu \simeq \varphi(\mu))$. Further, if $1 \leq \mu \leq d$ then $n \in N_s$ and $n \in N_c$ otherwise.

Definition 2.6 (Tensor Network Format, Tensor Network Representation). *Let* $G = (N, E)$ *be a tensor network graph of order* (d, L) *and* $m := \#E$. *Furthermore, let* $\mathcal{I}$ *be the incidence map and* g *the degree map of* G. *We define the following tensor format* U_G *as a* tensor network format in $\mathcal{V}$.

$$U_G : \bigtimes_{\mu=1}^{d} \mathcal{M}_0(\mathbb{N}^{g(\mu)}, V_\mu) \quad \times \quad \bigtimes_{\nu=1}^{L} \mathcal{M}_0(\mathbb{N}^{g(d+\nu)}, \mathbb{R}) \to \mathcal{V} \tag{11}$$

$$(v_1, \dots, v_d, w_1, \dots, w_L) \quad \mapsto \quad \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} \left(\prod_{\nu=1}^{L} w_\nu(\mathcal{I}(d+\nu, \underline{j})) \right) \bigotimes_{\mu=1}^{d} v_\mu(\mathcal{I}(\mu, \underline{j})).$$

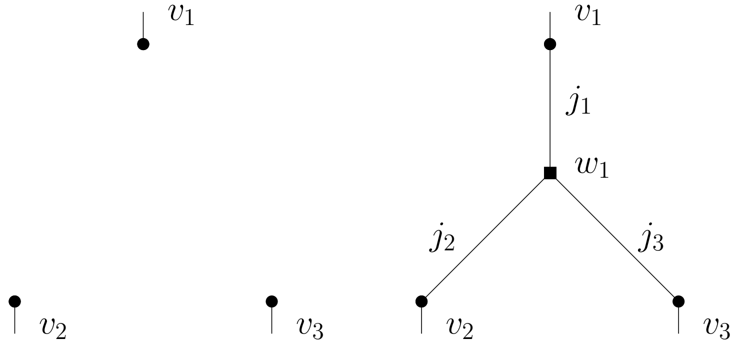
The tensor network representation $U_{G,\underline{r}}$ *with* representation rank $\underline{r} = (r_1, \dots, r_m) \in \mathbb{N}^m$ *is defined as*

$$U_{G,\underline{r}} : \bigtimes_{\mu=1}^{d} \mathcal{M}_0(\mathbb{N}^{g(\mu)}, V_\mu) \quad \times \quad \bigtimes_{\nu=1}^{L} \mathcal{M}_0(\mathbb{N}^{g(d+\nu)}, \mathbb{R}) \to \mathcal{V} \tag{12}$$

$$(v_1, \dots, v_d, w_1, \dots, w_L) \quad \mapsto \quad \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=1}^{L} w_\nu(\mathcal{I}(d+\nu, \underline{j})) \right) \bigotimes_{\mu=1}^{d} v_\mu(\mathcal{I}(\mu, \underline{j})).$$

We say $u = U_{G,\underline{r}}(v_1, \ldots, v_d, w_1, \ldots, w_L) \in \operatorname{Range} U_{G,\underline{r}} \subset \mathcal{V}$ *is represented in the tensor network format with representation rank* $\underline{r} \in \mathbb{N}^m$*. Furthermore, we call the tuple of parameters* $(v_1, \ldots, v_d, w_1, \ldots, w_L)$ *a* representation system *of* u *with representation rank* $\underline{r}$*.*

Note that due to the multilinearity of $U_{G,\underline{r}}$ a representation system is not uniquely determined. We want to illustrate the abstract definition of the tensor network on further examples. The most recent tensor representations are tensor networks, e.g. hierarchical tensor format [8, Hackbusch and Kühn, 2009], [6, Grasedyck, 2010], the tree Tucker format (TT) [19, 14, Oseledets and Tyrtyshnikov, 2009], where the TT tensor format is also called tensor train format. The Tucker decomposition is also a tensor network format, see Figure 2 for illustration. The canonical polyadic decomposition (CP) for tensor ranks greater than one and $d > 2$ is not a tensor network. But, it is easy to illustrate that the canonical polyadic tensor representation for $d = 2$ is a tensor network for any rank.



(a) Elementary Tensor (CP with $r = 1$)

(b) Tucker

Figure 2: The tensor network graph of the canonical polyadic (rank is one) and the Tucker format for $d = 3$.

Example 2.7. *Our first example of a tensor network is the hierarchical tensor format for* $d = 4$*. Where the tensor network graph of order* $(4, 3)$ *is shown in Figure 3. The map* $\mathcal{I}_H : N \times \times_{l=1}^{6} \mathbb{N} \to \bigcup_{l=1}^{6} \mathbb{N}^l$ *is defined*

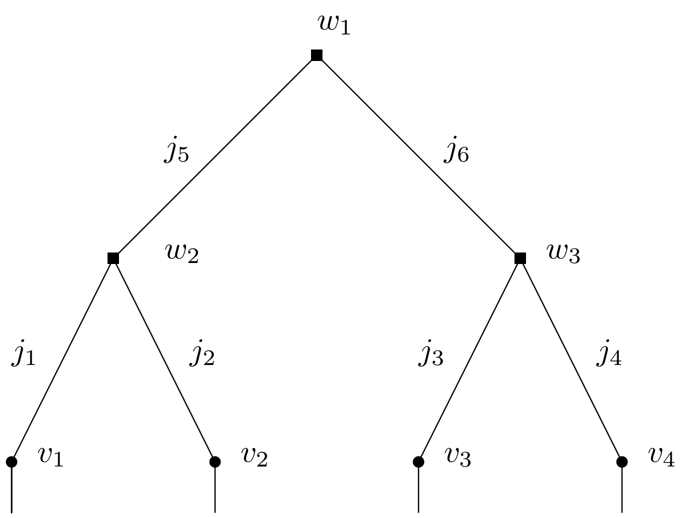


Figure 3: The tensor network graph of the hierarchical tensor format for $d = 4$.

by

$$\mathcal{I}_H(n,(j_1,\ldots,j_6)) := \begin{cases} (j_1), & n=1; \\ (j_2), & n=2; \\ (j_3), & n=3; \\ (j_4), & n=4; \\ (j_5,j_6), & n=5; \\ (j_1,j_2,j_5), & n=6; \\ (j_3,j_4,j_6), & n=7. \end{cases}$$

Furthermore, the multilinear map for the hierarchical tensor format is

$$U_H(v_1,\ldots,w_3) := \sum_{\underline{j}\in\mathbb{N}^6} w_1(j_5,j_6)w_2(j_1,j_2,j_5)w_3(j_3,j_4,j_6)\, v_1(j_1)\otimes v_2(j_2)\otimes v_3(j_3)\otimes v_4(j_4). \tag{13}$$

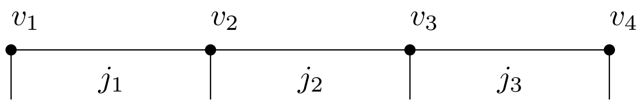


Figure 4: The tensor network graph of the tensor train format for $d=4$.

Next, we want to consider the tensor train format for $d=4$*. The tensor network graph of order* $(4,0)$ *is illustrated in Figure 4. We see that the degree of the nodes* v_1 *and* v_2 *is equal to* 1*. Furthermore, the degree of the nodes* v_2, v_3 *is* 2 *and the number of edges in the graph is* 3*. For this example, the map* $\mathcal{I}_{TT} : N \times \bigtimes_{l=1}^{3} \mathbb{N} \to \bigcup_{l=1}^{3} \mathbb{N}^l$ *is defined by*

$$\mathcal{I}_{TT}(n,(j_1,j_2,j_3)) := \begin{cases} (j_1), & n=1; \\ (j_1,j_2), & n=2; \\ (j_2,j_3), & n=3; \\ (j_3), & n=4. \end{cases}$$

Finally, for the tensor network representation with representation rank $\underline{r} = (r_1,r_2,r_3) \in \mathbb{N}^3$ *we have*

$$U_{TT,\underline{r}}(v_1,\ldots,v_4) := \sum_{j_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{j_3=1}^{r_3} v_1(j_1)\otimes v_2(j_1,j_2)\otimes v_3(j_2,j_3)\otimes v_4(j_3).$$

Another example of a tensor network is the tensor chain (see [11]). The network graph of the tensor chain

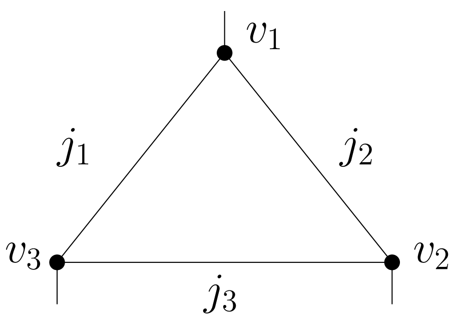


Figure 5: The tensor network graph of the tensor chain for $d=3$.

is presented in Figure 5 for $d = 3$ and the tensor network representation $U_{TC,\underline{r}}$ with representation rank $\underline{r} = (r_1, r_2, r_3) \in \mathbb{N}^3$ is defined by

$$U_{TC,\underline{r}}(v_1, \dots, v_3) := \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \sum_{j_3=1}^{r_3} v_1(j_1, j_2) \otimes v_2(j_2, j_3) \otimes v_3(j_1, j_3). \tag{14}$$

The so called projected entangled-pair states (PEPS) offers an efficient tensor network of certain many-body states of a lattice system, see e.g. [18], [17]. For $d = 6$, the tensor network graph of the PEPS tensor network is shown in Figure 6.

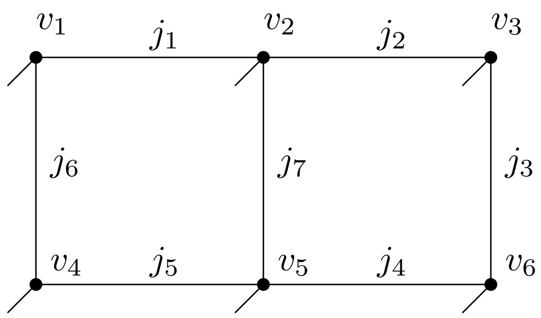


Figure 6: The tensor network graph of the PEPS for $d = 6$.

