DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Local convergence of alternating schemes for optimization of convex problems in the TT format

T. Rohwedder, A. Uschmajew

Preprint 112

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Local convergence of alternating schemes for optimization of convex problems in the TT format

T. Rohwedder, A. Uschmajew

Preprint 112

LOCAL CONVERGENCE OF ALTERNATING SCHEMES FOR OPTIMIZATION OF CONVEX PROBLEMS IN THE TT FORMAT

THORSTEN ROHWEDDER† AND ANDRÉ USCHMAJEW†

Abstract. Alternating linear schemes (ALS), with the Alternating Least Squares algorithm a notable special case, provide one of the simplest and most popular choices for the treatment of optimization tasks by tensor methods. An according adaptation of ALS for the recent TT (= tensor train) format (Oseledets, 2011) has recently been investigated in (Holtz, Rohwedder, Schneider, 2011). With the present work, the positive practical experiences with TT-ALS is backed up with an according local linear convergence theory for the optimization of convex functionals J. The main assumption entering the proof is that the redundancy introduced by the TT parametrization τ matches the null space of the Hessian of the induced functional $j = J \circ \tau$, and we give conditions under which this assumption can be expected to hold.

Key words. ALS, high-dimensional optimization, local convergence, nonlinear Gauss-Seidel, tensor product approximation, tensor train, TT decomposition

AMS subject classifications. 15A69, 65K10, 90C06

1. Introduction. In many application areas, the treatment of the respective governing equations amounts to the treatment of discrete tensors, i.e. of high dimensional quantities $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. For example, such problems arise in the context of life sciences and physics in the discretization of functions from tensor spaces, often defined implicitly as the solution of e.g. partial differential or integral equations, with various Schrödinger equation type models and the Fokker-Planck equation providing prominent examples. Another field where problems posed on high-dimensional spaces turn up naturally is the active field of data mining problems. Since the standard approaches to all such problems have a computational complexity growing exponentially in the dimension d of the tensors, the only effective remedy is often the use of a data-sparse representation or approximation of the tensors exploiting concepts of tensor product approximations, and the development of such concepts has consequently become an important and active field of mathematical research during the last years.

As a classical approach to multi-dimensional problems, the canonical format (also known as CANDECOMP/PARAFAC model, [26]) is extensively used in practical applications for extraction of information, i.e. linear least squares problems. This is contrasted by a lack of desirable theoretical properties [5]. Also, for treatment of more complex equations as optimization problems, application of this format is possible [10], but has to be stabilized by somewhat artificial techniques [9]; when it comes to treatment of equations based on a manifold approach [17], CANDECOMP lacks the basic property of being an embedded manifold, meaning that a stable, non-redundant parametrization of the set of rank-r-tensors cannot be given, ruling out the canonical format for such approaches. An alternative is provided by the – also quite classical – Tucker format [35], with the set of Tucker tensors of bounded rank forming a weakly closed set [11] with manifold properties [25], fostering for instance its application in quantum chemistry [4]. In practical application of the Tucker format, one is then unfortunately confronted with the problem that although it sometimes reduces complexity immensely, the ansatz still scales exponentially in d, often prohibiting its

†Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (rohwedde@math.tu-berlin.de, uschmajew@math.tu-berlin.de). This research was supported by DFG (Deutsche Forschungsgemeinschaft) through SPP 1324.

application to problems of higher dimension d. Hackbusch therefore generalized the idea of subspace approximation, being the basis of the Tucker decomposition, to a hierarchical ansatz proposed in [16]: High-dimensional tensors are therein successively decomposed to component tensors of order 3, which then can be treated separately, so that the complexity stays linear or at least polynomial in the dimension d. Many desirable theoretical features of this approach have recently been verified: Results on the existence of best approximations and minimizers of convex functionals from [11, 12] also apply to the hierarchical Tucker format; extending results from [19], manifold properties can be verified [37].

From the practical side, two notable special cases of the hierarchichal ansatz have received major attention: The HT format as used by Grasedyck [14], using a symmetric decomposition tree, and the TT format, as introduced independently by Tyrtyshnikov and Oseledets [30, 32], using a linear decomposition tree. For both approaches, the applicability to various problems has lately been investigated with promising results, see e.g. [1, 2, 20, 24, 27]; while the former approach seems to be superior on some problems in the sense that it sometimes provides lower ranks [15], the latter has advantage of being conceptually quite simple, thus often allowing for a simpler analysis from the theoretical point of view, and enabling, from the practical side, a robust sequential treatment of resulting equations to be solved for the component tensors. In particular, one of the authors of the present publication was co-author of the previous work [20], in which the well-known Alternating Least Squares (ALS) has been adapted to the treatment of more global high-dimensional optimization tasks like linear equations, eigenvalue equations etc. using the TT format. We found that stable equations can be derived from concept of left- and right-orthogonality; proceeding in "sweeps" inspired by the DMRG algorithm used in quantum chemistry [34, 40] gives an iteration scheme that we termed the Alternating Linear Scheme for optimization in the TT format (also abbreviated by ALS, introduced alongside with a modified variant MALS enabling dynamical rank-adaptation, see [20]). Our practical experience, as partly reported in [20], is that the ALS shows extraordinarily nice convergence behaviour similar to that of ALS applied to the Tucker format. Although other approaches based on tangent space of the TT manifold may be more suitable for treatment of e.g. differential equations [17, 19], and although ALS might be refined by using locally quadratically convergent methods near the sought optimum, the ALS approach is in fact at the moment our basic method of choice for the treatment of optimization problems in the sense that even with its striking simplicity, it converges stably, efficiently and reliable to stationary points of the parametrized problem. Of course, these points might not be the global minima on the approximation manifold under consideration – it has recently even be shown that the computation of global rank-1-best approximations is an NP-hard problem [18]. The occurrence of local minima is a (not solely theoretical) problem common to many tensor optimization methods; in the case of ALS for the TT format, it is often treated by random perturbation of computed stationary points (e,g, in practical applications from quantum chemistry), and has for the DMRG/MALS algorithm also readily been attacked in the recent publication [6] in a more systematic way.

The aim of the present paper is to back-up our positive practical experiences with ALS with an according theoretical investigation of the convergence behaviour of the algorithm when applied in the above form to the treatment of convex optimization problems in the TT format. Our practical observation in [20] was that ALS provides a very robust linear convergence behaviour.

In this work we deliver the according theoretical result, by rigorously proving local linear convergence under certain reasonable positive definiteness conditions on the Hessian of the functional to be minimized. Our proof is an adaptation of the concept followed in the recent paper [36], where one of the authors showed local convergence of PARAFAC-ALS: in the neighborhood of a solution, the ALS algorithm is identified as a perturbation of the linear block Gauss-Seidel method applied to the Hessian at the solution [3, 29, 33]. This Hessian is only semidefinite due to the non-uniqueness of the TT representation. In contrast to the linear case, one has to completely remove the null space of the Hessian from the iteration to keep the contractive properties of the Gauss-Seidel method. This can be achieved by introducing a local normalization operator which chooses a unique representation for the TT tensors. While this is necessary to link the convergence analysis to existing results on the nonlinear Gauss-Seidel method [23, 28], it turns out that the sequence of TT tensors generated by the ALS algorithm is independent of the choice of their representations therein (Proposition 3.3 below). From the convergence result for *one* choice of representations one hence obtains a practically relevant convergence result for whole equivalence class of *all* possible representations. In other words, a convergence result for ALS regarded as an algorithm on the manifold of tensors of fixed TT rank.

In short, the content of the paper can be outlined as follows: We start off in Section 2 with a collection of facts about the TT-format that later enter the proof of convergence; aside from ideas based on the previous work [19], we also introduce the concepts of orbits and normalization. In Section 3, we devise a generic ALS algorithm for which we prove convergence under assumptions on the solution and on the Hessian of the functional (Theorem 3.7 and Corollaries 3.8, 3.9). Based on this, convergence for the ALS from [20] is deduced (Theorem 3.10), and some supplementary results are provided in Sections 3.4 and 3.5. Sufficient conditions for the main assumption on the Hessian to hold are presented in Section 4.

2. The TT tensor format.

Although the basic idea of our convergence theorem is quite simple, a certain amount of technicalities and notions is needed to formulate it. In this section we first recall the TT (tensor train) format [30, 32].

2.1. The TT decomposition.

Let $d \in \mathbb{N}$ and $n_1, n_2, \ldots, n_d \in \mathbb{N}$ be given. We use the notation $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ for the space of d-th order tensors and treat its elements $\mathbf{X}$ as d-dimensional arrays, whose entries are indexed by

$$\mathbf{X}(x_1, x_2, \ldots, x_d), \qquad x_i = 1, 2, \ldots, n_i, \quad i = 1, 2, \ldots, d.$$

Let $\mathbf{r} = (r_1, r_2, \ldots, r_{d-1})$ where $0 < r_i \leq n_i$ are integers. Further we set $r_0 = r_d = 1$. The elements of the space

$$\overline{\mathcal{U}} = \bigtimes_{i=1}^{d} \mathbb{R}^{r_{i-1} \times n_i \times r_i}$$

are denoted by $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d)$. The x_i-th lateral slice of a third-order component tensor $\mathbf{U}_i \in \mathbb{R}^{r_{i-1} \times n_i \times r_i}$ will be denoted by $U_i(x_i)$, that is, for x_i fixed $U_i(x_i)$ is the $r_{i-1} \times r_i$ matrix given by

$$[U_i(x_i)]_{k_1 k_2} = \mathbf{U}_i(k_1, x_i, k_2).$$

Consider the mapping

$$\tau \colon \overline{\mathcal{U}} \to \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \colon \mathbf{U} \mapsto \tau(\mathbf{U}) = \tau(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d)$$

given pointwise by

$$\tau(\mathbf{U})(x_1, x_2, \dots, x_d) := U_1(x_1)U_2(x_2)\dots U_d(x_d). \tag{2.1}$$

Since $r_0 = r_d = 1$, this matrix product indeed results in a real number for each multi-index $(x_1, x_2, \dots, x_d)$.

