DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Optimization Problems in Contracted Tensor Networks

M. Espig, W. Hackbusch, S. Handschuh, R. Schneider

Preprint 108

Edited by

DFG-Schwerpunktprogramm 1324

Optimization Problems in Contracted Tensor Networks

M. Espig, W. Hackbusch, S. Handschuh, R. Schneider

Preprint 108

The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.

Optimization Problems in Contracted Tensor Networks

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

October 17, 2011

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

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

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

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 parametrised way, i.e. we are searching for*

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

where $\mathcal{U} \subset \mathcal{V}$ is the image of a multilinear map $U : P \to \mathcal{V}$. The map U is a tensor format from a parameter space P into the tensor product space, 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.

---

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

(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 $F(u) = \|Au - b\|_{\mathcal{V}'}^2$, 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:*

$$\text{find } \mathbf{u} \in \mathfrak{M}(J, M), \ J := F \circ U : P \to \mathcal{V} \to \mathbb{R} \text{ and } M \subseteq P. \tag{3}$$

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 $\mathbb{N}^\ell := \overset{\ell}{\underset{\nu=1}{\times}} \mathbb{N} \ (1 \leq \ell)$. The set of maps with finite support from $\mathbb{N}^\ell$ into A is defined by*

$$\mathcal{M}_0(\mathbb{N}^\ell, A) := \begin{cases} A, & \ell = 0, \\ \{u : \mathbb{N}^\ell \to A \mid \# \operatorname{supp}(u) \in \mathbb{N}\}, & \ell \geq 1. \end{cases} \tag{4}$$

The natural number ℓ is called the degree of $u \in \mathcal{M}_0(\mathbb{N}^\ell, A)$.

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

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 $\overset{\bullet}{\mid}$ stands for nodes of vector space meaning and the symbol $\blacksquare$*

$$
\begin{array}{ccccc}
v_1 & & w & & v_2 \\
\bullet & \rule{2cm}{0.4pt} & \blacksquare & \rule{2cm}{0.4pt} & \bullet \\
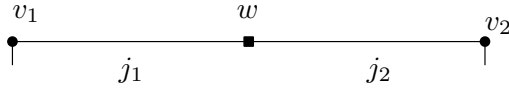& j_1 & & j_2 &
\end{array}
$$

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

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:

$$
U_G \quad : \quad \mathcal{M}_0(\mathbb{N}^2, \mathbb{R}) \times \mathcal{M}_0(\mathbb{N}, V_1) \times \mathcal{M}_0(\mathbb{N}, V_2) \to V_1 \otimes V_2
$$

$$
(w, v_1, v_2) \mapsto U_G(w, v_1, v_2) \quad := \quad \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \underbrace{w(j_1, j_2)}_{\in \mathbb{R}} \underbrace{v_1(j_1)}_{\in V_1} \otimes \underbrace{v_2(j_2)}_{\in V_2},
$$

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, Tensor Network Tree, Degree Map). *Let*

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

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

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

be a finite subset of nodes of interior parameters with $\#N_c = L$. *Further let* $N := N_s \cup N_c$ *and* $E \subset \{\{n_1, n_2\} : n_1, n_2 \in N, n_1 \ne n_2\} \subset \mathfrak{P}(N)$ *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* $e : \mathbb{N}_{\le m} \to E$, *where* $m := \#E$. *We call the map*

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

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

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

We will not distinguish between N and $\mathbb{N}_{\le d+L}$ such that we identify both sets with each other, i.e. there is a bijective map $\varphi : N_{\le d+L} \to N$ such that we can uniquely identify $\mu \in N_{\le d+L}$ 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 $1 \le \mu \le d$ 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* $\mathcal{V}$.

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

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

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

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

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

4

We say $u = U_{G,\underline{r}}(v_1, \ldots, v_d, w_1, \ldots, w_L) \in \text{Range } U_{G,\underline{r}} \subset \mathcal{V}$ *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) [18, 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.
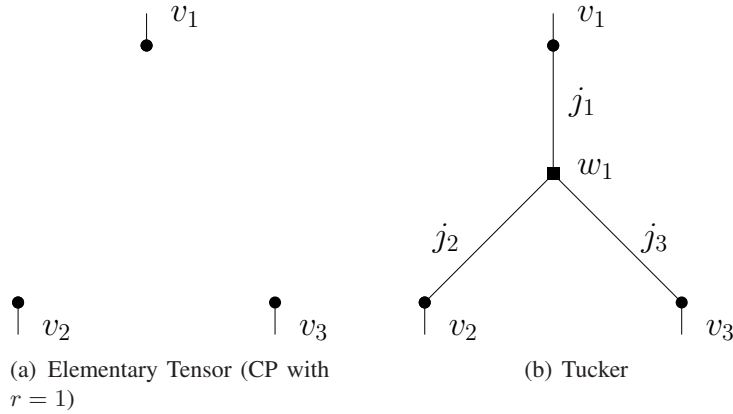