The multilinear map of the PEPS with equal representation ranks $r \in \mathbb{N}$ is given by

$$U_{PEPS,r}(v_1, \dots, v_6) := \sum_{\underline{j} \in \mathbb{N}^7_{\leq r}} \tilde{v}(\underline{j}) \otimes \hat{v}(\underline{j}), \tag{15}$$

where

$$\begin{aligned} \tilde{v}(\underline{j}) &:= v_1(j_1, j_6) \otimes v_2(j_1, j_2, j_7) \otimes v_3(j_2, j_3), \\ \hat{v}(\underline{j}) &:= v_4(j_5, j_6) \otimes v_5(j_4, j_5, j_7) \otimes v_6(j_3, j_4). \end{aligned}$$

3 Closedness of tensor network formats

The following section is of interest for optimization problems in tensor networks. The main statements of Theorem 3.2 and Proposition 3.4 can be summarized as follows. Assume, we have a sequence $(u_k)_{k \in \mathbb{N}}$ in $\mathcal{V}$ with $\lim_{k \to \infty} u_k = u$ and every u_k is presented in a tensor network $U_G : P_{d,L} \to \mathcal{V}$ with representation rank $\underline{r}$, i.e. there is $\hat{u}_k \in P_{d,L}$ with $u_k = U_{G,\underline{r}}(\hat{u}_k)$ (see Definition 2.6). The crucial question is whether we represent u in $U_{G,\underline{r}}$, i.e. is there $\hat{u} \in P_{d,L}$ such that $u = U_{G,\underline{r}}(\hat{u})$.
In the following let $G := (N, E)$ be a tensor network graph of order (d, L), $m := \#E$, and $U_G : P_{d,L} \to \mathcal{V}$ the tensor network introduced by the network graph G, as described in Definition 2.6.

Definition 3.1 (Closed)**.** *A tensor network format $U_G : P_{d,L} \to \mathcal{V}$ is called* closed*, if for every representation rank $\underline{r} \in \mathbb{N}^m$ the image of the corresponding tensor network representation $U_{G,\underline{r}} : P_{d,L} \to \mathcal{V}$ is a closed set in $(\mathcal{V}, \|\cdot\|)$.*

In order to prove the statement of Theorem 3.2 one needs further assumption on the norm of $(\mathcal{V} \|\cdot\|)$. The norm of $(\mathcal{V}, \|\cdot\|)$ is supposed to be not weaker then the induced injective norm $\|\cdot\|_\vee$, where the injective norm on $\mathcal{V}$ is defined by

$$\|x\|_{\vee} := \sup_{0 \neq v_\mu^* \in V_\mu^*, \mu \in \mathbb{N}_{\leq d}} \left\{ \frac{|(v_1^* \otimes \ldots \otimes v_d^*)(x)|}{\prod_{\mu=1}^d \|v_\mu\|_{V_\mu^*}} : 0 \neq v_\mu \in V_\mu^*, 1 \leq \mu \leq d \right\}, \tag{16}$$

see [5].

Falcó and Hackbusch showed in [5] that the tensor subspace representation is closed. Therefore, the Tucker, the tensor train and the hierarchical tensor format is a closed tensor format, see [7, Chapter 6]. The following Theorem 3.2 shows that arbitrary tree structured tensor networks are closed. The basic idea for the proof is not explained in [5] and [7].

Therorem 3.2. *Let the norm of $(\mathcal{V}, \|\cdot\|)$ be not weaker than $\|\cdot\|_\vee$ and $G = (N, E)$ a tensor network graph. Further, assume that the tensor network graph G is a tree. Then every tensor network U_G introduced by the tree G is a closed tensor format.*

Proof. (Induction over the cardinality of E, $m := \#E$) In order to make notations not more difficult than necessary, we assume that $r = r_1 = \cdots = r_m$. Initial Step: Follows direct from [7, Chapter 6]. Inductive Step: Let $G = (N, E)$ be a tensor network tree with $m + 1 = \#E$ and $\lim_{k\to\infty} U_G(\hat{u}^k) = u \in \mathcal{V}$. Choose an edge $e \in E$. Since G is a tree, the edge e subdivides G into two tensor network sub trees $G_1 = (N_1, E_1)$ and $G_2 = (N_2, E_2)$ with incidence maps $\mathcal{I}_1$, $\mathcal{I}_2$. Where $N_1 = N_{1s} \cup N_{1c}$ and $N_2 = N_{2s} \cup N_{2c}$ are parameter spaces of order (d_1, L_1) and (d_2, L_2) respectively, see Definition 2.3 and Definition 2.4, so the d-th node of G is in G_2. We introduce the following index sets:

$$I_1^c = \{\nu \in \mathbb{N} : w_\nu^k \in N_{1c}\}, I_1^s = \{\nu \in \mathbb{N} : v_\nu^k \in N_{1s}\},\ I_2^c = \{\nu \in \mathbb{N} : w_\nu^k \in N_{2c}\}, I_2^s = \{\nu \in \mathbb{N} : v_\nu^k \in N_{2s}\}.$$

We can assume without loss of generality that the edge e and the enumeration of the notes are chosen such that $e = \{v_{d_1}, v_d\}$. Furthermore, we have for $U_G(\hat{u}^k)$

$$u^k := U_G(\hat{u}^k) \quad = \quad \sum_{j_e=1}^{r} U_{G_1}(\hat{u}_1^k(j_e)) \otimes U_{G_2}(\hat{u}_2^k(j_e)),$$

with

$$U_{G_1}(\hat{u}_1^k(\cdot)) \quad = \quad \sum_{\underline{j} \in \mathbb{N}_{\leq r}^{m_1}} \left(\prod_{\nu \in I_1^c} w_\nu^k(\mathcal{I}_1(\nu, \underline{j})) \right) \bigotimes_{\mu \in I_1^s \setminus \{d_1\}} v_\mu^k(\mathcal{I}_1(\mu, \underline{j})) \otimes v_{d_1}^k((\mathcal{I}_1(\mu, \underline{j}), \cdot)) \in \tilde{\mathcal{V}}_1 \text{ and} \tag{17}$$

$$U_{G_2}(\hat{u}_2^k(\cdot)) \quad = \quad \sum_{\underline{j} \in \mathbb{N}_{\leq r}^{m_2}} \left(\prod_{\nu \in I_2^c} w_\nu^k(\mathcal{I}_2(\nu, \underline{j})) \right) \bigotimes_{\mu \in I_2^s \setminus \{d\}} v_\mu^k(\mathcal{I}_2(\mu, \underline{j})) \otimes v_d^k((\mathcal{I}_2(\mu, \underline{j}), \cdot)) \in \tilde{\mathcal{V}}_2, \tag{18}$$

where $\tilde{\mathcal{V}}_1 := \bigotimes_{\mu \in I_1^s \setminus \{d_1\}} V_\mu \otimes \mathcal{M}_0(\mathbb{N}_{\leq r}, V_{d_1})$ and $\tilde{\mathcal{V}}_2 := \bigotimes_{\mu \in I_2^s \setminus \{d\}} V_\mu \otimes \mathcal{M}_0(\mathbb{N}_{\leq r}, V_d)$.

The tensor space $\mathcal{V}$ is isomorphic to $\mathcal{V}_1 \otimes \mathcal{V}_2$, where $\mathcal{V}_1 = \bigotimes_{\mu \in I_1^s} V_\mu$ and $\mathcal{V}_2 = \bigotimes_{\mu \in I_2^s} V_\mu$. According to [7, Chapter 6], there exist a decomposition of $u = \lim_{k\to\infty} u^k$ in $\mathcal{V}_1 \otimes \mathcal{V}_2$ such that

$$u = \sum_{i=1}^{r'} u_1(i) \otimes u_2(i), \quad (r' \leq r) \tag{19}$$

with smallest sets $U_1 := \{u_1(i) \in \mathcal{V}_1 : 1 \leq i \leq r'\}$ and $U_2 := \{u_2(i) \in \mathcal{V}_2 : 1 \leq i \leq r'\}$ linearly independent. It remains to show that there are parameters $\hat{u}_1(\cdot)$ and $\hat{u}_2(\cdot)$ such that $u_1(\cdot) = U_{G_1}(\hat{u}_1(\cdot))$ and $u_2(\cdot) = U_{G_2}(\hat{u}_2(\cdot))$.

Let $U_1' = \{u_1'(i) \in \mathcal{V}_1' : 1 \leq i \leq r'\}$ and $U_2' = \{u_2'(i) \in \mathcal{V}_2' : 1 \leq i \leq r'\}$ be the dual basis of U_1 and U_2. In [7, Chapter 6] it is shown that $(\mathrm{id}_{\mathcal{V}_1} \otimes u_2'(i))(u^k) \xrightarrow[k\to\infty]{} u_1(i)$ and $(u_1'(i) \otimes \mathrm{id}_{\mathcal{V}_2})(u^k) \xrightarrow[k\to\infty]{} u_2(i)$ for all $1 \leq i \leq r'$. After short calculation (using tensor contractions) we have that

$$(\mathrm{id}_{\mathcal{V}_1} \otimes u_2'(\cdot))(U_G(\hat{u}^k)) = \sum_{\underline{j}\in\mathbb{N}^{m_1}_{\leq r}} \left(\prod_{\nu\in I_1^c} w_\nu^k(\mathcal{I}_1(\nu,\underline{j}))\right) \bigotimes_{\mu\in I_1^s\setminus\{d_1\}} v_\mu^k(\mathcal{I}_1(\mu,\underline{j})) \otimes \tilde{v}_{d_1}^k((\mathcal{I}_1(\mu,\underline{j}),\cdot)), \tag{20}$$

$$(u_1'(\cdot) \otimes \mathrm{id}_{\mathcal{V}_2})(U_G(\hat{u}^k)) = \sum_{\underline{j}\in\mathbb{N}^{m_2}_{\leq r}} \left(\prod_{\nu\in I_2^c} w_\nu^k(\mathcal{I}_2(\nu,\underline{j}))\right) \bigotimes_{\mu\in I_2^s\setminus\{d\}} v_\mu^k(\mathcal{I}_2(\mu,\underline{j})) \otimes \tilde{v}_{d}^k((\mathcal{I}_2(\mu,\underline{j}),\cdot)), \tag{21}$$

where we define $\tilde{v}_{d_1}^k((\mathcal{I}_1(\mu,\underline{j}),\cdot)) := \sum_{j_e=1}^r u_2'(\cdot)(U_{G_2}(u_2^k(j_e)))v_{d_1}^k((\mathcal{I}_1(\mu,\underline{j}),j_e)) \in \mathcal{M}_0(\mathbb{N}_{\leq r}, V_{d_1})$ and $\tilde{v}_d^k((\mathcal{I}_2(\mu,\underline{j}),\cdot)) := \sum_{j_e=1}^r u_1'(\cdot)(U_{G_1}(u_1^k(j_e)))v_d^k((\mathcal{I}_2(\mu,\underline{j}),j_e)) \in \mathcal{M}_0(\mathbb{N}_{\leq r}, V_d)$. Comparing the equations (17) and (18) with (20) and (21), we see that there are parameters $\hat{u}_1^k(\cdot)$ and $\hat{u}_2^k(\cdot)$ such that

$$\begin{aligned}(\mathrm{id}_{\mathcal{V}_1} \otimes u_2'(\cdot))(U_G(\hat{u}^k)) &= U_{G_1}(\hat{u}_1^k(\cdot))\\ (u_1'(\cdot) \otimes \mathrm{id}_{\mathcal{V}_2})(U_G(\hat{u}^k)) &= U_{G_2}(\hat{u}_2^k(\cdot)).\end{aligned}$$

Note that this is only possible if the network graph G is a tree. Since G_1 and G_2 are tensor network trees in $\tilde{\mathcal{V}}_1$ and $\tilde{\mathcal{V}}_2$, the induction hypothesis shows that there are parameters $\hat{u}_1(\cdot)$ and $\hat{u}_2(\cdot)$ such that $u_1(\cdot) = U_{G_1}(\hat{u}_1(\cdot))$ and $u_2(\cdot) = U_{G_2}(\hat{u}_2(\cdot))$. With Eq. (19) we finally have

$$u = \sum_{j_e=1}^{r'} u_1(j_e) \otimes u_2(j_e) = \sum_{j_e=1}^{r'} U_{G_1}(\hat{u}_1(j_e)) \otimes U_{G_2}(\hat{u}_2(j_e)) = U_G(\hat{u}),$$

where $\hat{u} := (\hat{u}_1, \hat{u}_2) \in P_{d,L}$ and $r' \leq r$.