DEFINITION 2.1 ($\mathrm{TT}_{\leq \mathbf{r}}$ format). *A tensor of the form* $\mathbf{X} = \tau(\mathbf{U})$ *is called a* $\mathrm{TT}_{\leq \mathbf{r}}$ tensor *or a* tensor in $\mathrm{TT}_{\leq \mathbf{r}}$ format*, and* $\mathbf{U}$ *is called a* $\mathrm{TT}_{\leq \mathbf{r}}$ decomposition *of* $\mathbf{X}$. *The image of* τ *is denoted by*

$$\mathcal{T}_{\leq \mathbf{r}} = \tau(\overline{\mathcal{U}}) \subseteq \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$$

and is called the set of TT tensors of rank at most $\mathbf{r}$.

The mapping $(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d) \mapsto \tau(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d)$ is multilinear, so that the $\mathrm{TT}_{\leq \mathbf{r}}$ decomposition has a lot of nice structural properties which are similar to those of rank-one tensors, and thus allow for an analogous treatment, at least to some extent. In fact, rank-one tensors are $\mathrm{TT}_{\leq \mathbf{r}}$ tensors where all $r_i = 1$.

2.2. Unfoldings and TT rank. If we treat a tensor $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ as a matrix

$$\mathbf{X}^{(i)} \in \mathbb{R}^{(n_1 \cdot \ldots \cdot n_i) \times (n_{i+1} \cdot \ldots \cdot n_d)},$$

then this is called the i-th canonical unfolding of $\mathbf{X}$. Its matrix rank is called the i-th separation rank of $\mathbf{X}$. To fix the ordering of multi-indices we choose the reverse lexicographical ordering, e.g., $(1, 1), (2, 1), (1, 2), (2, 2)$.

For $\mathbf{U} \in \overline{\mathcal{U}}$ and $i = 1, 2, \dots, d$ we further define the matrices

$$\mathbf{U}^{\leq i} = [U_1(x_1)U_2(x_2)\dots U_i(x_i)] \in \mathbb{R}^{n_1 n_2 \dots n_i \times r_i} \tag{2.2}$$

and

$$\mathbf{U}^{\geq i} = [U_i(x_i)U_{i+1}(x_{i+1})\dots U_d(x_d)] \in \mathbb{R}^{r_{i-1} \times n_i n_{i+1} \dots n_d}. \tag{2.3}$$

In $\mathbf{U}^{\leq i}$ the row vectors $U_1(x_1)U_2(x_2)\dots U_i(x_i)$ are stacked below each other (in reverse lexicographical order). Analogously, the columns $U_i(x_i)U_{i+1}(x_{i+1})\dots U_d(x_d)$ of $\mathbf{U}^{\geq i}$ are arranged. According to (2.1), the i-th unfolding of a $\mathrm{TT}_{\leq \mathbf{r}}$ tensor $\mathbf{X} = \tau(\mathbf{U})$ can be written as

$$\mathbf{X}^{(i)} = \mathbf{U}^{\leq i}\mathbf{U}^{\geq i+1}. \tag{2.4}$$

Hence, the i-th separation rank of a $\mathrm{TT}_{\leq \mathbf{r}}$ tensor $\mathbf{X} = \tau(\mathbf{U})$ is at most r_i. On the other hand, if $\operatorname{rank} \mathbf{X}^{(i)} \leq r_i$ for $i = 1, 2, \dots, d-1$, it is always possible, using the successive SVD algorithm of Oseldets [30], to find $\mathbf{U} \in \overline{\mathcal{U}}$ such that $\mathbf{X} = \tau(\mathbf{U})$ is a $\mathrm{TT}_{\leq \mathbf{r}}$ decomposition. This makes the following definition meaningful.

DEFINITION 2.2 (TT rank). *A tensor* $\mathbf{X}$ *has* TT rank $\mathbf{r} = (r_1, r_2, \dots, r_{d-1})$, *if* $\operatorname{rank} \mathbf{X}^{(i)} = r_i$ *for* $i = 1, 2, \dots, d-1$. *The values* r_i *are called* TT ranks *of* $\mathbf{X}$.

Note that we have the trivial bounds

$$r_i \leq \min\Big(\prod_{j=1}^{i} n_j, \prod_{j=i+1}^{d} n_j \Big) \tag{2.5}$$

for the TT ranks.

It follows from the above considerations that

$$\mathcal{T}_{\leq \mathbf{r}} = \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \mid \text{TT-rank}(\mathbf{X}) \leq \mathbf{r}\},$$

where the inequality is understood element-wise. By the semicontinuity of matrix rank, this shows that $\mathcal{T}_{\leq \mathbf{r}}$ is closed.

Another useful description of the TT rank was given in [19]. If we look at the component tensors $\mathbf{U}_i$ of a $\text{TT}_{\leq \mathbf{r}}$ decomposition $\mathbf{X} = \tau(\mathbf{U})$, we can unfold them either into the matrices

$$\mathbf{U}_i^L = \begin{bmatrix} U_i(1) \\ U_i(2) \\ \vdots \\ U_i(n_i) \end{bmatrix} \in \mathbb{R}^{(r_{i-1}n_i \times r_i)}, \tag{2.6}$$

or into

$$\mathbf{U}_i^R = \begin{bmatrix} U_i(1) & U_i(2) & \cdots & U_i(n_i) \end{bmatrix} \in \mathbb{R}^{r_{i-1} \times n_i r_i}.$$

These operations are called *left* and *right unfolding* of $\mathbf{U}_i$, respectively. With respect to the inverse lexicographical ordering the following recursive relations hold:

$$\mathbf{U}^{\leq i} = (I_{n_i} \otimes \mathbf{U}^{\leq i-1})\mathbf{U}_i^L \tag{2.7}$$

and

$$\mathbf{U}^{\geq i} = \mathbf{U}_i^R(I_{n_i} \otimes \mathbf{U}^{\geq i+1}) \tag{2.8}$$

(where one should set $\mathbf{U}^{\leq 0} = \mathbf{U}^{\geq d+1} = 1$). Using (2.1) and (2.7) inductively, one can characterize the TT ranks as follows [19].

PROPOSITION 2.3. *A tensor $\mathbf{X} = \tau(\mathbf{U})$ in the $TT_{\leq \mathbf{r}}$ format has TT rank $\mathbf{r}$ if and only if for $i = 1, 2, \ldots, d-1$ it holds*

$$\operatorname{rank} \mathbf{U}_i^L = \operatorname{rank} \mathbf{U}_{i+1}^R = r_i. \tag{2.9}$$

Note that this implies the relations

$$r_i \leq r_{i-1}n_i \qquad \text{and} \qquad r_i \leq r_{i+1}n_{i+1}, \tag{2.10}$$

which, by induction, are equivalent to (2.5). The TT ranks r_i of a tensor $\mathbf{X}$ are hence not unrelated among each other. Since $r_0 = r_d = 1$, they can first increase for growing i but have to decrease from a certain index on.

In what follows we will mainly focus on tensors of fixed TT rank $\mathbf{r}$. According to the above proposition, they are parametrized by the set

$$\mathcal{U} = \{\mathbf{U} \in \overline{\mathcal{U}} \mid \operatorname{rank} \mathbf{U}_i^L = \operatorname{rank} \mathbf{U}_{i+1}^R = r_i \text{ for } i = 1, 2, \ldots, d-1\},$$

which is open and dense in $\overline{\mathcal{U}}$. Here it is silently assumed that $\mathcal{U}$ is not empty.

PROPOSITION 2.4. *The set*

$$\mathcal{T}_{\mathbf{r}} = \{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \mid \text{TT-rank}(\mathbf{X}) = \mathbf{r}\} = \tau(\mathcal{U})$$

is an embedded submanifold in $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ of dimension $\dim \overline{\mathcal{U}} - \sum_{i=1}^{d-1} r_i^2$ and $\tau \colon \mathcal{U} \to \mathcal{T}_{\mathbf{r}}$ is a submersion, that is, $\tau|_{\mathcal{U}}$ is of constant rank $\dim \overline{\mathcal{U}} - \sum_{i=1}^{d-1} r_i^2$.

Note that submersion are always open maps [7, (16.7.5)]. Hence τ maps open sets in $\mathcal{U}$ onto open sets in $\mathcal{T}_{\mathbf{r}}$. We do not give a proof of the above proposition here, but refer to [37]. In the earlier work [19] it is shown that $\mathcal{T}_{\mathbf{r}}$ is an immersed submanifold.

We conclude with the remark that $\mathcal{T}_{\mathbf{r}}$ is not closed, but its closure is the set $\mathcal{T}_{\leq\mathbf{r}}$ of all decomposable tensors. This holds because, by the continuity of τ, the closure of $\mathcal{T}_{\mathbf{r}}$ contains $\tau(\overline{\mathcal{U}}) = \mathcal{T}_{\leq\mathbf{r}}$, which is already closed.

2.3. Equivalent TT representations. One difficulty of the $\mathrm{TT}_{\leq\mathbf{r}}$ format is that representations are highly nonunique. Clearly, by looking at (2.1), we have $\tau(\mathbf{U}) = \tau(\hat{\mathbf{U}})$ whenever it holds

$$\hat{U}_1(x_1) = U_1(x_1)A_1, \quad \hat{U}_d(x_d) = A_{d-1}^{-1}U_d(x_d), \quad \text{and} \quad \hat{U}_i(x_i) = A_{i-1}^{-1}U_i(x_i)A_i \tag{2.11}$$

for $i = 2, 3, \dots, d-1$, where the A_i are nonsingular $r_i \times r_i$ matrices. For rank-one tensors this is called *scaling indeterminacy* and we will use this terminology for TT tensors as well. The question whether the scaling indeterminacy (2.11) is the only kind of nonuniqueness has a simple answer.