(a) Elementary Tensor (CP with $r = 1$)

(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*
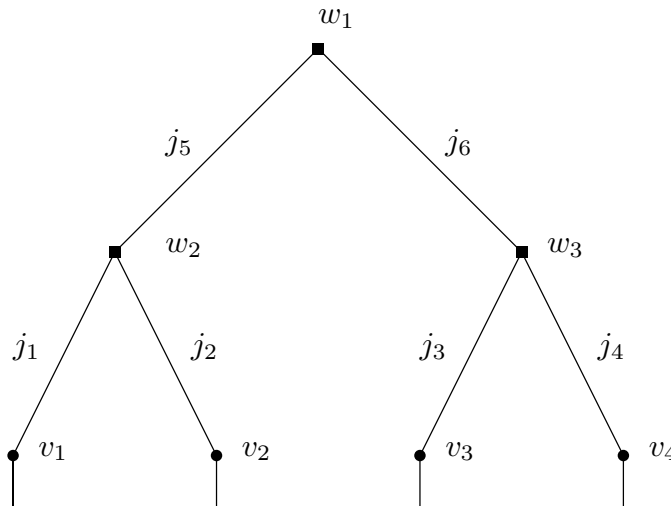

Figure 3: The tensor network graph of the hierarchical tensor format for $d = 4$.

5

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

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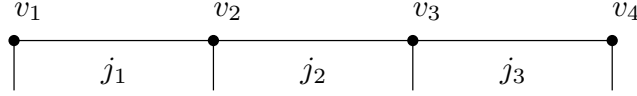

Figure 4: The tensor network graph of the tensor train format for $d = 4$.

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 $\mathcal{I}_{TT} : N \times \times_{l=1}^3 \mathbb{N} \to \bigcup_{l=1}^3 \mathbb{N}^l$ 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
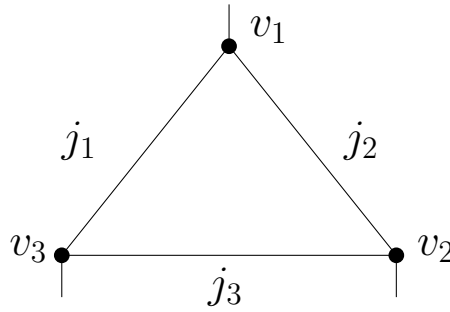

Figure 5: The tensor network graph of the tensor chain for $d = 3$.

6

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. [17], [16]. For $d = 6$, the tensor network graph of the PEPS tensor network is shown in Figure 6.
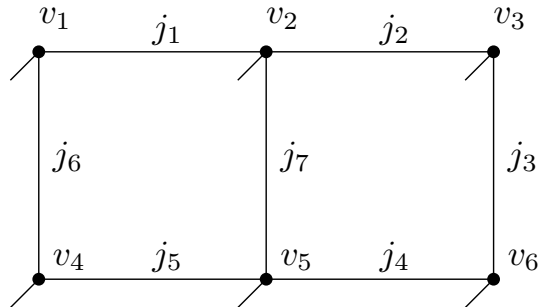

Figure 6: The tensor network graph of the PEPS for $d = 6$.

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].

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.*

The proof of Theorem 3.2 is deeply influenced by the ideas of Hackbusch published in [7, Chapter 6].

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. 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 $U_G(\hat{u}^k)$

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

with

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

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

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 $u_2(\cdot) = U_{G_2}(\hat{u}_2(\cdot))$.