■

If the tensor network graph G is not a tree (it contains cycles), then the induced tensor network U_G is in general not closed, see [12, Landsberg et al., 2011]. We want to mention that in the interesting case if $\dim(V_\mu) \leq 3$ (calculations in the second quantization of quantum mechanics), the analysis in [12] make no statement about the closedness of tensor network formats. If the tensor representation would be stable, we can ensure closedness.

Definition 3.3 (Stable). *Let $U_G : P \to \mathcal{V}$ be a tensor network in $\mathcal{V}$ and $\underline{r} \in \mathbb{N}^m$ a representation rank, where $G = (N, E)$ is a tensor network graph and $m := \#E$.*

(a) For $\hat{u} \in P$ we define

$$\chi_{U_G}(\hat{u}, r) := \frac{1}{\|U_G(\hat{u})\|} \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=d}^{L+d} |w_\nu(\mathcal{I}(\nu,\underline{j}))|\right) \prod_{\mu=1}^{d} \|v_\mu(\mathcal{I}(\mu,\underline{j}))\|.$$

(b) For $u \in \mathrm{Range}\,(U_G)$ we set

$$\chi_{U_G}(u, r) := \inf\left\{\chi_{U_G}(\hat{u}, r) : u = U_G(\hat{u})\right\}.$$

(c) The sequence $(u^k)_{k\in\mathbb{N}} \subset \mathrm{Range}\,(U_{G,\underline{r}})$ *is called* stable in $\mathrm{Range}\,(U_{G,\underline{r}})$*, if*

$$\chi_{U_G}((u^k)_{k\in\mathbb{N}}, r) := \sup_{k\in\mathbb{N}} \chi_{U_G}(u^k, r) < \infty;$$

otherwise, the sequence is called instable.

Proposition 3.4. *Let* $\mathcal{V} = \bigotimes_{\mu=1}^d V_\mu$ *and suppose that* $\dim V_\mu \in \mathbb{N}$*. Furthermore, let* $G = (N, E)$ *be a tensor network graph and* $U_{G,\underline{r}} : P \to \mathcal{V}$ *a tensor representation with representation rank* $\underline{r}$*. If a sequence* $(u^k)_{k\in\mathbb{N}} \subset \mathrm{Range}\,(U_{G,\underline{r}})$ *is stable and convergent, then* $\lim_{k\to\infty} u^k \in \mathrm{Range}\,(U_{G,\underline{r}})$.

Proof. Let $(u^k)_{k\in\mathbb{N}} \subset \mathrm{Range}\,(U_{G,\underline{r}})$ with $\lim_{k\to\infty} u^k$ and set $c := 2\chi_{U_G}((u^k)_{k\in\mathbb{N}}, r)$. After choosing a subsequence, $\lim_{k\to\infty} u^k = u$ holds with a representation system $\hat{u}^k := (w_1^k, \ldots, w_L^k, v_1^k, \ldots v_d^k)^t \in P$ such that

$$\sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=d}^{L+d} |w_\nu^k(\mathcal{I}(\nu, \underline{j}))| \right) \prod_{\mu=1}^{d} \|v_\mu^k(\mathcal{I}(\mu, \underline{j}))\| \le c\|u\|.$$

The components of the parameter space $w_\nu^k(\mathcal{I}(\nu, \underline{j}))$ and $v_\mu^k(\mathcal{I}(\mu, \underline{j}))$ can be scaled equally so that all $\{w_\nu^k(\mathcal{I}(\nu, \underline{j})) \in \mathbb{R} : k \in \mathbb{N}\}$ and $\{v_\mu^k(\mathcal{I}(\mu, \underline{j})) \in V_\mu : k \in \mathbb{N}\}$ are uniformly bounded. Choosing furthermore a subsequence, limits $\tilde{w}_\nu(\mathcal{I}(\nu, \underline{j})) := \lim_{k\to\infty} w_\nu^k(\mathcal{I}(\nu, \underline{j}))$ and $\tilde{v}_\mu(\mathcal{I}(\mu, \underline{j})) := \lim_{k\to\infty} v_\mu^k(\mathcal{I}(\mu, \underline{j}))$ exists and with the continuity of U_G it follows that $\lim_{k\to\infty} u^k = U_{G,\underline{r}}(\tilde{u})$, where $\tilde{u} := (\tilde{w}_1, \ldots, \tilde{w}_L, \tilde{v}_1, \ldots, \tilde{v}_d)^t \in P$. ■

4 Computation of derivatives in tensor representations

We would like to find a local minimizer by means of differential calculus in an arbitrary tensor format. Let $P := \bigtimes_{\nu=1}^{d+L} P_\nu$ a parameter space of order (d, L) and $U : P \to \mathcal{V}$ a tensor format. Before we can start with the computation of the derivatives, we need to introduce the following useful notation.

Notation 4.1. *Let* $D := d + L$, $\nu \in \mathbb{N}_{\le D}$ *and* $\hat{p} := (p_1, \ldots, p_D) \in P$*. We define the following substitution*

$$U_\nu(\hat{p}) : P_\nu \to \mathcal{V}, \quad u \mapsto U_\nu(\hat{p})(u) := U(p_1, \ldots, p_{\nu-1}, u, p_{\nu+1}, \ldots, p_D). \tag{22}$$

The Fréchet derivative $U'(\hat{p})$ of U at $\hat{p} \in P$ is a linear mapping from P to $\mathcal{V}$. Due to the multilinearity of U, it may be expressed by the partial derivatives of U in direction $p_\nu \in P_\nu$ which we will denote by $dU(\hat{p})/dp_\nu \in L(P, \mathcal{V}) := \{f : P \to \mathcal{V} : f \text{ is a homomorphism}\}$. The mapping $dU(\hat{p})/dp_\nu$ maps $u \in P_{\nu_1}$ to

$$\frac{dU(\hat{p})}{dp_{\nu_1}}(u) \;=\; \lim_{h\to 0} \frac{U_{\nu_1}(\hat{p})(p_{\nu_1} + hu) - U_{\nu_1}(\hat{p})(p_{\nu_1})}{h} = U_{\nu_1}(\hat{p})(u).$$

Corollary 4.2. *Let* U *be a tensor network as defined in Definition 2.6. For the partial derivatives we have*

$$\frac{dU(\hat{v}, \hat{w})}{dv_{\mu_1}}(u) \;=\; \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} \underline{w}(\underline{j}) \left(\bigotimes_{\mu=1}^{\mu_1-1} v_\mu(\mathcal{I}(\mu, \underline{j})) \right) \otimes u(\mathcal{I}(\mu_1, \underline{j})) \otimes \left(\bigotimes_{\mu=\mu_1+1}^{d} v_\mu(\mathcal{I}(\mu, \underline{j})) \right), \tag{23}$$

$$\frac{dU(\hat{v}, \hat{w})}{dw_{\nu_1}}(u) \;=\; \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} \prod_{\nu=1,\, \nu\neq\nu_1}^{L} w_\nu(\mathcal{I}(d+\nu, \underline{j}))\, \underline{v}(\underline{j}) u(\mathcal{I}(\nu_1, \underline{j})), \tag{24}$$

where $\underline{w}(\underline{j}) := \prod_{\nu=1}^{L} w_\nu(\mathcal{I}(d+\nu, \underline{j}))$ *and* $\underline{v}(\underline{j}) := \bigotimes_{\mu=1}^{d} v_\mu(\mathcal{I}(\mu, \underline{j}))$.

Corollary 4.3. *By the chain rule, the Fréchet derivative of the functional* $J := F \circ U : P \to \mathbb{R}$ *from (3) at point* $\hat{u} \in P$ *is given by*

$$J'(\hat{u}) = F'(U(\hat{u})) \circ U'(\hat{u}). \tag{25}$$

5 Tensor product subspaces and best approximation in tensor networks

Let $G = (N, E)$ be a tensor network graph of order (d, L) in $\mathcal{V}$ and $m := \#E$, where $\mathcal{V}$ is the tensor product of pre-Hilbert spaces $\left(V_v, \langle , \rangle_\mu\right)$. Furthermore, we define the two tensor network representations $U_R : P_R \to \mathcal{V}$ and $U_r : P_r \to \mathcal{V}$ introduced by G with representation ranks $R = (R_1, \dots, R_m)^t \in \mathbb{N}^m$ and $r = (r_1, \dots, r_m)^t \in \mathbb{N}^m$ respectively, where we have

$$r_l \leq R_l, \tag{26}$$

for all $1 \leq l \leq m$. Moreover, let $a \in \mathcal{V}$ be represented in U_R, i.e. there is $p_R = (v_{R,1}, \dots, w_{R,L}) \in P_R$ with

$$a = U_R(p_R) = \sum_{i_1=1}^{R_1} \cdots \sum_{i_m=1}^{R_m} \left(\prod_{\nu=1}^{L} w_{R,\nu}(\mathcal{I}(d+\nu, \underline{i})) \right) \bigotimes_{\mu=1}^{d} v_{R,\mu}(\mathcal{I}(\mu, \underline{i})). \tag{27}$$

In this section we are analyzing the following minimization problem.

Problem 5.1 (Representation Rank Minimization)**.** *For given representation ranks* R, r *with 26, find* $p_r^* \in P_r$ *such that*

$$\|U_R(p_R) - U_r(p_r^*)\|_{\mathcal{V}} = \inf_{p_r \in P_r} \|U_R(p_R) - U_r(p_r)\|_{\mathcal{V}}. \tag{28}$$

For a convenient description of our results we need an edge enumeration of the tensor network graph $G = (N, E)$, i.e. a bijective map $e : E \to N_{\leq m}$ from the set of edges E to the set $\mathbb{N}_{\leq m}$.