PROPOSITION 2.5. *A $TT_{\leq\mathbf{r}}$ decomposition $\mathbf{X} = \tau(\mathbf{U})$ is unique up to the scaling indeterminacy* (2.11) *if and only if $\mathbf{X}$ has TT rank $\mathbf{r}$.*

Proof. The necessity of this condition follows from the fact that a $\mathrm{TT}_{\leq\tilde{\mathbf{r}}}$ decomposition with $\tilde{r}_i \leq r_i$ can be artificially extended to a $\mathrm{TT}_{\leq\mathbf{r}}$ decomposition by adding zero blocks to the third-order components $\mathbf{U}_i$. This is then not covered by the operation (2.11). On the other hand, if $\operatorname{rank} \mathbf{X}^{(i)} = r_i$ for $i = 1, 2, \dots, d-1$, then (2.4) implies $\hat{\mathbf{U}}^{\leq i} = \mathbf{U}^{\leq i}A_i$ and $\hat{\mathbf{U}}^{\geq i+1} = A_i^{-1}\mathbf{U}^{\geq i+1}$ for nonsingular A_i. Using (2.7) one finds that this is equivalent to (2.11), which proves the sufficiency of the condition. ◻

The scaling operation (2.11) can be regarded as the action of the Lie group

$$\mathcal{G} = \bigtimes_{i=1}^{d-1} GL(r_i)$$

on $\mathcal{U}$, which we denote by

$$\theta \colon \mathcal{U} \times \mathcal{G} \to \mathcal{U} \colon (\mathbf{U}, \mathsf{A}) \mapsto \hat{\mathbf{U}} =: \theta(\mathbf{U}, \mathsf{A}),$$

with $\hat{\mathbf{U}}$ defined by (2.11). Obviously θ is continuous. For fixed $\mathbf{U} \in \mathcal{U}$ we use the shorthand $\theta_{\mathbf{U}}$ for the map $\mathsf{A} \mapsto \theta(\mathbf{U}, \mathsf{A})$, and denote by

$$\mathcal{M}_{\mathbf{U}} = \{\hat{\mathbf{U}} = \theta_{\mathbf{U}}(\mathsf{A}) \mid \mathsf{A} \in \mathcal{G}\}$$

the orbits of the group action. Proposition 2.5 states that

$$\mathcal{M}_{\mathbf{U}} = \{\hat{\mathbf{U}} \in \overline{\mathcal{U}} \mid \tau(\hat{\mathbf{U}}) = \tau(\mathbf{U})\}. \tag{2.12}$$

By showing that $\theta_{\mathbf{U}}$ is an embedding (that is, an immersion which is also an homeomorphism onto $\mathcal{M}_{\mathbf{U}}$), one can prove that $\mathcal{M}_{\mathbf{U}}$ is an embedded (not connected) submanifold of $\mathcal{U}$ of dimension $\sum_{i=1}^{d-1} r_i^2$ (cf. [37]). The tangent space of $\mathcal{M}_{\mathbf{U}}$ at $\mathbf{U}$ will be denoted by $T\mathcal{M}_{\mathbf{U}}$. Using the fact that the derivative of the matrix inverse $A \mapsto A^{-1}$ at a point A is the linear mapping $H \mapsto -A^{-1}HA^{-1}$, one calculates

$$\begin{aligned}\theta'_{\mathbf{U}}(\mathsf{A})[\mathsf{H}] = (&\mathbf{U}_1H_1\,,\ -A_1^{-1}H_1A_1^{-1}\mathbf{U}_2A_2 + A_1^{-1}\mathbf{U}_2H_2, \dots \\ &\dots, -A_{d-2}^{-1}H_{d-2}A_{d-2}^{-1}\mathbf{U}_{d-1}A_{d-1} + A_{d-2}^{-1}\mathbf{U}_{d-1}H_{d-1}\,,\ -A_{d-1}^{-1}H_{d-1}A_{d-1}^{-1}\mathbf{U}_{d-1}),\end{aligned}$$

where $\mathsf{H} = (H_1, H_2, \ldots, H_{d-1}) \in \bigtimes_{i=1}^{d-1} \mathbb{R}^{r_i \times r_i}$, and expressions like $A\mathbf{U}_i B$ with matrices A, B are understood slicewise as $AU_i(x_i)B$.[1] Evaluating this formula at the identity $\mathsf{I} = (I_{r_1}, I_{r_2}, \ldots, I_{r_{d-1}})$ gives

$$\begin{aligned} T\mathcal{M}_{\mathbf{U}} &= \left\{ \theta'_{\mathbf{U}}(\mathsf{I})[\mathsf{H}] \,\middle|\, \mathsf{H} \in \bigtimes_{i=1}^{d-1} \mathbb{R}^{r_i \times r_i} \right\} \\ &= \Big\{ (\mathbf{U}_1 H_1, -H_1 \mathbf{U}_2 + \mathbf{U}_2 H_2, \ldots \\ &\qquad \ldots, -H_{d-2}\mathbf{U}_{d-1} + \mathbf{U}_{d-1} H_{d-1}, -H_{d-1}\mathbf{U}_{d}) \,\Big|\, \mathsf{H} \in \bigtimes_{i=1}^{d-1} \mathbb{R}^{r_i \times r_i} \Big\}. \end{aligned} \tag{2.13}$$

2.4. Normalization. Elements in the same orbit parametrize the same TT tensor and are therefore called *equivalent*. It would be convenient and useful to find a normal form of TT rank $\mathbf{r}$ tensors which fixes a representation within the orbits. An ideal *normalization operator* $R\colon \mathcal{U} \to \mathcal{U}$ would have the properties (i) $R(\mathbf{U}) \in \mathcal{M}_{\mathbf{U}}$ and (ii) $R|_{\mathcal{M}_{\mathbf{U}}}$ is constant for all $\mathbf{U} \in \mathcal{U}$. Since $\mathcal{M}_{\mathbf{U}}$ is not connected, one might only require that $R|_{\mathcal{M}_{\mathbf{U}}}$ is constant on the connected components of $\mathcal{M}_{\mathbf{U}}$, but even then it is not clear whether such an operator exists. For our purposes, we only need a local variant. We use the shorthand $\mathsf{A}(\mathbf{U})$ for a function $\overline{\mathcal{U}} \to \mathcal{G}$.

Definition 2.6. *An operator of the form*

$$R\colon \mathcal{O} \subseteq \mathcal{U} \to \mathcal{U}\colon \mathbf{U} \mapsto \theta_{\mathbf{U}}(\mathsf{A}(\mathbf{U}))$$

defined on an open subset $\mathcal{O}$ of $\mathcal{U}$ is called local normalization operator *if*

(i) $R^2(\mathbf{U}) = R(\mathbf{U})$ *for all* $\mathbf{U} \in \mathcal{O}$,
(ii) R *is smooth in a neighborhood of its fixed points,*
(iii) $R|_{\mathcal{M}_{\mathbf{U}}}$ *is constant in a neighborhood of fixed points* $\mathbf{U}$.

The existence of local normalization operators is guaranteed.

Proposition 2.7. *For every $\mathbf{U}^* \in \mathcal{U}$ there exists an open neighborhood $\mathcal{O}$ of $\mathbf{U}^*$ and a local normalization operator $R_{\mathbf{U}^*}\colon \mathcal{O} \to \mathcal{U}$ such that $R_{\mathbf{U}^*}(\mathbf{U}^*) = \mathbf{U}^*$.*

Surely, this assertion can be pieced together from similar results in textbooks (see, e.g., [7, (16.10.3.2)], but one fist has to verify that the Lie group action is proper, which is trivial here). The intuitive idea behind the claim is that the orbits $\mathcal{M}_{\mathbf{U}}$ completely fill out $\mathcal{U}$. By putting for fixed $\mathbf{U}^*$ a submanifold $\mathcal{N}_{\mathbf{U}^*}$ of codimension $\dim \mathcal{M}_{\mathbf{U}^*}$ through $\mathbf{U}^*$ and transversal to $\mathcal{M}_{\mathbf{U}^*}$, we can define $R_{\mathbf{U}^*}$ by the property $R_{\mathbf{U}^*}(\mathbf{V}) = \theta_{\mathbf{V}}(\mathsf{A}(\mathbf{V})) \in \mathcal{N}_{\mathbf{U}^*}$. This is illustrated in Figure 2.1. For the most obvious choice $\mathcal{N}_{\mathbf{U}^*} = \mathbf{U}^* + W_{\mathbf{U}^*}$, where $W_{\mathbf{U}^*}$ is any complementary linear space to $T\mathcal{M}_{\mathbf{U}^*}$ (see for instance (4.1)), the existence of such a function $\mathsf{A}(\mathbf{V})$ for $\mathbf{V}$ in a neighborhood of $\mathbf{U}^*$ can be easily obtained from the implicit function theorem.