8

Let $U_1' = \{u_1'(i) \in \mathcal{V}_1' : 1 \le i \le r'\}$ and $U_2' = \{u_2'(i) \in \mathcal{V}_2' : 1 \le i \le r'\}$ 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 $1 \le i \le r'$. After short calculation (using tensor contractions) we have that

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

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

where we define $\tilde{v}_{d_1}^k((\mathcal{I}_1(\mu,\underline{j}),\cdot)) := \sum_{j_e=1}^r u_2'(\cdot)(U_{G_2}(u_2^k(j_e)))v_{d_1}^k((\mathcal{I}_1(\mu,\underline{j}),j_e)) \in \mathcal{M}_0(\mathbb{N}_{\le r}, V_{d_1})$ and $\tilde{v}_d^k((\mathcal{I}_2(\mu,\underline{j}),\cdot)) := \sum_{j_e=1}^r u_1'(\cdot)(U_{G_1}(u_1^k(j_e)))v_d^k((\mathcal{I}_2(\mu,\underline{j}),j_e)) \in \mathcal{M}_0(\mathbb{N}_{\le r}, V_d)$. 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 $r' \le r$.

∎

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 $\dim(V_\mu) \le 3$ (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

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

(c) The sequence $(u^k)_{k\in\mathbb{N}} \subset \operatorname{Range}(U_{G,\underline{r}})$ is called stable in $\operatorname{Range}(U_{G,\underline{r}})$, *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 $(u^k)_{k\in\mathbb{N}} \subset \operatorname{Range}(U_{G,\underline{r}})$ is stable and convergent, then $\lim_{k\to\infty} u^k \in \operatorname{Range}(U_{G,\underline{r}})$.*

Proof. Let $(u^k)_{k\in\mathbb{N}} \subset \operatorname{Range}(U_{G,\underline{r}})$ 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 $P := \underset{\nu=1}{\overset{d+L}{\times}} P_\nu$ 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 $\underline{R} = (R_1, \ldots, R_m)^t \in \mathbb{N}^m$ and $\underline{r} = (r_1, \ldots, r_m)^t \in \mathbb{N}^m$ 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 $p_R = (v_{R,1}, \ldots, w_{R,L}) \in P_R$ 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). *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 $p_r^* = (v_{r,1}^*, \ldots, w_{r,L}^*) \in P_r$ 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 $\underline{j} \in \bigtimes_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$*

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

Proof. Assume there is a $\mu^* \in \mathbb{N}_{\leq d}$ and a $\underline{j} \in \bigtimes_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$ 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

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

and 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

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

11

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 $\underline{i} \in \bigtimes_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq R_{e(l)}}$ and $\underline{j} \in \bigtimes_{1 \leq l \leq g(\mu)} \mathbb{N}_{\leq r_{e(l)}}$. 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

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

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 (NGS) 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 NGS 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 NGS, 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, 15] 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 NGS 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

$$U_{TT} \quad : \quad \mathcal{M}_0(\mathbb{N}_{\leq r}, V_1) \times \mathcal{M}_0(\mathbb{N}_{\leq r}^2, V_2) \times \mathcal{M}_0(\mathbb{N}_{\leq r}, V_3) \to \bigotimes_{\mu=1}^3 V_\mu$$

$$\hat{v} \quad = \quad (v_1, v_2, v_3) \mapsto U_{TT}(\hat{v}) = \sum_{j_1=1}^r \sum_{j_2=1}^r v_1(j_1) \otimes v_2(j_1, j_2) \otimes v_3(j_2),$$

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 $P = \overset{D}{\underset{\mu=1}{\times}} P_\mu$ a parameter space of a tensor representation. We call the set $\{X_l \subset P : 1 \le l \le p\}$ a* partition of coordinates *of P if:*

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

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

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

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

For a convenient description of the nonlinear block SOR method we introduce the function J_l. Where for a given partition of coordinates $\{X_l \subset P : 1 \le l \le p\}$, 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 $\{X_l \subset P : 1 \le l \le p\}$ 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 $\{X_l \subset P : 1 \le l \le p\}$ be a partition of coordinates of P and $J : P \to \mathbb{R}$ the objective function. The* nonlinear block Gauss-Seidel (GS) method *is described by Algorithm 1.*

Similar to the lineare 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 portioning 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

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

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

        where $\mathbf{x}_l^{\mathbf{c}k(l)} = \left(\mathbf{x}^{k(l)} - \mathbf{x}_l^{k(l)}\right) \in X_l^c$ and $\mathbf{x}^{k(l)} \in P$ is the current iterant.
5:         $\mathbf{x}_l^{k(l+1)} := \tilde{\mathbf{x}}_l$.
6:     **end for**
7:     $k \mapsto k + p$.
8: **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 fulfil $\hat{G}(\mathbf{x}^{k+1}, \mathbf{x}^k) = 0$ 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 $\rho(H(\mathbf{x}^*)) < 1$, 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 $\rho(H(\mathbf{x}^*))$.*

The statement of Theorem 6.4 provides useful a priori information, even when it is not possible to ascertain in advance that $\rho(H(\mathbf{x}^*)) < 1$. For quadratic functionals F and the canonical tensor format U_{CP}, the situation $J = F \circ U_{CP}$ is considered in [19].