Therorem 5.2. *Let* $p_r^* = (v_{r,1}^*, \dots, w_{r,L}^*) \in P_r$ *be a solution of the representation rank minimization problem (28) and* g *the degree map of* G *as defined in Definition 2.4. Then we have for all* $\mu \in \mathbb{N}_{\leq d}$ *and all* $\underline{j} \in \times_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$

$$v_{r,\mu}^*(\underline{j}) \in U_\mu := \operatorname{span}\left\{ v_{R,\mu}(\underline{i}) \in V_\mu : \underline{i} \in \bigtimes_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq R_{e(l)}} \right\}. \tag{29}$$

Proof. Assume there is a $\mu^* \in \mathbb{N}_{\leq d}$ and a $\underline{j} \in \times_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$ with $v_{r,\mu}^*(\underline{j}^*) \notin U_\mu$. Let $\mathcal{N}_\mu : V_\mu \to U_\mu$ be the orthonormal projection from V_μ onto U_μ. Then it is straightforward to show that $\mathcal{N} : \bigotimes_{\mu=1}^{d} V_\mu \to \bigotimes_{\mu=1}^{d} U_\mu$ is the orthonormal projection from $\bigotimes_{\mu=1}^{d} V_\mu$ onto $\bigotimes_{\mu=1}^{d} U_\mu$. After a short calculation, we have

$$\begin{aligned} \|U_R(p_R) - U_r(p_r^*)\|_{\mathcal{V}}^2 &= \|U_R(p_R) - \mathcal{N}U_r(p_r^*)\|_{\mathcal{V}}^2 + \|U_r(p_r^*) - \mathcal{N}U_r(p_r^*)\|_{\mathcal{V}}^2 \\ &> \|U_R(p_R) - \mathcal{N}U_r(p_r^*)\|_{\mathcal{V}}^2 \end{aligned}$$

since because of $v_{r,\mu}^*(\underline{j}^*) \notin U_\mu$ we can conclude $\|U_r(p_r^*) - \mathcal{N}U_r(p_r^*)\|_{\mathcal{V}} > 0$. Furthermore, we have

$$\begin{aligned} \mathcal{N}U_r(p_r^*) &= \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=1}^{L} w_{r,\nu}^*(\mathcal{I}(d+\nu, \underline{j})) \right) \bigotimes_{\mu=1}^{d} \underbrace{\mathcal{N}_\mu v_{r,\mu}^*}_{\hat{v}_{r,\mu}^* :=}(\mathcal{I}(\mu, \underline{j})) \\ &= U_r(\hat{p_r}^*), \end{aligned}$$

where $\hat{p_r}^* := (\hat{v}^*_{r,1}, \ldots, \hat{v}^*_{r,d}, w^*_{r,1}, \ldots, w^*_{r,L}) \in P_r$. Consequently $\|U_R(p_R) - U_r(\hat{p}^*_r)\|_{\mathcal{V}} < \|U_R(p_R) - U_r(p^*_r)\|_{\mathcal{V}}$, but this contradicts the fact that $\|U_R(p_R) - U_r(p^*_r)\|_{\mathcal{V}} = \inf_{p_r \in P_r} \|U_R(p_R) - U_r(p_r)\|_{\mathcal{V}}$. ■

Under the notations and premises of Theorem 5.2, let $\{z_{l\mu} \in U_\mu : l \in \mathbb{N}_{\leq t_\mu}\}$ be an orthonormal basis of U_μ, where we set $t_\mu := \dim U_\mu$. If we are looking for a solution of the Problem 5.1, with the use of Theorem 5.2, we can restrict our search to $\mathcal{U} := \bigotimes_{\mu=1}^d U_\mu$. Therefore, there are $\alpha_{R,\mu}(\underline{i}) \in \mathbb{R}^{t_\mu}$ and $\xi_{r,\mu}(\underline{j}) \in \mathbb{R}^{t_\mu}$ such that

$$v_{R,\mu}(\underline{i}) = \sum_{l_\mu=1}^{t_\mu} (\alpha_{R,\mu}(\underline{i}))_{l_\mu} z_{l_\mu \mu} \quad \text{and} \quad v_{r,\mu}(\underline{j}) = \sum_{l_\mu=1}^{t_\mu} (\xi_{r,\mu}(\underline{j}))_{l_\mu} z_{l_\mu \mu}.$$

These equations induce a linear mapping $Z_\mu : \mathbb{R}^{t_\mu} \to U_\mu$ with

$$v_{R,\mu}(\underline{i}) = Z_\mu \alpha_{R,\mu}(\underline{i}) \quad \text{and} \quad v_{r,\mu}(\underline{j}) = Z_\mu \xi_{r,\mu}(\underline{j}),$$

where $\underline{i} \in \times_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq R_{e(l)}}$ and $\underline{j} \in \times_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$. Furthermore, we have

$$\begin{aligned} U_R(p_R) &= \sum_{i_1=1}^{R_1} \cdots \sum_{i_m=1}^{R_m} \left(\prod_{\nu=1}^{L} w_{R,\nu}(\mathcal{I}(d+\nu, \underline{i})) \right) \bigotimes_{\mu=1}^{d} v_{R,\mu}(\mathcal{I}(\mu, \underline{i})) \\ &= \sum_{i_1=1}^{R_1} \cdots \sum_{i_m=1}^{R_m} \left(\prod_{\nu=1}^{L} w_{R,\nu}(\mathcal{I}(d+\nu, \underline{i})) \right) \bigotimes_{\mu=1}^{d} Z_\mu \alpha_{R,\mu}(\mathcal{I}(\mu, \underline{i})) \\ &= \left(\bigotimes_{\mu=1}^{d} Z_\mu \right) \underbrace{\left(\sum_{i_1=1}^{R_1} \cdots \sum_{i_m=1}^{R_m} \left(\prod_{\nu=1}^{L} w_{R,\nu}(\mathcal{I}(d+\nu, \underline{i})) \right) \bigotimes_{\mu=1}^{d} \alpha_{R,\mu}(\mathcal{I}(\mu, \underline{i})) \right)}_{\hat{U}_R(\hat{p}_R) :=} \end{aligned}$$

and

$$\begin{aligned} U_r(p_r) &= \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=1}^{L} w_{r,\nu}(\mathcal{I}(d+\nu, \underline{j})) \right) \bigotimes_{\mu=1}^{d} v_{r,\mu}(\mathcal{I}(\mu, \underline{j})) \\ &= \left(\bigotimes_{\mu=1}^{d} Z_\mu \right) \underbrace{\left(\sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \left(\prod_{\nu=1}^{L} w_{r,\nu}(\mathcal{I}(d+\nu, \underline{j})) \right) \bigotimes_{\mu=1}^{d} \xi_{r,\mu}(\mathcal{I}(\mu, \underline{j})) \right)}_{\hat{U}_r(\hat{p}_r) :=}. \end{aligned}$$

Corollary 5.3. *From the definition of $\hat{U}_R$ and $\hat{U}_r$ it is obvious that $\hat{U}_R$ and $\hat{U}_r$ are tensor networks in $\mathcal{S} := \bigotimes_{\mu=1}^d \mathbb{R}^{t_\mu}$, where the network topology is the same as for U_R and U_r respectively, since the incidence map is the same for all tensor networks. Further, we have*

$$U_R(p_R) = Z\hat{U}_R(\hat{p}_R) \quad \text{and} \quad U_r(p_r) = Z\hat{U}_r(\hat{p}_r), \tag{30}$$

where we set $Z : \mathcal{S} \to \mathcal{U}$, $Z := \bigotimes_{\mu=1}^d Z_\mu$. In addition

$$\|U_R(p_R) - U_r(p_r)\|^2_{\mathcal{V}} = \left\langle \hat{U}_R(\hat{p}_R) - \hat{U}_r(\hat{p}_r), Z^t Z(\hat{U}_R(\hat{p}_R) - \hat{U}_r(\hat{p}_r)) \right\rangle_{\mathcal{S}} = \|\hat{U}_R(\hat{p}_R) - \hat{U}_r(\hat{p}_r)\|^2_{\mathcal{S}}. \tag{31}$$

Corollary 5.4. *Under the notations and premises of Theorem 5.2, we have that*

$$\|U_R(p_R) - U_r(p_r^*)\|_{\mathcal{V}} = \inf_{p_r \in P_r} \|U_R(p_R) - U_r(p_r)\|_{\mathcal{V}} \tag{32}$$

is equivalent to

$$\|\hat{U}_R(\hat{p}_R) - \hat{U}_r(\hat{p}_r^*)\|_{\mathcal{S}} = \inf_{\hat{p}_r \in \hat{P}_r} \|\hat{U}_R(\hat{p}_R) - \hat{U}_r(\hat{p}_r)\|_{\mathcal{S}}. \tag{33}$$

For this reason, it is sufficient to consider the original approximation only in $\mathcal{S}$. Hereby we have to assume that in practice the computation of the orthonormal basis of U_μ and the coefficients α_R is reasonable. This fact reduces the original potentially infinite dimensional approximation to a finite minimization task.

6 Nonlinear block Gauss-Seidel method

So far we have developed all ingredients for applying steepest decent type algorithms. In the following section let $P = \times_{\mu=1}^{D} P_\mu$ be a parameter space of order (d, L), where $D := d + L$, and $U : P \to \mathcal{V}$ a tensor representation. Further, let $J := F \circ U : P \to \mathcal{V} \to \mathbb{R}$ be an objective function as defined in Problem 1.3. In the following analysis, it is not required that U is a tensor network.

The nonlinear Gauss-Seidel (GS) method arises from iterative methods used for linear systems of equations. Intuitively, we may think of a generalized linear method which reduces to a feasible iteration for nonlinear systems. The direct extension of the linear Gauss-Seidel method to the nonlinear GS method is obvious. Suppose that the k-th iterate $x^k = (x_1^k, \ldots, x_n^k)^T$ and the first $l-1$ components $x_1^{k+1}, \ldots, x_{l-1}^{k+1}$ of the $(k+1)$-th iterate x^{k+1} have been determined. If $H : \Omega \subset R^n \to \mathbb{R}^n$ has components functions $h_1, \ldots, h_n$, then the basic step of the the nonlinear GS, in analogy to the linear case, is to solve the l-th equation

$$h_l(x_1^{k+1}, \ldots, x_{l-1}^{k+1}, x_l, x_{l+1}^k, \ldots, x_n^k) = 0,$$

for x_l, and to set $x_l^{k+1} = x_l$. Thus, in order to obtain x^{k+1} from x^k, we have to solve successively the n one-dimensional nonlinear equations.

From a mathematical point of view, the established alternating least square (ALS) method [2, 3] and the density matrix renormalization group (DMRG) algorithm [9, 10, 16] are nonlinear block Gauss-Seidel methods, where the DMRG algorithm is also called modified alternating least square method (MALS). In the DMRG method we allow an enlargement of the parameter space and the partitioning (blocking) of the parameter space is not disjoint.