Alternatively, taking Proposition 2.4 for granted, it follows from standard theorems (e.g. [7, (16.8.8)]) that a neighborhood $\mathcal{O}(\mathbf{U}^*)$ of $\mathbf{U}^*$ in $\mathcal{N}_{\mathbf{U}^*} = \mathbf{U}^* + W_{\mathbf{U}^*}$ is diffeomorphic to a neighborhood $\mathcal{O}(\mathbf{X})$ of $\mathbf{X}^* = \tau(\mathbf{U}^*)$ in $\mathcal{T}_{\mathbf{r}}$ via the mapping $\tau|_{\mathcal{N}_{\mathbf{U}^*}}$. One then can define $R_{\mathbf{U}^*} = (\tau|_{\mathcal{N}_{\mathbf{U}^*}})^{-1} \circ \tau$ on $\mathcal{O}(\mathbf{U}^*)$. Note that $R_{\mathbf{U}^*}$ then will be a local normalization operator for all $\mathbf{V} \in \tau|_{\mathcal{N}_{\mathbf{U}^*}}^{-1}(\mathcal{O}(\mathbf{X}))$. These $\mathbf{V}$ are the unique intersection points of $\mathbf{U}^* + W_{\mathbf{U}^*}$ with orbits close to $\mathbf{U}^*$. Conversely, one can show that

[1] Denoting the inverse operations of the left and right unfolding by superscipts $-L$ and $-R$ respectively, it holds $A\mathbf{U}_i B = \{A[(\mathbf{U}_i^L B)^{-L}]^R\}^{-R}$.

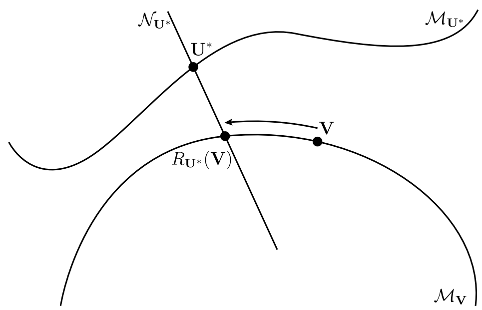

FIG. 2.1. *The local normalization operator* $R_{\mathbf{U}^*}$

for any $\mathbf{X}^* = \tau(\mathbf{U}^*) \in \mathcal{T}_{\mathbf{r}}$ and $\mathbf{X}$ close enough to $\mathbf{X}^*$, the corresponding neighborhood $\mathcal{O}(\mathbf{X})$ contains $\mathbf{X}^*$. In this way one can at least construct a local normalization operator for some $\mathbf{U}^* \in \mathcal{M}_{\mathbf{U}^*}$ without exactly knowing $\mathbf{U}^*$, which might be the unknown solution of a minimization problem as discussed later.

Finding an explicit instance of a "self consistent" local normalization operator, which does not depend on a certain $\mathbf{U}^*$, is a nontrivial task. Usually, one uses (2.11) to choose a representation which satisfies

$$(\mathbf{U}_i^L)^T \mathbf{U}_i^L = I_{r_i}, \tag{2.14}$$

which implies $(\mathbf{U}^{\leq i})^T \mathbf{U}^{\leq i} = 0$. If $\mathbf{U} \in \mathcal{U}$, then such a representation is unique up to scaling by orthogonal matrices. We call a $\mathrm{TT}_{\leq \mathbf{r}}$ representation satisfying (2.14) *left orthonormal*.

The successive SVD algorithm of Oseledets [30], which decomposes a tensor of TT rank $\mathbf{r}$ into the TT format produces a certain normal form which is also used in the quantum physics community [39, 38]. This representation can be characterized by the property that the columns of $\mathbf{U}^{\leq i}$ consist of left singular vectors of the ith unfolding $\mathbf{X}^{(i)} = \mathbf{U}^{\leq i} \mathbf{U}^{\geq i+1}$, that is,

$$\mathbf{U}^{\geq i+1} = \Sigma_i V_i,$$

such that

$$\mathbf{X}^{(i)} = \mathbf{U}^{\leq i} \Sigma_i V_i \tag{2.15}$$

is an SVD with singular values arranged in *descending* order. We call such a representation *left normal*. In the case that all unfoldings $\mathbf{X}^{(i)}$ have rank r_i and the nonzero singular vectors are mutually distinct (have multiplicity one), the left normal form of a $\mathrm{TT}_{\leq \mathbf{r}}$ tensor is unique up to scaling by orthogonal diagonal matrices (distributions of signs). We will call tensors with this property *non-degenerate*.

As shown in [22], a left normal representation satisifies the two gauge conditions

$$(\mathbf{U}_i^L)^T \mathbf{U}_i^L = I_{r_i}, \quad \mathbf{U}_{i+1}^L \Sigma_{i+1} (\mathbf{U}_{i+1}^L)^T = \Sigma_i. \tag{2.16}$$

On the other hand, if $\mathbf{X}$ is non-degenerate, a solution of (2.16) also is unique up to scaling by orthogonal diagonal matrices, which shows that (2.16) is equivalent to (2.15).

We can define a so called *left normalization operator*

$$R_L\colon \mathcal{U} \to \mathcal{U},$$

which scales a $\mathrm{TT}_{\leq \mathbf{r}}$ representation $\mathbf{U}$ into the left normal form, using for instance the successive SVD algorithm. The properties of this operator then depend on the concrete choice of the SVD solver, which might not be known in practice when a black box SVD is applied.

PROPOSITION 2.8. *Assume that*

(i) *R_L is smooth in a neighborhood of its fixed points,*

(ii) *left normal representations are fixed points of R_L.*

Then R_L is a local normalization operator in the neighborhood of left normal representations $\mathbf{U} \in \mathcal{U}$, if $\mathbf{X} = \tau(\mathbf{U})$ is non-degenerate.

Proof. Since, in the case of non-degeneracy, the left normal TT decomposition is unique up to scaling by orthogonal diagonal matrices, which form a discrete group, the fixed points of $R|_{\mathcal{M}_\mathbf{U}}$ are isolated points of its range. The assertion then follows from property (i) above. ∎

Similar statements can be made for the analogously defined *right orthonormal* and *right normal* representations.

3. The local convergence of the alternating linear scheme.

In this section we present an abstract alternating optimization scheme which involves rescalings during the iteration process. First, local convergence of a particular fixed point version of this algorithm (using local normalization) can be proved under the assumption that the Hessian of the loss function j to be defined below (see (3.2)) is positive definite at the solution modulo the null space caused by the nonuniquenes(2.11) of the $\mathrm{TT}_{\leq \mathbf{r}}$ representation. Since we will see that convergence in the sense of orbits is not affected when the iterates are moved on their orbits during the iteration, we then deduce a convergence result for the general version of the algorithm, and, as a special case, for the alternating linear scheme introduced in [20]. We also give a short treatise on ALS for nonconvex functionals and on its application to reconstruction problems. The discussion of the positive definiteness assumption is postponed to the next section.

3.1. Alternating optimization and scaling.

Let

$$J\colon \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{R}$$

be a strictly convex C^2-functional to be minimized. Due to the high-dimensionality of the domain $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the task is restricted to the set of $\mathrm{TT}_{\leq \mathbf{r}}$ decomposable tensors:

$$J(\mathbf{X}) = \min, \quad \mathbf{X} \in \mathcal{T}_{\leq \mathbf{r}}. \tag{3.1}$$

In the parametrized version, let

$$j = J \circ \tau\colon \overline{\mathcal{U}} \to \mathbb{R}\colon \mathbf{U} \mapsto J(\tau(\mathbf{U})), \tag{3.2}$$

then we seek for a solution of

$$j(\mathbf{U}) = j(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d) = \min. \tag{3.3}$$

A prominent example is the problem of finding a best approximation of a tensor $\mathbf{Y}$ by a $\mathrm{TT}_{\leq \mathbf{r}}$ decomposable tensor in the Frobenius (Euclidian) norm, such that the task becomes $\|\mathbf{Y} - \tau(\mathbf{U})\|_F^2 = \min$.

Obviously, if $\mathbf{X}^* = \tau(\mathbf{U}^*)$ is a local minimum (3.1) (local in $\mathcal{T}_{\leq \mathbf{r}}$), then $\mathbf{U}^*$ has to be a local minimum of j in $\overline{\mathcal{U}}$. The converse is at least true if $\mathbf{U}^* \in \mathcal{U}$. Namely, since submersions are open maps, it follows from Proposition 2.4 that $\mathbf{X}^*$ is a local minimum of J in $\mathcal{T}_{\mathbf{r}}$. Since this set is open in $\mathcal{T}_{\leq \mathbf{r}}$ (for $\mathcal{T}_{\leq \tilde{\mathbf{r}}}$ is closed for every $\tilde{\mathbf{r}} \leq \mathbf{r}$), $\mathbf{X}^*$ is also a local minimum in $\mathcal{T}_{\leq \mathbf{r}}$.

Our results will only hold for (local) solutions $\mathbf{U}^* \in \mathcal{U}$ of (3.3), that is, for (local) solutions $\mathbf{X}^* \in \mathcal{T}_{\mathbf{r}}$. We assume such solutions to exist. In particular, $\mathcal{U}$ shall not be empty.

Note that j is constant on orbits $\mathcal{M}_{\mathbf{U}} \subseteq \mathcal{U}$, which has the following consequence.