7 Numerical Experiments

In this section, we want to explicitly describe the nonlinear block Gauss-Seidel method for two different partitions of the coordinates.

We want 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. In terms of Eq. (14) U_{TC} with representation

rank $(r_1, \ldots, r_d) \in \mathbb{N}^d$ is given as

$$U_{TC} : \underset{\mu=1}{\overset{d-1}{\times}} (\mathbb{R}^{n_\mu})^{(r_\mu, r_{\mu+1})} \times (\mathbb{R}^{n_d})^{(r_1, r_d)} \quad \rightarrow \quad \mathcal{V},$$

$$(u_1, \ldots, u_d) \quad \mapsto \quad \sum_{j_1=1}^{r_1} \ldots \sum_{j_d=1}^{r_d} \bigotimes_{\mu=1}^{d-1} u_\mu(j_\mu, j_{\mu+1}) \otimes u_d(j_1, j_d) =: u$$

so J of (3) is defined as

$$J : \underset{\mu=1}{\overset{d-1}{\times}} (\mathbb{R}^{n_\mu})^{(r_\mu, r_{\mu+1})} \times (\mathbb{R}^{n_d})^{(r_1, r_d)} \quad \rightarrow \quad \mathbb{R},$$

$$(u_1, \ldots, u_d) \quad \mapsto \quad (F_a \circ U_{TC})(u_1, \ldots, u_1)$$

which we want to minimize in our experiments. We define

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

such that

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

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 as

$$X_\ell = \{0\} \times \ldots \times \{0\} \times P_\ell \times \{0\} \times \ldots \times \{0\} \tag{37}$$

for $1 \le \ell \le d$, where

$$P_\ell = \begin{cases} (\mathbb{R}^{n_\ell})^{(r_\ell, r_{\ell+1})} & \ell = 1, \ldots, d-1, \\ (\mathbb{R}^{n_d})^{(r_1, r_d)} & \ell = d. \end{cases}$$

16

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

$$A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) := \sum_{i_1,\ldots,i_{\ell-1}=1}^{R_1,\ldots,R_{\ell-1}} \sum_{i_{\ell+2},\ldots,i_d=1}^{R_{\ell+2},\ldots,R_d} \sum_{j_1,\ldots,j_{\ell-1}=1}^{r_1,\ldots,r_{\ell-1}} \sum_{j_{\ell+2},\ldots,j_d=1}^{r_{\ell+2},\ldots,r_d}$$

$$\left(\prod_{\mu=1,\mu\neq\ell}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_d, i_1, j_d, j_1)$$

and

$$B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) := \sum_{j_1=1}^{r_1} \cdots \sum_{j_{\ell-1}=1}^{r_{\ell-1}} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \cdots \sum_{j_d=1}^{r_d} \sum_{j'_1=1}^{r_1} \cdots \sum_{j'_{\ell-1}=1}^{r_{\ell-1}} \sum_{j'_{\ell+2}=1}^{r_{\ell+2}} \cdots \sum_{j'_d=1}^{r_d}$$

$$\left(\prod_{\mu=1,\mu\neq\ell}^{d-1} B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) \right) B_d(j_d, j_1, j'_d, j'_1)$$

which leads to a structure of equation (36) that pays respect to the partitioning (37):

$$J_\ell(u_\ell, u_\ell^c) = \frac{1}{\|a\|^2} \left(- \sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+1}=1}^{R_{\ell+1}} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+1}=1}^{r_{\ell+1}} A_\ell(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) \right.$$

$$\left. + \frac{1}{2} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+1}=1}^{r_{\ell+1}} \sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+1}=1}^{r_{\ell+1}} B_\ell(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) B_{[\ell]}(j_\ell, j_{\ell+1}, j'_\ell, j'_{\ell+1}) \right).$$