For the nonlinear block GS method, we want to describe the situation by an explicit example in order to motivate the abstract setting defined below. For this purpose consider a simple structured tensor network for $d = 3$, e.g. the tensor train format as defined in Example 2.7. The tensor train representation is described by the multilinear map

$$\begin{aligned} U_{TT} \quad &: \quad \mathcal{M}_0(\mathbb{N}_{\leq r}, V_1) \times \mathcal{M}_0(\mathbb{N}_{\leq r}^2, V_2) \times \mathcal{M}_0(\mathbb{N}_{\leq r}, V_3) \to \bigotimes_{\mu=1}^{3} V_\mu \\ \hat{v} \quad &= \quad (v_1, v_2, v_3) \mapsto U_{TT}(\hat{v}) = \sum_{j_1=1}^{r} \sum_{j_2=1}^{r} v_1(j_1) \otimes v_2(j_1, j_2) \otimes v_3(j_2), \end{aligned}$$

i.e. in our setting we have the parameter space $P = P_1 \times P_2 \times P_3$, where $P_1 = V_1^r$, $P_2 = V_2^{r^2}$, and $P_3 = V_3^r$. The ALS and the DMRG method are introduced by a partitioning of the parameter space P. The partitioning

$\{\tilde{X}_1, \tilde{X}_2, \tilde{X}_3\}$ for the ALS method is given by

$$P \quad = \quad \underbrace{P_1 \times \{0\} \times \{0\}}_{=:\tilde{X}_1} \quad + \quad \underbrace{\{0\} \times P_2 \times \{0\}}_{=:\tilde{X}_2} \quad + \quad \underbrace{\{0\} \times \{0\} \times P_3}_{=:\tilde{X}_3}$$

and the partitioning $\{X_1, X_2\}$ for the DMRG method is defined by

$$P \quad = \quad \underbrace{P_1 \times P_2 \times \{0\}}_{=:X_1} \quad + \quad \underbrace{\{0\} \times P_2 \times P_3}_{=:X_2}.$$

For general cases, a partition of the coordinates is defined as follows.

Definition 6.1 (Partition of Coordinates)**.** *Let $p \in \mathbb{N}$ and $P = \bigtimes_{\mu=1}^{D} P_\mu$ a parameter space of a tensor representation. We call the set $\{X_l \subset P : 1 \leq l \leq p\}$ a* partition of coordinates *of P if:*

(i) $P = \sum_{l=1}^{p} X_l$.

(ii) Every X_l is of the form $X_l = \bigtimes_{\mu=1}^{D} X_{l,\mu}$, where $X_{l,\mu}$ is either equal to P_μ or equal to the null space $\{0_{P_\mu}\}$ of P_μ.

We say $\{X_l \subset P : 1 \leq l \leq p\}$ is a disjoint partition of coordinates *if we have*

$$X_l \cap X_{l'} = \{0_P\} \text{ for all } 1 \leq l, l' \leq D.$$

For a convenient description of the nonlinear block GS method we introduce the function J_l. Where for a given partition of coordinates $\{X_l \subset P : 1 \leq l \leq p\}$, the function J_l can be viewed as the restriction of J to the the subset X_l, see Notation 6.2.

Notation 6.2. *Let $\{X_l \subset P : 1 \leq l \leq p\}$ be a partition of coordinates of P and $J : P \to \mathbb{R}$ the objective function from Problem 1.3. Further, let $X_l^c := P \setminus X_l$ be the complement of X_l in P. We define*

$$J_l : X_l \times X_l^c \to \mathbb{R},\ (x_l, x_l^c) \mapsto J_l(x_l, x_l^c) := J(x_l + x_l^c).$$

Definition 6.3 (Nonlinear block Gauss-Seidel method)**.** *Let $\{X_l \subset P : 1 \leq l \leq p\}$ be a partition of coordinates of P and $J : P \to \mathbb{R}$ the objective function. The* nonlinear block Gauss-Seidel method *is described by Algorithm 1.*

Similar to the linear case one can extend the nonlinear Gauss-Seidel method to the nonlinear successive over-relaxation method. The convergence analysis of the nonlinear GS method is already discussed in the literature, e.g. in [13, Ortega and Rheinboldt]. Generally speaking, the convergence of the nonlinear GS method is locally assigned by the convergence of the linear GS method applied to the Hessian $J''(\mathbf{x}^*)$ of J at a point $\mathbf{x}^* \in P$ with $J'(\mathbf{x}^*) = 0$. Let us consider a block decomposition of the Hessian $J''(x)$

$$J''_{block}(\mathbf{x}) = D_{block}(\mathbf{x}) - L_{block}(\mathbf{x}) - L^t{}_{block}(\mathbf{x})$$

into its block diagonal, strictly block lower-, and strictly block upper-triangular parts, where the blocking is introduced by a disjoint partitioning of coordinates, and suppose that $D_{block}(\mathbf{x}^*)$ is nonsingular. Furthermore, let $H(\mathbf{x})$ be defined by

$$H(\mathbf{x}) := [D_{block}(\mathbf{x}) - L_{block}(\mathbf{x})]^{-1} L^t_{block}(\mathbf{x}), \tag{35}$$

Algorithm 1 Nonlinear block GS method

Choose initial $\mathbf{x}^1 \in P$, and define $k := 1$.
while Stop Condition **do**
  for $1 \leq l \leq p$ **do**
    Compute $\tilde{\mathbf{x}}_l \in X_l$ such that

$$\partial_1 J_l \left(\tilde{\mathbf{x}}_l, \mathbf{x}^{\mathbf{c}}{}_l^{k(l)} \right) = 0_{X_l}, \tag{34}$$

    where $\mathbf{x}^{\mathbf{c}}{}_l^{k(l)} = \left(\mathbf{x}^{k(l)} - \mathbf{x}_l^{k(l)} \right) \in X_l^c$ and $\mathbf{x}^{k(l)} \in P$ is the current iteration point.
    $\mathbf{x}_l^{k(l+1)} := \tilde{\mathbf{x}}_l$.
  end for
  $k \mapsto k + p$.
end while

where $H(\mathbf{x})$ is simply the GS iteration matrix for the linear system $J''(\mathbf{x})\tilde{x} = b$. We can establish the following Theorem 6.4, whose proof follows directly from the arguments used in [13, Theorem 10.3.5, p. 326]. Unfortunately, the proof of Theorem 10.3.5 does not match for nonlinear block GS methods with a non-disjoint partition of the coordinates, since the function $\hat{G}$ defined in [13, Eq. (15), p. 326] does not fulfill $\hat{G}(\mathbf{x}^{k+1}, \mathbf{x}^k) = 0$ for all $k \in \mathbb{N}_0$. Therefore, we cannot apply Theorem 10.3.5 for methods with overlapping partition of the coordinates like the DMRG method.

Therorem 6.4. *Let $\{X_l \subset P : 1 \leq l \leq p\}$ be a partition of coordinates of P, $J \in C^2(P, \mathbb{R})$ and $\mathbf{x}^* \in P$ a parameter for which $J'(\mathbf{x}^*) = 0$ and $\rho\left(H(\mathbf{x}^*)\right) < 1$, where $H(\mathbf{x}^*)$ is defined in Eq. (35) and $D_{block}(\mathbf{x}^*)$ is nonsingular. Then there exists an environment $B(\mathbf{x}^*)$ of $\mathbf{x}^*$ such that, for any initial guess $\mathbf{x^1} \in B(\mathbf{x}^*)$, there is a unique sequence $(\mathbf{x^k})_{k\in\mathbb{N}} \subset B(\mathbf{x}^*)$ which satisfies the description of the nonlinear block GS method from Algorithm 1. Furthermore, $\lim_{k\to\infty} \mathbf{x^k} = \mathbf{x}^*$ is R-linear with R-convergence factor $\rho\left(H(\mathbf{x}^*)\right)$.*

The statement of Theorem 6.4 provides useful a priori information, even when it is not possible to ascertain in advance that $\rho\left(H(\mathbf{x}^*)\right) < 1$. For quadratic functionals F and the canonical tensor format U_{CP}, the situation $J = F \circ U_{CP}$ is considered in [20].

7 Efficient numerical treatment

In this section, we want to explicitly describe the nonlinear block Gauss-Seidel method for two different partitions of the coordinates. We want to express that the nonlinear block GS method can be performed efficiently on certain formats.

We are going to consider a tensor chain $a \in S := \bigotimes_{\mu=1}^d \mathbb{R}^{n_\mu}$ with $(n_1, \ldots, n_d) \in \mathbb{N}^d$ and representation rank $(R_1, \ldots, R_d) \in \mathbb{N}^d$ similarly to the tensor chain (TC) example in Section 2. We want to minimize

$$\|a - u\|$$

which is equivalent to minimizing

$$F_a(x) := \frac{1}{\|a\|^2} \left(-\langle a, x \rangle + \frac{1}{2} \langle x, x \rangle \right),$$

where the quotient $\|a\|^2$ has been added for numerical reasons. We consider U_{TC} from Eq. (14) with representation rank $(r_1, \ldots, r_d) \in \mathbb{N}^d$. So J of Eq. (3) is defined as $F_a \circ U_{TC}$, which we want to minimize in our experiments. We define

$$\begin{aligned} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) &:= \langle a_\mu(i_\mu, i_{\mu+1}), u_\mu(j_\mu, j_{\mu+1})\rangle, \qquad 1 \le \mu \le d-1 \\ A_d(i_d, i_1, j_d, j_1) &:= \langle a_d(i_1, i_d), u_d(i_1, j_d)\rangle \\ B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) &:= \langle u_\mu(j_\mu, j_{\mu+1}), u_\mu(j'_\mu, j'_{\mu+1})\rangle, \qquad 1 \le \mu \le d-1 \\ B_d(j_d, j_1, j'_d, j'_1) &:= \langle u_d(j_1, j_d), u_d(j'_1, j'_d)\rangle \end{aligned}$$

such that

$$\begin{aligned} J(u_1, \ldots, u_d) = \frac{1}{\|a\|^2} \Bigg(&- \sum_{\underline{i} \in \mathbf{I}} \sum_{\underline{j} \in \mathbf{J}} \left(\prod_{\mu=1}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_d, i_1, j_d, j_1) \\ &+ \frac{1}{2} \sum_{\underline{j} \in \mathbf{J}} \sum_{\underline{j}' \in \mathbf{J}} \left(\prod_{\mu=1}^{d-1} B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) \right) B_d(j_d, j_1, j'_d, j'_1) \Bigg) \end{aligned} \tag{36}$$

whereas $\mathbf{J} := \{(l_1, \ldots, l_d) : l_\mu = 1, \ldots, r_\mu, 1 \le \mu \le d\}, \mathbf{I} := \{(l_1, \ldots, l_d) : l_\mu = 1, \ldots, R_\mu, 1 \le \mu \le d\}$ and j_k denotes the k-th component of multi-index $\underline{j}$.

7.1 Alternating least squares for the tensor chain format

As stated earlier, the ALS method is the nonlinear block Gauss-Seidel method with disjoint partition of the coordinates that is defined in Section 6.