PROPOSITION 3.1. *Let $\mathbf{U}^* \in \mathcal{U}$ be a local minimizer of j. Then every $\hat{\mathbf{U}}^* \in \mathcal{M}_{\mathbf{U}^*}$ is also a local minimizer.*

Proof. Assume this is not true for some $\hat{\mathbf{U}}^*$. Then there exists a sequence $(\hat{\mathbf{V}})_n$ converging to $\hat{\mathbf{U}}^* = \theta_{\mathbf{U}^*}(\mathsf{A})$ such that $j(\hat{\mathbf{V}}_n) > j(\hat{\mathbf{U}}^*)$ for all n. By the continuity of θ, the sequence $\mathbf{V}_n = \theta_{\hat{\mathbf{V}}_n}(\mathsf{A}^{-1})$ converges to $\mathbf{U}^*$ and satisifies $j(\mathbf{V}_n) = j(\hat{\mathbf{V}}_n) > j(\hat{\mathbf{U}}^*) = j(\mathbf{U}^*)$. This contradicts the local optimality of $\mathbf{U}^*$. □

In light of this observation, we will call $\mathcal{M}_{\mathbf{U}^*}$ a *local solution orbit* of (3.3). The alternating optimization approach to find a representant of $\mathcal{M}_{\mathbf{U}^*}$ consists in iterating the cycle

$$\begin{aligned}
\mathbf{U}_1^{(n+1)} &= \operatorname*{argmin}_{\mathbf{V}_1} \ j(\mathbf{V}_1, \mathbf{U}_2^{(n)}, \dots, \mathbf{U}_d^{(n)}), \\
\mathbf{U}_2^{(n+1)} &= \operatorname*{argmin}_{\mathbf{V}_2} \ j(\mathbf{U}_1^{(n+1)}, \mathbf{V}_2, \mathbf{U}_3^{(n)} \dots, \dots, \mathbf{U}_d^{(n)}), \\
&\ \ \vdots \\
\mathbf{U}_d^{(n+1)} &= \operatorname*{argmin}_{\mathbf{V}_d} \ j(\mathbf{U}_1^{(n+1)}, \dots, \mathbf{U}_{d-1}^{(n+1)}, \mathbf{V}_d).
\end{aligned} \tag{3.4}$$

In the case of best approximation (in the Frobenius norm) this method is called *alternating least squares* (ALS), since every micro-step in (3.4) is a linear least squares problem then. More generally, this type of method (more precisely its gradient version) is referred to as nonlinear block Gauss-Seidel, SOR or relaxation method [3, 29, 33]. In fact, it locally equals the linear Gauss-Seidel iteration applied to the Hessian $j''(\mathbf{U}^*)$ up to second order terms.

For computational and also, as we will see later, for analytical reasons, it can be useful to rescale the iterates during the process. Given sequences of scaling operators $R_1^{(n)}, R_2^{(n)}, \dots, R_d^{(n)}$ of the form

$$R_i^{(n)}(\mathbf{U}) = \theta_{\mathbf{U}}(\mathsf{A}_i^{(n)}(\mathbf{U})),$$

the algorithm we now analyze is displayed as Algorithm 1. The ALS used in [20] is a symmetric extension of this and will be discussed in Section 3.3.

To study Algorithm 1 in a precise mathematical framework, we shall first convince ourselves that every ALS micro-step of (3.4) has a unique solution if the current micro-iterate is in $\mathcal{U}$, that is, satisfies the full rank condition (2.9).

PROPOSITION 3.2. *Let $\mathbf{U} \in \mathcal{U}$. Then for $i = 1, 2, \dots, d$ the linear maps*

$$P_{\mathbf{U},i} \colon \mathbb{R}^{r_{i-1} \times n_i \times r_i} \to \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \colon \mathbf{V}_i \mapsto \tau(\mathbf{U}_1, \dots, \mathbf{U}_{i-1}, \mathbf{V}_i, \mathbf{U}_{i+1}, \dots, \mathbf{U}_d)$$

are injective and the operators

$$S_i \colon \mathcal{U} \to \overline{\mathcal{U}} \colon \mathbf{U} \mapsto (\mathbf{U}_1, \dots, \mathbf{U}_{i-1}, \operatorname*{argmin}_{\mathbf{V}_i} \ J(P_{\mathbf{U},i}(\mathbf{V}_i)), \mathbf{U}_{i+1}, \dots, \mathbf{U}_d)$$

Algorithm 1 TT-ALS with rescaling

Require: $\mathbf{U}^{(0)}$
  for $n = 0, 1, 2, \dots$ **do**
    for $i = 1, 2, \dots, d$ **do**
      1. Perform one ALS micro-step:
        $\widetilde{\mathbf{U}}_i^{(n+1)} = \underset{\mathbf{V}_i}{\operatorname{argmin}}\ j(\mathbf{U}_1^{(n+1)}, \dots, \mathbf{U}_{i-1}^{(n+1)}, \mathbf{V}_i, \mathbf{U}_{i+1}^{(n)}, \dots, \mathbf{U}_d^{(n)}),$
        $S_i(\mathbf{U}^{(n)}) = (\mathbf{U}_1^{(n+1)}, \dots, \widetilde{\mathbf{U}}_i^{(n+1)}, \mathbf{U}_{i+1}^{(n)}, \dots, \mathbf{U}_d^{(n)}).$
      2. Rescale TT representation:
        $\mathbf{U}^{(n+1)} = R_i^{(n)}(S_i(\mathbf{U}^{(n)})).$
    end for
  end for

are well-defined and smooth on $\mathcal{U}$.

Proof. By (2.4) and (2.7), the i-th canonical unfolding of $\mathbf{X} = P_{\mathbf{U},i}(\mathbf{V}_i)$ reads

$$\mathbf{X}^{(i)} = (I_{n_i} \otimes \mathbf{U}^{\leq i-1})\mathbf{V}_i^L \mathbf{U}^{\geq i+1}.$$

Since the matrices $(I_{n_i} \otimes \mathbf{U}^{\leq i-1})$ and $\mathbf{U}^{\geq i+1}$ are of full column and row rank, respectively, and since unfoldings are isomorphisms, the injectivity of $P_{\mathbf{U},i}$ follows. Consequently, the assumed strict convexity and smoothness of J implies that the mapping $\mathbf{V}_i \mapsto J(P_{\mathbf{U},i}(\mathbf{V}_i))$ is strictly convex and hence possesses a unique global minimizer which depends smoothly on $\mathbf{U}$. Hence the S_i are well-defined and smooth. ∎

Assuming for the moment that the iteration process does not leave $\mathcal{U}$ at any stage, we may write Algorithm 1 as

$$\mathbf{U}^{(n+1)} = (R_d^{(n)} \circ S_d \circ R_{d-1}^{(n)} \circ S_{d-1} \circ \cdots \circ R_1^{(n)} \circ S_1)(\mathbf{U}^{(n)}). \tag{3.5}$$

A crucial observation is that the true object of interest, the sequence of the TT tensors $\tau(\mathbf{U}^{(n)})$, does not depend on the choice of the scaling operations $R_i^{(n)}$. For illustration, let us recall the rank-one case. If a^* is a unique minimizer of $a \mapsto J(a \otimes b \otimes c)$ for fixed b, c, then $\beta^{-1}\gamma^{-1}a^*$ is the unique minimizer of $a \mapsto J(a \otimes \beta b \otimes \gamma c)$. Hence one ends up with equivalent representations $a^* \otimes b \otimes c$ and $\beta^{-1}\gamma^{-1}a^* \otimes \beta b \otimes \gamma c$. Interestingly, this generalizes to TT tensors.

PROPOSITION 3.3. *Let $\mathbf{U} \in \mathcal{U}$ and assume $S_i(\mathbf{U}) \in \mathcal{U}$. Then equivalent TT representations $\hat{\mathbf{U}} \in \mathcal{M}_{\mathbf{U}}$ are mapped to equivalent representations $S_i(\hat{\mathbf{U}}) \in \mathcal{M}_{S_i(\mathbf{U})}$, that is,*

$$S_i(\mathcal{M}_{\mathbf{U}}) \subseteq \mathcal{M}_{S_i(\mathbf{U})}.$$

Proof. The main argument is that the linear operators $P_{\mathbf{U},i}$ and $P_{\hat{\mathbf{U}},i}$ have the same range [20, Lemma 3.2], and that J, by virtue of its strict convexity, has a unique minimizer $\mathbf{X}$ on that space. Hence, denoting by $\mathbf{V}_i$ and $\hat{\mathbf{V}}_i$ the minimizers of $J \circ P_{\mathbf{U},i}$ and $J \circ P_{\hat{\mathbf{U}},i}$, respectively, it holds

$$\tau(S_i(\mathbf{U})) = P_{\mathbf{U},i}(\mathbf{V}_i) = \mathbf{X} = P_{\hat{\mathbf{U}},i}(\hat{\mathbf{V}}_i) = \tau(S_i(\hat{\mathbf{U}})).$$

The assertion follows from Proposition 2.5 in the form of (2.12). ∎

The preceeding proposition identifies Algorithm 1 as an algorithm on the manifold $\mathcal{T}_{\mathbf{r}}$ of tensors of fixed TT rank. The value for the convergence analysis lies in the possibility to restrict the attention to particular choices of scaling strategies which are easy to investigate. Using this idea, we first can prove that $\mathbf{U}^{(n+1)}$ in (3.5) is uniquely defined.

LEMMA 3.4. *Let $\mathbf{U}^* \in \mathcal{U}$ be a local solution of minimization problem (3.3). Then $\mathcal{M}_{\mathbf{U}^*}$ possesses an open neighborhood $\mathcal{O} \subseteq \mathcal{U}$ in which the composed operator*

$$R_d^{(n)} \circ S_d \circ R_{d-1}^{(n)} \circ S_{d-1} \circ \cdots \circ R_1^{(n)} \circ S_1 \colon \mathcal{O} \to \mathcal{O} \tag{3.6}$$

is well-defined for all sequences of scaling operators $(R_i^{(n)})$. Hence, the result of one full step of Algorithm 1 is uniquely defined if the current iterate $\mathbf{U}^{(n)}$ is in $\mathcal{O}$. The next iterate is given by (3.5).

Every $\hat{\mathbf{U}}^* \in \mathcal{M}_{\mathbf{U}^*}$ is a fixed point of every S_i. Consequently, it holds

$$(R_d^{(n)} \circ S_d \circ R_{d-1}^{(n)} \circ S_{d-1} \circ \cdots \circ R_1^{(n)} \circ S_1)(\hat{\mathbf{U}}^*) = (R_d^{(n)} \circ R_{d-1}^{(n)} \circ \cdots \circ R_1^{(n)})(\hat{\mathbf{U}}^*).$$

Proof. We first show the second part. As we have seen in the proof of Proposition 3.2, the $S_i(\mathbf{U}^*)$ are constructed by replacing $\mathbf{U}_i^*$ with the unique global minimum of the strictly convex functional $\mathbf{V}_i \mapsto J \circ P_{\mathbf{U}^*,i}$. Since by assumption $\mathbf{U}_i^*$ has to be a local minimum of this function, it is already the global one. Due to Proposition 3.1, this argument works for all $\hat{\mathbf{U}}^* \in \mathcal{M}_{\mathbf{U}^*}$.