Now we can formulate the derivative with respect to the partitioning in a shorter notation, resulting in the derivative with respect to X_ℓ, with $u_\ell \in X_\ell$ and $u_\ell^c \in X_\ell^c$

$$\|a\|^2 \frac{\partial}{\partial u_\ell} J_l(u_\ell, u_\ell^c) = \left(- \sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+1}=1}^{R_{\ell+1}} a_\ell(i_\ell, i_{\ell+1}) A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) \right.$$

$$\left. + \sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+1}=1}^{r_{\ell+1}} u_\ell(j'_\ell, j'_{\ell+1}) B_{[\ell]}(j'_\ell, j'_{\ell+1}, j_\ell, j_{\ell+1}) \right)_{j_\ell, j_{\ell+1}}$$

such that setting this derivative equal to zero as in Eq. (34) in Algorithm 1, one has to solve the equation

$$\left(\sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+1}=1}^{R_{\ell+1}} a_\ell(i_\ell, i_{\ell+1}) A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) \right)_{j_\ell, j_{\ell+1}}$$

$$= \left(\sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+1}=1}^{r_{\ell+1}} u_\ell(j'_\ell, j'_{\ell+1}) B_{[\ell]}(j'_\ell, j'_{\ell+1}, j_\ell, j_{\ell+1}) \right)_{j_\ell, j_{\ell+1}},$$

which is equivalent to

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

where

$$\mathbf{a}_\ell \;:=\; \begin{pmatrix} a_\ell(1,1) \\ \vdots \\ a_\ell(R_\ell, R_{\ell+1}) \end{pmatrix}$$

$$\mathbf{A}_{[\ell]} \;:=\; \Big(A_{[\ell]}(i_\ell, i_{\ell+1}, j_\ell, j_{\ell+1}) \Big)_{(j_\ell, j_{\ell+1}),(i_\ell, i_{\ell+1})}$$

$$\mathbf{u}_\ell \;:=\; \begin{pmatrix} u_\ell(1,1) \\ \vdots \\ u_\ell(r_\ell, r_{\ell+1}) \end{pmatrix}$$

$$\mathbf{B}_{[\ell]} \;:=\; \Big(B_{[\ell]}(j'_\ell, j'_{\ell+1}, j_\ell, j_{\ell+1}) \Big)_{(j_\ell, j_{\ell+1}),(j'_\ell, j'_{\ell+1})}$$

and consequently, we have to solve

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

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 $r := \max_{1 \leq \mu \leq d}\{r_\mu\}$, $R := \max_{1 \leq \mu \leq d}\{R_\mu\}$ and $n := \max_{1 \leq \mu \leq d}\{n_\mu\}$.

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 chain inner product. If we take into account the connection between each entry, we can improve the complexity significantly. First, we introduce the definitions

$$A_\mu \;:=\; \begin{cases} \Big(A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \Big)_{(i_\mu, j_\mu),(i_{\mu+1}, j_{\mu+1})} \in \mathbb{R}^{R_\mu r_\mu \times R_{\mu+1} r_{\mu+1}} & 1 \leq \mu \leq d-1 \\[2mm] \Big(A_\mu(i_d, i_1, j_d, j_1) \Big)_{(i_d, j_d),(i_1, j_1)} \in \mathbb{R}^{R_d r_d \times R_1 r_1} & \mu = d \end{cases}$$

and

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

such that we can formulate the following lemma.

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

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

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.

18

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

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

$$y((i_2, j_2), (i_3, j_3)) := A_2(i_2, i_3, j_2, j_3)$$

which results in

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

such that we see

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

$$= A_2 \cdot \left(x((i_3, j_3), (i_1, j_1)) \right)_{(i_3,j_3),(i_1,j_1)}.$$

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*

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

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

$$A_\ell^{(k)} := \left(\langle a_\ell(i_\ell, i_{\ell+1}), u_\ell^{(k)}(j_\ell, j_{\ell+1}) \rangle \right)_{(i_{\ell+1}, j_{\ell+1}),(i_\ell, j_\ell)}$$

$$A_{>\ell}^{(k)} := \prod_{\mu=\ell+1}^{d} A_\mu^{(k)}, \qquad A_{<\ell}^{(k)} := \prod_{\mu=1}^{\ell-1} A_\mu^{(k)}$$

$$B_\ell^{(k)} := \left(\langle u_\ell^{(k)}(j_\ell, j_{\ell+1}), u_\ell^{(k)}(j_\ell', j_{\ell+1}') \rangle \right)_{(j_{\ell+1}, j_{\ell+1}'),(j_\ell, j_\ell')}$$

$$B_{>\ell}^{(k)} := \prod_{\mu=\ell+1}^{d} B_\mu^{(k)}, \qquad B_{<\ell}^{(k)} := \prod_{\mu=1}^{\ell-1} B_\mu^{(k)}$$

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*

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

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

19

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

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

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

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

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 $1 \leq \ell \leq d-1$.