For convenience, we will assume $\ell \neq d$ in our formulae and lemmata. In the following part, we want to introduce some abbreviations, that will become handy later on. We define

$$\begin{aligned} A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) &:= \sum_{i_1,\ldots,i_{\ell-1}=1}^{R_1,\ldots,R_{\ell-1}} \sum_{i_{\ell+2},\ldots,i_d=1}^{R_{\ell+2},\ldots,R_d} \sum_{j_1,\ldots,j_{\ell-1}=1}^{r_1,\ldots,r_{\ell-1}} \sum_{j_{\ell+2},\ldots,j_d=1}^{r_{\ell+2},\ldots,r_d} \\ &\quad \left(\prod_{\mu=1,\mu\neq\ell}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_d, i_1, j_d, j_1) \end{aligned}$$

and

$$\begin{aligned} B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) &:= \sum_{j_1=1}^{r_1} \cdots \sum_{j_{\ell-1}=1}^{r_{\ell-1}} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \cdots \sum_{j_d=1}^{r_d} \sum_{j'_1=1}^{r_1} \cdots \sum_{j'_{\ell-1}=1}^{r_{\ell-1}} \sum_{j'_{\ell+2}=1}^{r_{\ell+2}} \cdots \sum_{j'_d=1}^{r_d} \\ &\quad \left(\prod_{\mu=1,\mu\neq\ell}^{d-1} B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) \right) B_d(j_d, j_1, j'_d, j'_1) \end{aligned}$$

which leads to a structure of Eq. (36) that pays respect to the partitioning:

$$\begin{aligned} J_\ell(u_\ell, u_\ell^c) = \frac{1}{\|a\|^2} \Bigg(&- \sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+1}=1}^{R_{\ell+1}} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+1}=1}^{r_{\ell+1}} A_\ell(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) \\ &+ \frac{1}{2} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+1}=1}^{r_{\ell+1}} \sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+1}=1}^{r_{\ell+1}} B_\ell(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) \Bigg) . \end{aligned}$$

Setting this derivative $\frac{\partial}{\partial u_\ell} J_\ell(u_\ell, u_\ell^c)$ equal to zero as in Eq. (34) in Algorithm 1, one has to solve the equation

$$\left(\mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right) \mathbf{a}_\ell = \left(\mathbf{B}_{[\ell]} \otimes Id_{n_\ell}\right) \mathbf{u}_\ell$$

where

$$\begin{aligned}
\mathbf{a}_\ell &:= \begin{pmatrix} a_\ell(1,1) \\ \vdots \\ a_\ell(R_\ell, R_{\ell+1}) \end{pmatrix} \\
\mathbf{A}_{[\ell]} &:= \left(A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1})\right)_{(j_\ell, j_{\ell+1}),(i_\ell, i_{\ell+1})} \\
\mathbf{u}_\ell &:= \begin{pmatrix} u_\ell(1,1) \\ \vdots \\ u_\ell(r_\ell, r_{\ell+1}) \end{pmatrix} \\
\mathbf{B}_{[\ell]} &:= \left(B_{[\ell]}(j'_\ell, j'_{\ell+1}, j_\ell, j_{\ell+1})\right)_{(j_\ell, j_{\ell+1}),(j'_\ell, j'_{\ell+1})}
\end{aligned}$$

and consequently, we have to solve

$$\mathbf{u}_\ell \stackrel{!}{=} \left(\mathbf{B}_{[\ell]}^{-1} \otimes Id_{n_\ell}\right) \left(\mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right) \mathbf{a}_\ell = \left(\mathbf{B}_{[\ell]}^{-1} \mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right) \mathbf{a}_\ell.$$

Remark 7.1. *The existence of $\mathbf{B}_{[\ell]}^{-1}$ is not guaranteed in all cases. If $\mathbf{B}_{[\ell]}$ is not regular, its matrix-rank is smaller than $r_\ell \cdot r_{\ell+1}$ and since $\mathbf{B}_{[\ell]}$ is a Gramian matrix we can reduce the rank of $B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1})$.*

To make compact statements about the complexity of the algorithm, we want to define $r := \max_{1 \le \mu \le d}\{r_\mu\}$, $R := \max_{1 \le \mu \le d}\{R_\mu\}$ and $n := \max_{1 \le \mu \le d}\{n_\mu\}$.

The question may arise, how to efficiently compute $\mathbf{A}_{[\ell]}$ and $\mathbf{B}_{[\ell]}$. For one single entry of $\mathbf{A}_{[\ell]}$, the naive approach (compute each term separately) is in $\mathcal{O}(r^{d-2} R^{d-2}(d-2))$ such that the complete cost would be in $\mathcal{O}(r^d R^d (d-2))$ which we want to avoid. A better approach it to treat each matrix entry as an inner product of two tensors in the MPS/TT format which is $\mathcal{O}((d-2) r^2 R^2)$. This improves the complete complexity to be $\mathcal{O}((d-2) r^4 R^4)$ but this still allows improvements since we have considered each entry as a separate tensor train inner product. If we take into account the connection between each entry, we can improve the complexity significantly. First, we introduce the definitions

$$A_\mu := \begin{cases} \left(A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1})\right)_{(i_\mu, j_\mu),(i_{\mu+1}, j_{\mu+1})} \in \mathbb{R}^{R_\mu r_\mu \times R_{\mu+1} r_{\mu+1}} & 1 \le \mu \le d-1 \\ \left(A_\mu(i_d, i_1, j_d, j_1)\right)_{(i_d, j_d),(i_1, j_1)} \in \mathbb{R}^{R_d r_d \times R_1 r_1} & \mu = d \end{cases}$$

and

$$B_\mu := \begin{cases} \left(B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1})\right)_{(j_\mu, j'_\mu),(j_{\mu+1}, j'_{\mu+1})} \in \mathbb{R}^{r_\mu^2 \times r_{\mu+1}^2} & 1 \le \mu \le d-1 \\ \left(B_\mu(j_d, j_1, j'_d, j'_1)\right)_{(j_d, j'_d),(j_1, j'_1)} \in \mathbb{R}^{r_d^2 \times r_1^2} & \mu = d \end{cases}$$

such that we can formulate the following lemma.

Lemma 7.2. *For $1 \le \ell \le d-1$ and A_μ and B_μ for $1 \le \mu \le d$ as defined above,*

$$\left(A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1})\right)_{(i_{\ell+1}, j_{\ell+1}),(i_\ell, j_\ell)} = \prod_{\mu=\ell+1}^{d} A_\mu \prod_{\mu=1}^{\ell-1} A_\mu$$

and

$$\Big(B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1})\Big)_{(j_{\ell+1},j'_{\ell+1}),(j_\ell,j'_\ell)} = \prod_{\mu=\ell+1}^{d} B_\mu \prod_{\mu=1}^{\ell-1} B_\mu,$$

hold true, so $\mathbf{A}_{[\ell]}$ and $\mathbf{B}_{[\ell]}$ can be interpreted as a product of matrices.

Proof. Without loss of generality, ℓ will be set equal to 1 and we will only prove the equation for $A_{[\ell]}$. As the first step, let us abbreviate

$$x((i_3, j_3), (i_1, j_1)) := \sum_{i_4,\dots,i_d=1}^{R_4,\dots,R_d} \sum_{j_4,\dots,j_d=1}^{r_4,\dots,r_d} \left(\prod_{\mu=3}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_1, i_d, j_1, j_d)$$

which results in

$$A_{[1]}(i_1, i_2, j_1, j_2) = \sum_{(i_3,j_3)=(1,1)}^{(R_3,r_3)} A_2(i_2, i_3, j_2, j_3) x((i_3, j_3), (i_1, j_1))$$

such that we see

$$\Big(A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1})\Big)_{(i_{\ell+1},j_{\ell+1}),(i_\ell,j_\ell)} = \Big(A_2(i_2, i_3, j_2, j_3)\Big)_{(i_2,j_2),(i_3,j_3)} \Big(x((i_3, j_3), (i_1, j_1))\Big)_{(i_3,j_3),(i_1,j_1)}.$$

Applying this procedure successively to $x((i_3, j_3), (i_1, j_1))$ finishes the proof, since analogous arguments hold for $B_{[\ell]}$. ■

Corollary 7.3. *The computational cost of $\mathbf{A}_{[\ell]}$ is at most*

$$\mathcal{O}(dr^3R^3).$$

Analogously, $\mathbf{B}_{[\ell]} \in \mathcal{O}(dr^6)$.

Note that $\mathbf{B}_{[\ell]}$ and $\mathbf{A}_{[\ell]}$ are only after reshaping representable as a product of matrices (compare the definition of $\mathbf{B}_{[\ell]}$ and $\mathbf{A}_{[\ell]}$ with Lemma 7.2).

We want to give the concrete algorithm for ALS in the TC format, which is a specialized version of Algorithm 1. First, we have to give four short definitions

$$\begin{aligned}
A_\ell^{(k)} &:= \Big(\langle a_\ell(i_\ell, i_{\ell+1}), u_\ell^{(k)}(j_\ell, j_{\ell+1})\rangle\Big)_{(i_{\ell+1},j_{\ell+1}),(i_\ell,j_\ell)} \\
A_{>\ell}^{(k)} &:= \prod_{\mu=\ell+1}^{d} A_\mu^{(k)}, \qquad A_{<\ell}^{(k)} := \prod_{\mu=1}^{\ell-1} A_\mu^{(k)} \\
B_\ell^{(k)} &:= \Big(\langle u_\ell^{(k)}(j_\ell, j_{\ell+1}), u_\ell^{(k)}(j'_\ell, j'_{\ell+1})\rangle\Big)_{(j_{\ell+1},j'_{\ell+1}),(j_\ell,j'_\ell)} \\
B_{>\ell}^{(k)} &:= \prod_{\mu=\ell+1}^{d} B_\mu^{(k)}, \qquad B_{<\ell}^{(k)} := \prod_{\mu=1}^{\ell-1} B_\mu^{(k)}
\end{aligned}$$

for $k \in \mathbb{N}$ and $1 \leq \ell \leq d-1$, where $u_\ell^{(k)}$ is that u_ℓ which has been computed in cycle k. Additionally, we set $A_{>d}^{(k)} = B_{>d}^{(k)} = A_0^{(k)} = B_0^{(k)} = Id$.

Lemma 7.4. *The computational cost of* $\left(\mathbf{B}_{[\ell]}^{-1}\mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right)\mathbf{a}_\ell$ *is at most*

$$\mathcal{O}(dr^6) + \mathcal{O}(dr^3R^3) + \mathcal{O}(n(r^2R^2 + r^4))$$

if the matrices A_μ *and* B_μ *are given for* $1 \le \mu \le d$ *and if we consider reshaping of a matrix as a free operation.*

Proof. From Lemma 7.2, we can conclude that the computational cost for $\mathbf{B}_{[\ell]}$ is equal to the computational cost of $d-2$ matrix-matrix multiplications of $r^2 \times r^2$ matrices such that

$$cost(\mathbf{B}_{[\ell]}) \in \mathcal{O}((d-2)r^6)$$

and an analogous argument holds for $\mathbf{A}_{[\ell]}$, such that

$$cost(\mathbf{A}_{[\ell]}) \in \mathcal{O}((d-2)r^3R^3)$$

since $\mathbf{A}_{[\ell]}$ can be calculated as a product of $rR \times rR$ matrices. Computing $\mathbf{B}_{[\ell]}^{-1}$ from $\mathbf{B}_{[\ell]}$ has a complexity of $\mathcal{O}(r^6)$. The computation of $\left(\mathbf{B}_{[\ell]}^{-1}\mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right)\mathbf{a}_\ell$ can be done by one matrix-matrix multiplication without having to perform $\otimes Id_{n_\ell}$ by considering $a_\ell(i_\ell, i_{\ell+1})$ as columns of $\tilde{\mathbf{a}}_\ell \in \mathbb{R}^{n_\ell \times R_\ell \cdot R_{\ell+1}}$ such that

$$\left(\mathbf{B}_{[\ell]}^{-1}\mathbf{A}_{[\ell]} \otimes Id_{n_\ell}\right)\mathbf{a}_\ell \cong \tilde{\mathbf{a}}_\ell \left(\mathbf{B}_{[\ell]}^{-1}\mathbf{A}_{[\ell]}\right)^T = \tilde{\mathbf{a}}_\ell \mathbf{A}_{[\ell]}^T \mathbf{B}_{[\ell]}^{-1^T}$$

which finishes the proof by $\tilde{\mathbf{a}}_\ell \mathbf{A}_{[\ell]}^T \mathbf{B}_{[\ell]}^{-1^T}$ being in

$$\mathcal{O}(n(r^2R^2 + r^4))$$

if we compute $\tilde{\mathbf{a}}_\ell \mathbf{A}_{[\ell]}^T$ first. ■

In cycle k, in the ℓ-th step of Algorithm 1, we have to compute $B_{>\ell}^{(k)} B_{<\ell}^{(k+1)}$ and $A_{>\ell}^{(k)} A_{<\ell}^{(k+1)}$. Therefore, it is more efficient to compute and store $B_{>\ell}^{(k)}$ and $A_{>\ell}^{(k)}$ in a prephase. Additionally, we will store $B_{<\ell}^{(k+1)}$ and $A_{<\ell}^{(k+1)}$ in each ℓ-step since $B_{<\ell+1}^{(k+1)} = B_{<\ell}^{(k+1)} B_\ell^{(k+1)}$ and $A_{<\ell+1}^{(k+1)} = A_{<\ell}^{(k+1)} A_\ell^{(k+1)}$ for $1 \le \ell \le d-1$.