Set $\mathcal{O}_d = \mathcal{U}$. Since $\mathcal{U}$ is open and $\mathbf{U}^*$ is a fixed point of the continuous map S_{d-1}, there exists an open neighborhood $\mathcal{O}_{d-1}$ of $\mathbf{U}^*$ such that $S_{d-1}(\mathcal{O}_{d-1}) \subseteq \mathcal{O}_d$. Hence $S_d \circ S_{d-1}$ is defined on $\mathcal{O}_{d-1}$. By induction, we find a neighborhood $\mathcal{O}_1$ of $\mathbf{U}^*$ such that

$$S_d \circ S_{d-1} \circ \cdots \circ S_1$$

is defined on $\mathcal{O}_1$ and even, by Proposition 3.3, on

$$\mathcal{O} := \theta(\mathcal{O}_1, \mathcal{G}),$$

which is a neighborhood of $\mathcal{M}_{\mathbf{U}^*}$. Proposition 3.3 also proves that operator (3.6) is well-defined on $\mathcal{O}$ for any choice of scaling operators $R_i^{(n)}$. □

3.2. Main assumption and convergence results. Formally, Algorithm 1 is an algorithm in the parameter space $\overline{\mathcal{U}}$ and produces a sequence $(\mathbf{U}^{(n)}) \subseteq \mathcal{U}$ if the starting point is close enough to a local solution $\mathbf{U}^* \in \mathcal{U}$ of the minimization problem (3.3). The standard approach to prove local convergence $\mathbf{U}^{(n)} \to \mathbf{U}^*$ would be a contraction argument. But since we made no assumptions on the scaling operators $R_i^{(n)}$ so far, no point of the solution orbit $\mathcal{M}_{\mathbf{U}^*}$ has to be a fixed point of the iteration at all. On the other hand, we initially started with the minimization problem (3.1) on the set $\mathcal{T}_{\leq \mathbf{r}}$. In practice, we are thus only interested in the convergence $\tau(\mathbf{U}^{(n)}) \to \tau(\mathbf{U}^*)$. According to Proposition 3.3 the latter sequence is independent of the scaling operators $R_i^{(n)}$. Our trick is now to use a scaling strategy that enables us to use fixed-point arguments in the parameter space.

As a particular instance of (3.5), we investigate the iteration

$$\mathbf{U}^{(n+1)} = (R_{\mathbf{U}^*} \circ S)(\mathbf{U}^{(n)}), \tag{3.7}$$

where $R_{\mathbf{U}^*}$ is a local normalization operator as described in Proposition 2.7 and

$$S := S_d \circ S_{d-1} \circ \cdots \circ S_1$$

is the "pure" ALS operator. By Lemma 3.4, $R_{\mathbf{U}^*} \circ S$ is well-defined and smooth in a neighborhood of $\mathbf{U}^*$. Note that every point in $\mathcal{M}_{\mathbf{U}^*}$ is a fixed-point of S, but only $\mathbf{U}^*$ is a fixed point of $R_{\mathbf{U}^*}$, at least locally. Hence $\mathbf{U}^*$ is the only fixed point of $R_{\mathbf{U}^*} \circ S$ in $\mathcal{M}_{\mathbf{U}^*}$ in a neighborhood of itself.

If j would be a quadractic form, then the alternating optimization given by operator S is just the standard block Gauss-Seidel iteration applied to the Hessian matrix $j''(\mathbf{U}^*)$ of j at the minimizer $\mathbf{U}^*$. If j is arbitrary, then this is at least asymptotically true (Lemma 3.6 below). Since j is constant on the orbit $\mathcal{M}_{\mathbf{U}^*}$, the null space of $j''(\mathbf{U}^*)$ at least contains the $(\sum_{i=1}^{d-1} r_i^2)$-dimensional tangent space $T\mathcal{M}_{\mathbf{U}^*}$. Thus, as known from theory, elements in $T\mathcal{M}_{\mathbf{U}^*}$ will not be damped by the Gauss-Seidel method. As we will see later, this drawback is compensated by the local normalization operator $R_{\mathbf{U}^*}$. What we need to assume for our convergence proof is that $j''(\mathbf{U}^*)$ is positive definite in all other directions. Such an assumption is natural and in line with usual results concerning the convergence of the nonlinear SOR, if one takes the scaling indeterminacy in our setting into account.

MAIN ASSUMPTION. *At the local minimizer $\mathbf{U}^*$, the Hessian $j''(\mathbf{U}^*)$ is of full possible rank,*

$$\operatorname{rank} j''(\mathbf{U}^*) = \sum_{i=1}^{d} r_{i-1} n_i r_i - \sum_{i=1}^{d-1} r_i^2, \quad \textit{that is,} \quad \ker j''(\mathbf{U}^*) = T\mathcal{M}_{\mathbf{U}^*}. \tag{MA}$$

We should make clear that the particular choice of $\mathbf{U}^*$ has no qualitative influence. In accordance to our viewpoint of the ALS algorithm as an ,,iteration of orbits", assumption (MA) is in fact an assumption on the whole solution orbit $\mathcal{M}_{\mathbf{U}^*}$.

PROPOSITION 3.5. *Let $\mathcal{M}_{\mathbf{U}^*} \subseteq \mathcal{U}$ be a local solution orbit. If* (MA) *holds for* $\mathbf{U}^* \in \mathcal{M}_{\mathbf{U}^*}$*, then it holds for all $\hat{\mathbf{U}}^* \in \mathcal{M}_{\mathbf{U}^*}$.*

Proof. Let $\hat{\mathbf{U}}^* = \theta_{\mathsf{A}^{-1}}(\mathbf{U}^*)$ for some $\mathsf{A} \in \mathcal{G}$.[2] Then for all $\mathbf{H} \in \overline{\mathcal{U}}$ we have

$$j(\hat{\mathbf{U}}^* + \mathbf{H}) = j(\theta_{\mathsf{A}^{-1}}(\mathbf{U}^* + \theta_{\mathsf{A}}(\mathbf{H})) = j(\mathbf{U}^* + \theta_{\mathsf{A}}(\mathbf{H})),$$

where we used that j is constant on orbits. Taking into account that $j(\hat{\mathbf{U}}^*) = j(\mathbf{U}^*)$ and $j'(\hat{\mathbf{U}}^*) = j'(\mathbf{U}^*) = 0$, the above relation implies

$$j''(\hat{\mathbf{U}}^*)[\mathbf{H}, \mathbf{H}] = j''(\mathbf{U}^*)[\theta_{\mathsf{A}}(\mathbf{H}), \theta_{\mathsf{A}}(\mathbf{H})].$$

Now note that θ_{A} is an isomorphism from $\overline{\mathcal{U}}$ onto itself (with inverse $\theta_{\mathsf{A}^{-1}}$). Hence $j''(\hat{\mathbf{U}}^*)$ and $j''(\mathbf{U}^*)$ are of same rank. □

We now present the convergence analysis of Algorithm 1 under assumption (MA). The validity of this assumption will be discussed in the next section.

LEMMA 3.6. *Let $\mathbf{U}^* \in \mathcal{U}$ be a local minimum of* (3.3) *for which main assumption* (MA) *holds. Partition $j''(\mathbf{U}^*) = L + D + U$ according to the block variables $\mathbf{U}_i$ into lower block triangular, block diagonal and upper block triangular matrices L, D and U, respectively. Then D is positive definite and it holds*

$$S'(\mathbf{U}^*) = -(D + L)^{-1} U.$$

[2] In this proof we use the shorthand θ_{A} for the map $\mathbf{U} \mapsto \theta(\mathbf{U}, \mathsf{A})$.

Proof. Clearly, $j''(\mathbf{U}^*)$ is positive semidefinite. Using the full rank properties (2.9) of $\mathbf{U}^*$, it is easily verified from (2.13), that vectors of the form $\mathbf{H} = (0,\dots,0,\mathbf{H}_i,0,\dots,0)$ do not belong to $T\mathcal{M}_{\mathbf{U}^*}$ unless $\mathbf{H}_i = 0$. It follows from (MA) that D is positive definite, and hence $D+L$ invertible. A nice calculation of $S'(\mathbf{U}^*)$ which leads to the asserted formula can be found in [3, Lemma 2]. □

In the following, we denote by

$$|\mathbf{H}|_E = (j''(\mathbf{U}^*)[\mathbf{H},\mathbf{H}])^{1/2}$$

the energy seminorm of $j''(\mathbf{U}^*)$. Let further $\|\cdot\|$ denote a norm on $\overline{\mathcal{U}}$. If main assumption (MA) holds, then

$$\|\mathbf{H}\|_* = \|(I - R'_{\mathbf{U}^*}(\mathbf{U}^*))\mathbf{H}\| + |\mathbf{H}|_E$$

also defines a norm, for if $|\mathbf{H}|_E = 0$, then $\mathbf{H} \in \ker j''(\mathbf{U}^*) = T\mathcal{M}_{\mathbf{U}^*}$, and since $R_{\mathbf{U}^*}$ is constant on a neighborhood of $\mathbf{U}^*$ in $T\mathcal{M}_{\mathbf{U}^*}$, $\|(I - R'_{\mathbf{U}^*}(\mathbf{U}^*))\mathbf{H}\| = \|\mathbf{H}\| = 0$ implies $\mathbf{H} = 0$.