Algorithm 2 Alternating Least Squares (ALS) Method for TC

1: 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 $\mathbf{g} := J(u^{(1)}), k := 1$.
2: **while** $\Delta \mathbf{g} > \varepsilon$ **do**
3:    $\tilde{B} := Id, \tilde{A} := Id$
4:    **for** $d-1 \geq \ell \geq 1$ **do**
5:       store $B_{>\ell}^{(k)} = B_{\ell+1}^{(k)} B_{>\ell+1}^{(k)}$ and $A_{>\ell}^{(k)} = A_{\ell+1}^{(k)} A_{>\ell+1}^{(k)}$
6:    **end for**
7:    **for** $1 \leq \ell \leq d$ **do**
8:       $\tilde{B} \mapsto \tilde{B} B_{\ell-1}^{(k+1)} \quad \{\Rightarrow \tilde{B} = B_{<\ell}^{(k+1)}\}$
9:       $\tilde{A} \mapsto \tilde{A} A_{\ell-1}^{(k+1)} \quad \{\Rightarrow \tilde{A} = A_{<\ell}^{(k+1)}\}$
10:      $u_\ell^{(k+1)} := \left(\left(reshape\left(B_{>\ell}^{(k)}\tilde{B}\right)\right)^{-1} reshape\left(A_{>\ell}^{(k)}\tilde{A}\right) \otimes Id_{n_\ell}\right)\mathbf{a}_\ell$
11:    **end for**
12:    $\mathbf{g} \leftarrow J(u^{(k+1)})$
13:    $k \mapsto k+1$
14: **end while**

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

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

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 $dr^4 + dr^2 R^2$.*

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

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

in terms of the complexity.

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 neighboured ones. So we are using the following partition of coordinates:

$$X_\ell = \begin{cases} \{0\} \times \ldots \times \{0\} \times P_\ell \times P_{\ell+1} \times \{0\} \times \ldots \times \{0\}, & 1 \leq \ell \leq d-1, \\ P_1 \times \{0\} \times \ldots \times \{0\} \times P_d, & \ell = d. \end{cases}$$

In contrary to ALS, we do not have a disjoint partition since

$$\begin{aligned} X_\ell \cap X_{\ell+1} &= \{0\} \times \ldots \times \{0\} \times P_{\ell+1} \times \{0\} \times \ldots \times \{0\} \text{ for } 1 \leq \ell \leq d-1 \\ &\text{and} \\ X_d \cap X_1 &= P_1 \times \{0\} \times \ldots \times \{0\}. \end{aligned}$$