Lemma 7.5. *One complete ALS cycle with prephase, as described in Algorithm 2, is at most*

$$\mathcal{O}(dr^6) + \mathcal{O}(dr^3R^3) + \mathcal{O}(dn(r^2R^2 + r^4))$$

in terms of complexity.

Proof. Follows from the described prephase and the proof of Lemma 7.4. ■

Remark 7.6. *The prephase described above needs additional storage of* $dr^4 + dr^2R^2$.

Remark 7.7. *Computing the initially needed* B_μ *and* A_μ *for* $2 \le \mu \le d$ *in Lemma 7.4 and 7.5 is in*

$$\mathcal{O}\left(dn(r^2R^2 + r^4)\right)$$

in terms of the complexity.

Algorithm 2 Alternating Least Squares (ALS) Method for the TC format

Choose initial $u^{(1)} = (u_1^{(1)}, \dots, u_d^{(1)}) \in \times_{\mu=1}^{d} P_\mu$ and parameter $\varepsilon \in \mathbb{R}_{>0}$. Define $\mathbf{g} := J(u^{(1)}), k := 1$.
while $\Delta\mathbf{g} > \varepsilon$ **do**
  $\tilde{B} := Id, \tilde{A} := Id$
  for $d-1 \geq \ell \geq 1$ **do**
    store $B_{>\ell}^{(k)} = B_{\ell+1}^{(k)} B_{>\ell+1}^{(k)}$ and $A_{>\ell}^{(k)} = A_{\ell+1}^{(k)} A_{>\ell+1}^{(k)}$
  end for
  for $1 \leq \ell \leq d$ **do**
    $\tilde{B} \mapsto \tilde{B} B_{\ell-1}^{(k+1)} \; \{\Rightarrow \tilde{B} = B_{<\ell}^{(k+1)}\}$
    $\tilde{A} \mapsto \tilde{A} A_{\ell-1}^{(k+1)} \; \{\Rightarrow \tilde{A} = A_{<\ell}^{(k+1)}\}$
    $\mathbf{u}_\ell^{(k+1)} := \left(\left(reshape \left(B_{>\ell}^{(k)} \tilde{B} \right) \right)^{-1} reshape \left(A_{>\ell}^{(k)} \tilde{A} \right) \otimes Id_{n_\ell} \right) \mathbf{a}_\ell$
  end for
  $\mathbf{g} \leftarrow J(u^{(k+1)})$
  $k \mapsto k+1$
end while

7.2 DMRG for the tensor chain format

ALS does not adjust the ranks of the edges, so now, we want to choose a slightly different approach: Instead of fixing all nodes but one, we are fixing all nodes but two neighbored ones. So we are using the following partition of coordinates as stated in Section 6.

In contrary to ALS, we do not have a disjoint partitioning which gives us the opportunity to adjust the rank between nodes ℓ and $\ell+1$ for $1 \leq \ell \leq d-1$ and between nodes d and 1 since we do not have to fix $r_{\ell+1}$ and r_1, respectively.

From now on, $1 \leq \ell \leq d-2$ in order to keep the readability of the upcoming notations.

Similar to the previous section, we want to define some useful abbreviations

$$\begin{aligned}
A_{[\ell)}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) &:= \sum_{i_1,\dots,i_{\ell-1}=1}^{R_1,\dots,R_{\ell-1}} \sum_{i_{\ell+3},\dots i_d=1}^{R_{\ell+3},\dots,R_d} \sum_{j_1,\dots,j_{\ell-1}=1}^{r_1,\dots,r_{\ell-1}} \sum_{j_{\ell+3},\dots,j_d=1}^{r_{\ell+3},\dots,r_d} \\
&\quad \left(\prod_{\mu=1,\mu\notin\{\ell,\ell+1\}}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_1, i_d, j_1, j_d), \\
B_{[\ell)}(j_\ell, j_{\ell+2}, j'_\ell, j'_{\ell+2}) &:= \sum_{j_1,\dots,j_{\ell-1}=1}^{r_1,\dots,r_{\ell-1}} \sum_{j_{\ell+3},\dots,j_d=1}^{r_{\ell+3},\dots,r_d} \sum_{j'_1,\dots,j'_{\ell-1}=1}^{r_1,\dots,r_{\ell-1}} \sum_{j'_{\ell+3},\dots,j'_d=1}^{r_{\ell+3},\dots,r_d} \\
&\quad \left(\prod_{\mu=1,\mu\notin\{\ell,\ell+1\}}^{d-1} B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) \right) B_d(j_1, j_d, j'_1, j'_d), \\
a_{\ell,\ell+1}(i_\ell, i_{\ell+2}) &:= \sum_{i_{\ell+1}=1}^{R_{\ell+1}} a_\ell(i_\ell, i_{\ell+1}) \otimes a_{\ell+1}(i_{\ell+1}, i_{\ell+2}) \in \mathbb{R}^{n_\ell \times n_{\ell+1}} \\
&\text{and} \\
u_{\ell,\ell+1}(j_\ell, j_{\ell+2}) &:= \sum_{j_{\ell+1}=1}^{r_{\ell+1}} u_\ell(j_\ell, j_{\ell+1}) \otimes u_{\ell+1}(j_{\ell+1}, j_{\ell+2}) \in \mathbb{R}^{n_\ell \times n_{\ell+1}} \qquad (37)
\end{aligned}$$

such that Eq. (36) with respect to the above written partitioning is

$$\begin{aligned}
J_\ell(u_{\ell,\ell+1}, u^c_{\ell,\ell+1}) = \frac{1}{\|a\|^2} \Bigg(&- \sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+2}=1}^{R_{\ell+2}} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \langle a_{\ell,\ell+1}(i_\ell, i_{\ell+2}), u_{\ell,\ell+1}(j_\ell, j_{\ell+2}) \rangle A_{[\ell)}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) \\
&+ \frac{1}{2} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+2}=1}^{r_{\ell+2}} \langle u_{\ell,\ell+1}(j_\ell, j_{\ell+2}), u_{\ell,\ell+1}(j'_\ell, j'_{\ell+2}) \rangle B_{[\ell)}(j_\ell, j_{\ell+2}, j'_\ell, j'_{\ell+2}) \Bigg).
\end{aligned}$$

Setting this derivative $\frac{\partial}{\partial u_{\ell,\ell+1}} J_\ell(u_{\ell,\ell+1}, u^c_{\ell,\ell+1})$ equal to zero results in

$$\left(\mathbf{A}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\right) \mathbf{a}_{\ell,\ell+1} = \left(\mathbf{B}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\right) \mathbf{u}_{\ell,\ell+1}$$

where

$$\begin{aligned}
\mathbf{a}_{\ell,\ell+1} &:= \begin{pmatrix} a_{\ell,\ell+1}(1,1) \\ \vdots \\ a_{\ell,\ell+1}(R_\ell, R_{\ell+2}) \end{pmatrix} \\
\mathbf{A}_{[\ell)} &:= \left(A_{[\ell)}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) \right)_{(j_\ell, j_{\ell+2}),(i_\ell, i_{\ell+2})} \\
\mathbf{u}_{\ell,\ell+1} &:= \begin{pmatrix} u_{\ell,\ell+1}(1,1) \\ \vdots \\ u_{\ell,\ell+1}(r_\ell, r_{\ell+2}) \end{pmatrix} \\
\mathbf{B}_{[\ell)} &:= \left(B_{[\ell)}(j'_\ell, j'_{\ell+2}, j_\ell, j_{\ell+2}) \right)_{(j_\ell, j_{\ell+2}),(j'_\ell, j'_{\ell+2})}
\end{aligned}$$

such that we have to solve

$$\mathbf{u}_{\ell,\ell+1} \stackrel{!}{=} \left(\mathbf{B}^{-1}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\right) \left(\mathbf{A}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\right) \mathbf{a}_{\ell,\ell+1} = \left(\mathbf{B}^{-1}_{[\ell)} \mathbf{A}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\right) \mathbf{a}_{\ell,\ell+1}$$

in order to improve the approximation. This formula will give us all $u_{\ell,\ell+1}$ but what we need are all u_ℓ and $u_{\ell+1}$. So we have to separate $u_{\ell,\ell+1}$ and the obvious way to do this is by using the singular value decomposition (SVD). If we reorder $\mathbf{u}_{\ell,\ell+1}$ such that i_ℓ with the component dimension of u_ℓ are the row index and $i_{\ell+1}$ with the component dimension of $u_{\ell+1}$ are the column index:

$$\Big(u_{\ell,\ell+1}(i_\ell, i_{\ell+2})_{m_\ell,m_{\ell+1}}\Big)_{(m_\ell,i_\ell),(m_{\ell+1},i_{\ell+2})} \overset{SVD}{=} \sum_{j_{\ell+1}=1}^{\tilde{r}_{\ell+1}} \begin{pmatrix} u_\ell(1,i)_1 \\ u_\ell(1,i)_2 \\ \vdots \\ u_\ell(r_\ell,i)_{n_\ell} \end{pmatrix} \otimes \begin{pmatrix} u_{\ell+1}(i,1)_1 \\ u_{\ell+1}(i,1)_2 \\ \vdots \\ u_{\ell+1}(i,r_{\ell+2})_{n_\ell+1} \end{pmatrix}$$

where we obtain the terms separated. Note that $\tilde{r}_{\ell+1}$ is the new rank for the edge between the optimized nodes.

Just as before it is now necessary to compute $\mathbf{A}_{[\ell)}$ and $\mathbf{B}_{[\ell)}$ in an efficient way. That can be done similarly to Lemma 7.2.