THEOREM 3.7. *Let $\mathbf{U}^* \in \mathcal{U}$ be a local minimum of* (3.3) *for which main assumption* (MA) *holds. Then the iteration* (3.7) *is locally q-linearly convergent to $\mathbf{U}^*$ in the norm $\|\cdot\|_*$ at an asymptotic rate of at least $q = |S'(\mathbf{U}^*)|_E < 1$. That is, for every $\epsilon > 0$ there exists a neighborhood $\mathcal{O}^*_\epsilon \subseteq \mathcal{U}$ of $\mathbf{U}^*$ such that*

$$\|\mathbf{U}^* - \mathbf{U}^{(n+1)}\|_* \le (q+\epsilon)\|\mathbf{U}^* - \mathbf{U}^{(n)}\|_*$$

for all n, if $\mathbf{U}^{(0)}$ is in $\mathcal{O}^*_\epsilon$.

Proof. Since $R^2_{\mathbf{U}^*} = R_{\mathbf{U}^*}$ in a neighborhood of $\mathbf{U}^*$, it holds

$$(I - R'_{\mathbf{U}^*}(\mathbf{U}^*))R'_{\mathbf{U}^*}(\mathbf{U}^*) = 0.$$

Further, for sufficiently small $\mathbf{H} \in \overline{\mathcal{U}}$, we have $j(\mathbf{U}^* + \mathbf{H}) = j(R_{\mathbf{U}^*}(\mathbf{U}^* + \mathbf{H}))$, which together with $j'(\mathbf{U}^*) = j'(R_{\mathbf{U}^*}(\mathbf{U}^*)) = 0$ shows

$$j''(\mathbf{U}^*)[\mathbf{H},\mathbf{H}] = j''(\mathbf{U}^*)[R'_{\mathbf{U}^*}(\mathbf{U}^*)\mathbf{H}, R'_{\mathbf{U}^*}(\mathbf{U}^*)\mathbf{H}],$$

that is,

$$|R'_{\mathbf{U}^*}(\mathbf{U}^*)|_E = 1.$$

Consequently, we can estimate

$$\|(R_{\mathbf{U}^*} \circ S)'(\mathbf{U}^*)\mathbf{H}\|_* = \|R'_{\mathbf{U}^*}(\mathbf{U}^*)S'(\mathbf{U}^*)\mathbf{H}\|_* \le |S'(\mathbf{U}^*)|_E|\mathbf{H}|_E \le |S'(\mathbf{U}^*)|_E\|\mathbf{H}\|_*.$$

By the previous Lemma 3.6, $S'(\mathbf{U}^*)$ is the error iteration matrix of the linear block Gauss-Seidel method applied to $j''(\mathbf{U}^*)$ and is known to be a contraction in the energy seminorm, that is, $|S'(\mathbf{U}^*)|_E < 1$, see [23, Eq. (9)] or [28, Theorem 3.2]. The theorem is thus a consequence of the contraction principle. □

Generic estimates for $q = |S'(\mathbf{U}^*)|_E$ are given in [41].

We can now give two equivalent convergence statements for arbitrary scalings.

COROLLARY 3.8. *Let $\mathcal{M}_{\mathbf{U}^*} \subseteq \mathcal{U}$ be a local solution orbit of* (3.3) *for which the main assumption* (MA) *holds.*[3] *Then the sequence $(\mathcal{M}_{\mathbf{U}^{(n)}})$ of orbits produced by*

[3]This phrase makes sense by Proposition 3.5.

Algorithm 1 is locally r-linear convergent to $\mathcal{M}_{\mathbf{U}^*}$ at an asymptotic rate of at least $q = |S'(\mathbf{U}^*)|_E < 1$. That is, for every $\epsilon > 0$ there exists a neighborhood $\mathcal{O}_\epsilon \subseteq \mathcal{U}$ of $\mathcal{M}_{\mathbf{U}^*}$ such that

$$\limsup_{n\to\infty} \sqrt[n]{\operatorname{dist}(\mathcal{M}_{\mathbf{U}^*}, \mathcal{M}_{\mathbf{U}^{(n)}})} \le q + \epsilon,$$

if $\mathbf{U}^{(0)}$ is in $\mathcal{O}_\epsilon$. (The distance is measured in any norm $\|\cdot\|$ on $\overline{\mathcal{U}}$.)

Proof. Since it is independent from the choice of norm, the assertion follows immediately from Theorem 3.7 and Proposition 3.3 by choosing a suitable neighborhood $\mathcal{O}_\epsilon^*$ of any $\mathbf{U}^* \in \mathcal{M}_{\mathbf{U}^*}$ and setting $\mathcal{O}_\epsilon := \theta(\mathcal{O}_\epsilon^*, \mathcal{G})$. □

COROLLARY 3.9. *Let $\mathbf{X}^* = \tau(\mathbf{U}^*)$ be a local solution of* (3.1) *with full TT rank* $\mathbf{r}$*, that is, $\mathbf{U}^* \in \mathcal{U}$. Assume main assumption* (MA) *holds for $\mathbf{X}^*$.*[3] *Then the sequence $(\mathbf{X}^{(n)})$ of TT tensors produced by Algorithm 1 is locally r-linear convergent to $\mathbf{X}^*$ at an asymptotic rate of at least $q = |S'(\mathbf{U}^*)|_E < 1$. That is, for every $\epsilon > 0$ there exists a neighborhood $\mathcal{X}_\epsilon \subseteq \mathcal{T}_{\le \mathbf{r}}$ of $\mathbf{X}^*$ such that*

$$\limsup_{n\to\infty} \sqrt[n]{\|\mathbf{X}^* - \tau(\mathbf{U}^{(n)})\|} \le q + \epsilon,$$

if $\mathbf{X}^{(0)} = \tau(\mathbf{U}^{(0)})$ is in $\mathcal{X}_\epsilon$ (with $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$).

Proof. This follows from Theorem 3.7 and Proposition 3.3 by choosing a suitable neighborhood $\mathcal{O}_\epsilon^*$ of any $\mathbf{U}^* \in \mathcal{M}_{\mathbf{U}^*}$ and a constant $C_\epsilon > 0$ such that

$$\|\tau(\mathbf{U}^*) - \tau(\mathbf{U})\| \le C_\epsilon \|\mathbf{U}^* - \mathbf{U}\|_*$$

for all $\mathbf{U} \in \mathcal{O}_\epsilon^*$. Then, by Proposition 2.4, $\mathcal{X}_\epsilon := \tau(\mathcal{O}_\epsilon^*)$ is a neighborhood of $\mathbf{X}^*$ in $\mathcal{T}_\mathbf{r}$ for which the assertion holds. Since $\mathcal{T}_\mathbf{r}$ is open in $\mathcal{T}_{\le\mathbf{r}}$, $\mathcal{X}_\epsilon$ is also a neighborhood in $\mathcal{T}_{\le\mathbf{r}}$. □

3.3. Convergence of ALS from [20]. The alternating linear scheme (also ALS) as introduced in [20] is a symmetric extension of Algorithm 1. The algorithm is described in Algorithm 2.

As one can see, the components $\mathbf{U}_i$ are first optimized from left to right and then backwards. After each micro-step the representation is changed according to a QR decomposition of the unfolding. Shifting of R and R^T was noted for convenience of the reader and has not to be implemented, since the affected component will be updated in the next micro-step anyway. After the first inner loop, the representation is left orthonormal, and after a full loop right orthonormal. The details of the algorithm are described in [20].

The convergence analysis of Algorithm 2 proceeds by the same lines as for Algorithm 1. It can be shown that it is well defined in a neighborhood of a full rank solution orbit $\mathcal{M}_{\mathbf{U}^*} \subseteq \mathcal{U}$. More precisely, let $S_L = S$ denote the left to right ALS operator from above, and $S_R = S_d \circ S_{d-1} \circ \cdots \circ S_1$ the right to left operator. Then Algorithm 2 produces the same orbits as

$$R_{\mathbf{U}^*} \circ S_R \circ S_L$$

(where formally $\mathbf{U}_d$ and $\mathbf{U}_1$ are updated twice in a row). The local behaviour is governed by the spectral properties of the matrix $(S_R \circ S_L)'(\mathbf{U}^*) = S_R'(\mathbf{U}^*) S_L'(\mathbf{U}^*)$, which is the error iteration matrix of the symmetric linear block Gauss-Seidel iteration. In the same way as $S_L'(\mathbf{U}^*)$, matrix $S_R'(\mathbf{U}^*)$ is a contraction in the energy seminorm [28, Theorem 3.2], and hence

$$|(S_R \circ S_L)'(\mathbf{U}^*)|_E \le |S_R'(\mathbf{U}^*)|_E \cdot |S_L'(\mathbf{U}^*)|_E < 1.$$

Algorithm 2 Alternating linear scheme[4]

Require: $\mathbf{U}^{(0)}$

for $n = 0, 1, 2, \dots$ **do**

  for $i = 1, 2, \dots, d-1$ **do**

    1. Perform one ALS micro-step:
       $\widetilde{\mathbf{U}}_i^{(n+1)} = \underset{\mathbf{V}_i}{\operatorname{argmin}}\ j(\widetilde{\mathbf{U}}_1^{(n+1)}, \dots, \widetilde{\mathbf{U}}_{i-1}^{(n+1)}, \mathbf{V}_i, \mathbf{U}_{i+1}^{(n)}, \dots, \mathbf{U}_d^{(n)})$.
    2. Make a (tall) QR decomposition:
       $(\widetilde{\mathbf{U}}_i^{(n+1)})^L = QR$.
    3. Keep Q, shift R to the right:
       $\widetilde{\mathbf{U}}_i^{(n+1)} \leftarrow Q^{-L}, \qquad \mathbf{U}_{i+1}^{(n)} \leftarrow R^{-1}\mathbf{U}_{i+1}^{(n)}$.

  end for

  for $i = d, d-1, \dots, 2$ **do**

    1. Perform one ALS micro-step:
       $\mathbf{U}_i^{(n+1)} = \underset{\mathbf{V}_i}{\operatorname{argmin}}\ j(\widetilde{\mathbf{U}}_1^{(n+1)}, \dots, \widetilde{\mathbf{U}}_{i-1}^{(n+1)}, \mathbf{V}_i, \mathbf{U}_{i+1}^{(n+1)}, \dots, \mathbf{U}_d^{(n+1)})$.
    2. Make a (tall) QR decomposition:
       $((\mathbf{U}_i^{(n+1)})^R)^T = QR$.
    3. Keep Q^T, shift R^T to the left:
       $\mathbf{U}_i^{(n+1)} \leftarrow (Q^T)^{-R}, \qquad \widetilde{\mathbf{U}}_{i-1}^{(n)} \leftarrow \widetilde{\mathbf{U}}_{i-1}^{(n)} R^{-T}$.

  end for

end for

We conclude the following.