This partitioning 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

$$\begin{aligned} A_{[\ell]}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) &:= \sum_{i_1=1}^{R_1} \cdots \sum_{i_{\ell-1}=1}^{R_{\ell-1}} \sum_{i_{\ell+3}=1}^{R_{\ell+3}} \cdots \sum_{i_d=1}^{R_d} \sum_{j_1=1}^{r_1} \cdots \sum_{j_{\ell-1}=1}^{r_{\ell-1}} \sum_{j_{\ell+3}=1}^{r_{\ell+3}} \cdots \sum_{j_d=1}^{r_d} \\ &\quad \left(\prod_{\mu=1, \mu \notin \{\ell, \ell+1\}}^{d-1} A_\mu(i_\mu, i_{\mu+1}, j_\mu, j_{\mu+1}) \right) A_d(i_1, i_d, j_1, j_d), \\[2ex] B_{[\ell]}(j_\ell, j_{\ell+2}, j'_\ell, j'_{\ell+2}) &:= \sum_{j_1=1}^{r_1} \cdots \sum_{j_{\ell-1}=1}^{r_{\ell-1}} \sum_{j_{\ell+3}=1}^{r_{\ell+3}} \cdots \sum_{j_d=1}^{r_d} \sum_{j'_1=1}^{r_1} \cdots \sum_{j'_{\ell-1}=1}^{r_{\ell-1}} \sum_{j'_{\ell+3}=1}^{r_{\ell+3}} \cdots \sum_{j'_d=1}^{r_d} \\ &\quad \left(\prod_{\mu=1, \mu \notin \{\ell, \ell+1\}}^{d-1} B_\mu(j_\mu, j_{\mu+1}, j'_\mu, j'_{\mu+1}) \right) B_d(j_1, j_d, j'_1, j'_d), \\[2ex] a_{\ell, \ell+1}(i_\ell, i_{\ell+2}) &:= \sum_{i_{\ell+1}=1}^{R_{\ell+1}} a_\ell(i_\ell, i_{\ell+1}) \otimes a_{\ell+1}(i_{\ell+1}, i_{\ell+2}) \in \mathbb{R}^{n_\ell \times n_{\ell+1}} \\ &\text{and} \\ u_{\ell, \ell+1}(j_\ell, j_{\ell+2}) &:= \sum_{j_{\ell+1}=1}^{r_{\ell+1}} u_\ell(j_\ell, j_{\ell+1}) \otimes u_{\ell+1}(j_{\ell+1}, j_{\ell+2}) \in \mathbb{R}^{n_\ell \times n_{\ell+1}} \end{aligned} \tag{38}$$

21

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

$$J_\ell(u_{\ell,\ell+1}, u^c_{\ell,\ell+1}) = \frac{1}{\|a\|^2} \left(-\sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+2}=1}^{R_{\ell+2}} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \langle a_{\ell,\ell+1}(i_\ell, i_{\ell+2}), u_{\ell,\ell+1}(j_\ell, j_{\ell+2}) \rangle A_{[\ell]}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) \right.$$

$$\left. +\frac{1}{2} \sum_{j_\ell=1}^{r_\ell} \sum_{j_{\ell+2}=1}^{r_{\ell+2}} \sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+2}=1}^{r_{\ell+2}} \langle u_{\ell,\ell+1}(j_\ell, j_{\ell+2}), u_{\ell,\ell+1}(j'_\ell, j'_{\ell+2}) \rangle B_{[\ell]}(j_\ell, j_{\ell+2}, j'_\ell, j'_{\ell+2}) \right).$$

The derivative with respect to the partition of coordinates is defined as

$$\|a\|^2 \frac{\partial}{\partial u_{\ell,\ell+1}} J_l(u_{\ell,\ell+1}, u^c_{\ell,\ell+1}) = \left(-\sum_{i_\ell=1}^{R_\ell} \sum_{i_{\ell+2}=1}^{R_{\ell+2}} a_{\ell,\ell+1}(i_\ell, i_{\ell+2}) A_{[\ell]}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) \right.$$

$$\left. +\sum_{j'_\ell=1}^{r_\ell} \sum_{j'_{\ell+2}=1}^{r_{\ell+2}} u_{\ell,\ell+1}(j'_\ell, j'_{\ell+2}) B_{[\ell]}(j'_\ell, j'_{\ell+2}, j_\ell, j_{\ell+2}) \right)_{j_\ell, j_{\ell+2}}$$

and setting this derivative 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

$$\mathbf{a}_{\ell,\ell+1} := \begin{pmatrix} a_{\ell,\ell+1}(1,1) \\ \vdots \\ a_{\ell,\ell+1}(R_\ell, R_{\ell+2}) \end{pmatrix}$$

$$\mathbf{A}_{[\ell]} := \left(A_{[\ell]}(i_\ell, i_{\ell+2}, j_\ell, j_{\ell+2}) \right)_{(j_\ell, j_{\ell+2}),(i_\ell, i_{\ell+2})}$$

$$\mathbf{u}_{\ell,\ell+1} := \begin{pmatrix} u_{\ell,\ell+1}(1,1) \\ \vdots \\ u_{\ell,\ell+1}(r_\ell, r_{\ell+2}) \end{pmatrix}$$

$$\mathbf{B}_{[\ell]} := \left(B_{[\ell]}(j'_\ell, j'_{\ell+2}, j_\ell, j_{\ell+2}) \right)_{(j_\ell, j_{\ell+2}),(j'_\ell, j'_{\ell+2})}$$

such that we have to solve

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

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:

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

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,*

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

and

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

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 $\left(\mathbf{B}_{[\ell]}^{-1} \mathbf{A}_{[\ell]} \otimes Id_{n_\ell \times n_{\ell+1}} \right) \mathbf{a}_{\ell,\ell+1}$ is in*

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

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.