Lemma 7.8. *For $1 \leq \ell \leq d$ and A_μ and B_μ for $1 \leq \mu \leq d$ as defined in Section 7.1,*

$$\Big(A_{[\ell)}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2})\Big)_{(i_{\ell+2},j_{\ell+2}),(i_\ell,j_\ell)} = \prod_{\mu=\ell+2}^{d} A_\mu \prod_{\mu=1}^{\ell-1} A_\mu$$

and

$$\Big(B_{[\ell)}(j_\ell, j_{\ell+2}, j'_\ell, j'_{\ell+2})\Big)_{(j_{\ell+2},j'_{\ell+2}),(j_\ell,j'_\ell)} = \prod_{\mu=\ell+2}^{d} B_\mu \prod_{\mu=1}^{\ell-1} B_\mu,$$

hold true, so $A_{[\ell)}$ and $B_{[\ell)}$ can be interpreted as a product of matrices.

Proof. Analogous to Lemma 7.2. ■

Lemma 7.9. *Computing $\Big(\mathbf{B}_{[\ell)}^{-1}\mathbf{A}_{[\ell)} \otimes Id_{n_\ell \times n_{\ell+1}}\Big)\mathbf{a}_{\ell,\ell+1}$ is in*

$$\mathcal{O}(dr^6) + \mathcal{O}(dr^3R^3) + \mathcal{O}(n^2(r^2R^2 + r^4)).$$

Proof. Analogous to the proof of Lemma 7.4. ■

Similar to section 7.1, we add a prephase, which computes and stores $A_{>\ell}^{(k)}$ and $B_{>\ell}^{(k)}$ for $2 \leq \ell \leq d-1$ before the k-th cycle. Then one complete DMRG cycle has a complexity linear in d.

Lemma 7.10. *One DMRG cycle with prephase is in*

$$\mathcal{O}(dr^6) + \mathcal{O}(dr^3R^3) + \mathcal{O}(dn^2(r^2R^2 + r^4)) + \mathcal{O}(dn^3R^3).$$

Proof. Follows directly from Lemma 7.9 and the complexity of the singular value decomposition being in $\mathcal{O}(n^3R^3)$. ■

Algorithm 3 DMRG Method for the TC format

Choose initial $u^{(1)} = (u_1^{(1)}, \ldots, u_d^{(1)}) \in \times_{\mu=1}^{d} P_\mu$ and parameter $\varepsilon \in \mathbb{R}_{>0}$. Define $\mathbf{g} := J(u^{(1)}), k := 1$.
while $\Delta\mathbf{g} > \varepsilon$ **do**
  $\mathbf{u}_{d,1}^{(k+1)} := \left(\left(reshape\left(\prod_{\mu=2}^{d-1} B_\mu\right)\right)^{-1} reshape\left(\prod_{\mu=2}^{d-1} A_\mu\right) \otimes Id_{n_d \times n_1}\right) \mathbf{a}_{d,1}$
  $[u_d^{(k+1)}, u_1^{(k+1)}] := SVD\left(reshape\left(\mathbf{u}_{d,1}^{(k+1)}\right)\right)$
  $\tilde{B} := Id, \tilde{A} := Id$
  for $d-1 \geq \ell \geq 2$ **do**
    store $B_{>\ell}^{(k)} = B_{\ell+1}^{(k)} B_{>\ell+1}^{(k)}$ and $A_{>\ell}^{(k)} = A_{\ell+1}^{(k)} A_{>\ell+1}^{(k)}$
  end for
  for $1 \leq \ell \leq d-1$ **do**
    $\tilde{B} \mapsto \tilde{B} B_{\ell-1}^{(k+1)} \; \{\Rightarrow \tilde{B} = B_{<\ell}^{(k+1)}\}$
    $\tilde{A} \mapsto \tilde{A} A_{\ell-1}^{(k+1)} \; \{\Rightarrow \tilde{A} = A_{<\ell}^{(k+1)}\}$
    $\mathbf{u}_{\ell,\ell+1}^{(k+1)} := \left(\left(reshape\left(B_{>\ell+1}^{(k)} \tilde{B}\right)\right)^{-1} reshape\left(A_{>\ell+1}^{(k)} \tilde{A}\right) \otimes Id_{n_\ell \times n_{\ell+1}}\right) \mathbf{a}_{\ell,\ell+1}$
    $[u_\ell^{(k+1)}, u_{\ell+1}^{(k)}] := SVD\left(reshape\left(\mathbf{u}_{\ell,\ell+1}^{(k+1)}\right)\right)$
  end for
  $\mathbf{g} \leftarrow J(u^{(k+1)})$
  $k \mapsto k+1$
end while

7.3 Numerical experiments

After describing the ALS and DMRG algorithms for the TC format, we will perform some very basic experiments with data that has been obtained from two–electron integrals.

Since the algorithmic description is about approximating a TC tensor with another TC tensor, we first have to convert the data, which is a full order 4 tensor, into a TC tensor without introducing a relevant error. A TT tensor is a TC tensor where one representation rank component is equal to one, so we will perform the conversion with the procedure which is described in [15] as *TT-SVD algorithm* and in [21] as *Vidal decomposition* in the MPS context where we choose an approximation accuracy of 10^{-12}. This procedure gives us a TT tensor that we will approximate with a TC tensor with the help of the described algorithms ALS and DMRG.

We are using different molecules (NH_3 and H_2O) as well as different (chemical) basis sets (STO-3G and 6-31G) that lead to different vector space dimensions. The initial $\mathbf{r} = (R_1, \ldots, R_4)$ is the representation rank that we obtain while converting the full tensor to the TT format and ϵ is the relative error of the initial TT tensor and the approximated TC tensor. We will write $(a)^4$ instead of (a, a, a, a) in order to save space.

Another important factor for the experiments is the number of complete ALS/DMRG iterations that are needed to compute the approximations (we will denote them by "#iter.") which immediately gives an impression about the consumed time (together with the complexity estimates).

The initial guess for the TC approximation is chosen with the following procedure:

$$\sum_{i_\mu=1}^{R_\mu} u_{\mu-1}(\cdot, i_\mu) \otimes u_\mu(i_\mu, \cdot) \overset{ACA}{\approx:} \sum_{i_\mu=1}^{r_\mu} u_{\mu-1}(\cdot, i_\mu) \otimes u_\mu(i_\mu, \cdot)$$

for $\mu = 2, 3, 4$ ($\mu = 1$ analogously), such that $u_{\mu-1}$ and u_μ are getting reassigned, which is known as adaptive

cross approximation (or ACA, see [1]). All computations are preformed with the library [4].
The subsequent tables display our results that we could obtain using Algorithms 2 and 3.

Table 1: Reduced representation ranks for AO integrals in H_2O using ALS

Basis set	$\dim(V_\mu)$	Initial $\mathbf{r}$	ϵ 10^{-2}		10^{-4}		10^{-6}	
			$\mathbf{r}$	#iter.	$\mathbf{r}$	#iter.	$\mathbf{r}$	#iter.
STO-3G	7	$(7, 49, 7, 1)$	$(12)^4$	34	$(14)^4$	61	$(16)^4$	31
6-31G	13	$(13, 169, 13, 1)$	$(27)^4$	71	$(40)^4$	82	$(44)^4$	42

Table 2: Reduced representation ranks for AO integrals in NH_3 using ALS

Basis set	$\dim(V_\mu)$	Initial $\mathbf{r}$	ϵ 10^{-2}		10^{-4}		10^{-6}	
			$\mathbf{r}$	#iter.	$\mathbf{r}$	#iter.	$\mathbf{r}$	#iter.
STO-3G	8	$(8, 64, 8, 1)$	$(14)^4$	97	$(16)^4$	195	$(18)^4$	90
6-31G	15	$(15, 225, 15, 1)$	$(33)^4$	29	$(50)^4$	72	$(55)^4$	39

Table 3: Reduced representation ranks for AO integrals in H_2O using DMRG

Basis set	$\dim(V_\mu)$	Initial $\mathbf{r}$	$\mathbf{r}$	ϵ	#iter.
STO-3G	7	$(7, 49, 7, 1)$	$(43, 7, 7, 7)$	$4.6 \cdot 10^{-6}$	1
6-31G	13	$(13, 169, 13, 1)$	$(162, 13, 13, 13)$	$2.3 \cdot 10^{-6}$	1

The conclusion of these experiments is, that we can use the ALS method to even out the representation rank. This can become an advantage when a TC tensor is treated with algorithms that run in parallel over multiple dimensions.

References

[1] M. Bebendorf. Approximation of boundary element matrices. *Numer. Math.*, 86(4):565–589, 2000.

[2] G. Beylkin and M. J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proc. Natl. Acad. Sci. USA*, 99(16):10246–10251, 2002.

[3] G. Beylkin and M. J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM J. Sci. Comput.*, 26(6):2133–2159, 2005.

[4] M. Espig, M. Schuster, A. Killaitis, N. Waldren, P. Wähnert, S. Handschuh, and H. Auer. TensorCalculus library, 2008–2012. `http://gitorious.org/tensorcalculus`.

[5] A. Falcó and W. Hackbusch. On minimal subspaces in tensor representations. 2010. MIS Preprint 70.

[6] L. Grasedyck. Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.*, 2010.

[7] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer, 2012.

[8] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.*, 5(15):706–722, 2009.

[9] S. Holtz, T. Rohwedder, and R. Schneider. The alternating linear scheme for tensor optimisation in the TT format. 2010. Preprint 71.

[10] T. Huckle, K. Waldherr, and T. Schulte-Herbrüggen. Computations in quantum tensor networks. 2010. Preprint.

[11] B. N. Khoromskij. $\mathcal{O}(d \log N)$-Quantics Approximation of $N-d$ Tensors in High-Dimensional Numerical Modeling. *Const. Approx.*, 34:1–24, 2010.

[12] J. M. Landsberg, Y. Qi, and K. Ye. On the geometry of tensor network states. *arXiv:1105.4449 [math.AG]*, 2011.

[13] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial Mathematics, 1970.

[14] I. V. Oseledets. Compact matrix form of the d-dimensional tensor decomposition. 2009. INM RAS preprint 01.

[15] I. V. Oseledets. Tensor-Train Decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.

[16] I. V. Oseledets and B. N. Khoromskij. DMRG and QTT approach to high-dimensional quantum molecular dynamics. 2010. MIS preprint 69.

[17] N. Schuch, I. Cirac, and D. Pérez-García. PEPS as ground states: degeneracy and topology. *arXiv:1001.3807v2 [quant-ph]*, 2010.

[18] S. Singh, R. N. C. Pfeifer, and G. Vidal. Tensor network decompositions in the presence of a global symmetry. *arXiv:0907.2994v1 [cond-mat.str-el]*, 2009.

[19] E. E. Tyrtyshnikov and I. V. Oseledets. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.*, 31:3744–3759, 2009.

[20] A. Uschmajew. Local convergence of the alternating least squares algorithm for canonical tensor approximation. 2011. Preprint 103.

[21] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91(14), 2003.