Algorithm 3 DMRG Method for TC

1: 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 $\mathbf{g} := J(u^{(1)})$, $k := 1$.
2: **while** $\Delta \mathbf{g} > \varepsilon$ **do**
3:     $\mathbf{u}_{d,1}^{(k+1)} := \left(\left(reshape\left(\prod_{\mu=2}^{d-1} B_\mu \right) \right)^{-1} reshape\left(\prod_{\mu=2}^{d-1} A_\mu \right) \otimes Id_{n_d \times n_1} \right) \mathbf{a}_{d,1}$
4:     $[u_d^{(k+1)}, u_1^{(k+1)}] := SVD\left(reshape\left(\mathbf{u}_{d,1}^{(k+1)} \right) \right)$
5:     $\tilde{B} := Id, \tilde{A} := Id$
6:     **for** $d-1 \geq \ell \geq 2$ **do**
7:         store $B_{>\ell}^{(k)} = B_{\ell+1}^{(k)} B_{>\ell+1}^{(k)}$ and $A_{>\ell}^{(k)} = A_{\ell+1}^{(k)} A_{>\ell+1}^{(k)}$
8:     **end for**
9:     **for** $1 \leq \ell \leq d-1$ **do**
10:        $\tilde{B} \mapsto \tilde{B} B_{\ell-1}^{(k+1)} \{\Rightarrow \tilde{B} = B_{<\ell}^{(k+1)}\}$
11:        $\tilde{A} \mapsto \tilde{A} A_{\ell-1}^{(k+1)} \{\Rightarrow \tilde{A} = A_{<\ell}^{(k+1)}\}$
12:        $\mathbf{u}_{\ell,\ell+1}^{(k+1)} := \left(\left(reshape\left(B_{>\ell+1}^{(k)} \tilde{B} \right) \right)^{-1} reshape\left(A_{>\ell+1}^{(k)} \tilde{A} \right) \otimes Id_{n_\ell \times n_{\ell+1}} \right) \mathbf{a}_{\ell,\ell+1}$
13:        $[u_\ell^{(k+1)}, u_{\ell+1}^{(k)}] := SVD\left(reshape\left(\mathbf{u}_{\ell,\ell+1}^{(k+1)} \right) \right)$
14:     **end for**
15:     $\mathbf{g} \leftarrow J(u^{(k+1)})$
16:     $k \mapsto k + 1$
17: **end while**

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

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

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

Remark 7.11 (Initial guesses)**.** *It is especially important for algorithms that do not adjust the rank (like ALS) to start with an approximation that is already relatively close to the solution. Let $(r_1, \ldots, r_d)$ be our representation ranks just as before. To efficiently generate a starting value, we are taking the given tensor a as the initial approximation value such that u has the same initial representation rank as a. For each summation $1 \leq \mu \leq d$ we want to perform an adaptive cross approximation (ACA, see [1]) as follows:*

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

After approximating edge μ, the representation rank of u is reduced to $(R_1, \ldots, R_{\mu-1}, r_\mu, R_{\mu+1}, \ldots, R_d)$. Successively applying this scheme to all edges is resulting in u with the desired representation rank.

For the DMRG algorithm however, we do not need to compute an initial guess with non trivial ranks, we simply use a rank 1 tensor as the init guess and let the algrithm find the ranks.

7.3 Tables

The results, that we want to present here were are in the following way. In the first step, we convert the full tensor with the scheme that is described in [20] into a MPS/TT (with an accuracy of 10^{-12}) which we interpret as a tensor chain where the last summation index is 1. The source code can be found in [4]. One iteration is one complete cycle.

Table 1: Reduced representation ranks for AO integrals in H_2O using different basis sets 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 different basis sets 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

References

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

Table 3: Reduced representation ranks for AO integrals in H_2O using different basis sets 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

[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, 2011. http://gitorious.org/tensorcalculus.

[5] A. Falcó and W. Hackbusch. On minimal subspaces in tensor representations. 2010. MIS Prerint 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. to be published.

[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 nd 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 and B. N. Khoromskij. DMRG and QTT approach to high-dimensional quantum molecular dynamics. 2010. MIS preprint 69.

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

[17] 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.

[18] 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.

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

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

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.