THEOREM 3.10. *The convergence results of Theorem 3.7 and Corollaries 3.8, 3.9 hold for the ALS Algorithm 2 from [20], with the convergence rate replaced by* $q = |(S_R \circ S_L)'(\mathbf{U}^*)|_E$.

3.4. Decomposition of tensors with known rank. If in the case of least squares approximation, the rank of the approximand $\mathbf{Y}$ matches that of the used TT manifold, the ALS from Algorithms 1 and 2 usually returns a TT decomposition of $\mathbf{Y}$ after one run over all components, see [20] for numerics. As a generalization of [20, Lemma 4.2], we prove that this holds independent of the scaling and of orthogonality of the components.

PROPOSITION 3.11. *Assume* $\mathbf{Y} = \tau(\mathbf{V}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *has TT-rank* $\mathbf{r}$. *Then Algorithm 1 applied to*

$$j \colon \overline{\mathcal{U}} \to \mathbb{R} \colon \mathbf{U} \mapsto \frac{1}{2}\|\mathbf{Y} - \tau(\mathbf{U})\|_F^2$$

finishes with a TT decomposition $\mathbf{Y} = \tau(\mathbf{U}^{(1)})$ *after one loop over* $i = 1, 2, \dots, d$, *if the starting point* $\mathbf{U}^{(0)}$ *is in*

$$\mathcal{V} = \{\mathbf{U} \in \mathcal{U} \mid \operatorname{rank}(\mathbf{V}_{i+1}^R (\mathbf{U}_{i+1}^R)^T) = r_i \text{ for } i = 1, 2, \dots, d-1\}.$$

This set is open and dense in $\overline{\mathcal{U}}$. *The complement* $\overline{\mathcal{U}} \setminus \mathcal{V}$ *has measure zero.*

Proof. Let $\mathbf{Y} = \tau(\mathbf{V})$ be a left orthonormal $\mathrm{TT}_{\leq \mathbf{r}}$ decomposition (see (2.14)). We set $(\mathbf{U}^{(1)})^{\leq 0} = \mathbf{V}^{\leq 0} = 1$ and assume $(\mathbf{U}^{(1)})^{\leq i-1} = \mathbf{V}^{\leq i-1}$ for some $1 \leq i \leq d$.

[4]See footnote 1 on page 7 for notation.

Using (2.4) and (2.7), it holds for the i-th micro-step

$$(\mathbf{U}_i^{(1)})^L = \operatorname*{argmin}_{Z\in\mathbb{R}^{r_{i-1}n_i\times r_i}} \|(I_{n_i}\otimes\mathbf{V}^{\leq i-1})[\mathbf{V}_i^L\mathbf{V}^{\geq i+1} - Z\cdot(\mathbf{U}^{(0)})^{\geq i+1}]\|_F$$
$$= \operatorname*{argmin}_{Z\in\mathbb{R}^{r_{i-1}n_i\times r_i}} \|\mathbf{V}_i^L\mathbf{V}^{\geq i+1} - Z\cdot(\mathbf{U}^{(0)})^{\geq i+1}\|_F,$$

where we used that $(I_{n_i}\otimes\mathbf{V}^{\leq i-1})$ has orthonormal columns (see (2.7)). Hence,

$$(\mathbf{U}_i^{(1)})^L = \mathbf{V}_i^L A$$

with[5] $A = \mathbf{V}^{\geq i+1}((\mathbf{U}^{(0)})^{\geq i+1})^\dagger$. It is easy to deduce from (2.8), that, by the choice of $\mathbf{U}^{(0)}$, matrix $\mathbf{V}^{\geq i+1}((\mathbf{U}^{(0)})^{\geq i+1})^T$ and with that matrix A are invertible. In consideration of Proposition 3.3 we can assume $A = I_{r_i}$ (for $i = d$ this necessarily holds) and proceed by induction.

For completeness we prove the assertions on $\mathcal{V}$. While it is obvious that $\mathcal{V}$ is open, we could not find a reference for $\overline{\mathcal{U}}\setminus\mathcal{V}$ having measure zero. We therefore deduce it from the following lemma by choosing, for each i, $E = \ker\mathbf{V}_{i+1}^R$ and $s = r = r_i$ therein. ∎

LEMMA 3.12. *Let $1\leq s\leq r\leq n$ and E be an $(n-r)$-dimensional subspace of $\mathbb{R}^n$. Consider the set $\Omega_s = \{A\in\mathbb{R}^{n\times s}\mid \operatorname{rank}A = s,\ \operatorname{ran}A\cap E = \{0\}\}$. Then the complement Ω_s^c has (Borel) measure zero in $\mathbb{R}^{n\times s}$.*

Proof. We proceed by induction over s. For $s = 1$ the assertion is clear. Assume it holds for some $s < r$. Then, $\Omega_s^c\times\mathbb{R}^n$, interpreted as subset of $\mathbb{R}^{n\times(s+1)}$, is a null set by Fubini's theorem. It hence suffices to show that $\Omega_{s+1}^c\cap(\Omega_s\times\mathbb{R}^n)$ has (Borel) measure zero in $\mathbb{R}^{n\times(s+1)}$. But this follows again from Fubini's theorem, since for each $A_s\in\Omega_s$ the cut

$$\{a\in\mathbb{R}^n\mid \begin{bmatrix}A_s & a\end{bmatrix}\in\Omega_{s+1}^c\} = \{a\in\mathbb{R}^n\mid a\in\operatorname{ran}A_s\oplus E\}$$

is a (Borel) null set in $\mathbb{R}^n$ (because $\dim(\operatorname{ran}A_s\oplus E) < n$). ∎

3.5. Non-convex functionals. The strict convexity of the functional J was needed in Proposition 3.2 to guarantee the unique solvability of the optimization steps (3.4). On the other hand, we have seen in the proof of Lemma 3.6, that the main assumption (MA) implies the positive definiteness of the block diagonal of $j''(\mathbf{U}^*)$. This is then true for all $\mathbf{U}$ in a suitable neighborhood of $\mathcal{M}_{\mathbf{U}^*}$. Hence, if $j = J\circ\tau$ would be a quadratic form, the unique solvability of the micro least squares problems could be locally guaranteed by (MA). This observation opens the door to an extension of Algorithm 1 to nonconvex functionals, in which at each micro step of (3.4), the functional

$$\iota_{n,i}(\mathbf{V}_i) = j_i'(\mathbf{U}^{n,i})\mathbf{V}_i + \frac{1}{2}j_i''(\mathbf{U}^{n,i})[\mathbf{V}_i,\mathbf{V}_i]$$

is minimized instead of j, where $\mathbf{U}^{n,i} = (\mathbf{U}_1^{(n+1)},\dots,\mathbf{U}_{i-1}^{(n+1)},\mathbf{U}_i^{(n)},\dots,\mathbf{U}_d^{(n)})$ denotes the current micro iterate, j_i the i-th partial derivative of j and j_i'' the i-th diagonal block of the Hessian. Minimizing $\iota_{n,i}$ is nothing else than taking one Newton step to solve

$$j'(\mathbf{U}_1^{(n+1)},\dots,\mathbf{U}_{i-1}^{(n+1)},\mathbf{V}_i,\mathbf{U}_{i+1}^{(n)},\dots,\mathbf{U}_d^{(n)}) = 0,$$

[5] By $M^\dagger = M^T(MM^T)^{-1}$ we denote the (right) pseudo inverse of a matrix M with full row rank.

for $\mathbf{V}_i$, starting from $\mathbf{U}^{n,i}$. This kind of procedure is called approximate nonlinear relaxation in [33] and Gauss-Seidel-Newton method in [29].

By the above considerations, a corresponding iteration operator $\hat{S}$ can be defined in a neighborhood of $\mathcal{M}_{\mathbf{U}^*}$, given of course that assumption (MA) is satisfied. It is then again true that $\hat{S}'(\mathbf{U}^*)$ is the error iteration matrix of the linear block Gauss-Seidel method for $j''(\mathbf{U}^*)$, see the proof of [29, Theorem 10.3.3]. Note that the claim of Proposition 3.3 is not true anymore. If we hence restrict ourselves to rescaling iterates only after one complete loop by a local normalization operator $R_{\mathbf{U}^*}$, that is, consider the iteration

$$\mathbf{U}^{(n+1)} = (R_{\mathbf{U}^*} \circ \hat{S})(\mathbf{U}^{(n)}),$$

then we can immediately conclude that the iterates will converge linearly to $\mathbf{U}^*$, provided that the starting point is close enough to $\mathbf{U}^*$ and assumption (MA) holds.

For the practical aspects concerning the choice of local normalization operators we refer to Section 2.4 again.

4. Estimates for the Hessian and discussion of the main assumption.

The main assumption (MA) entering the proof of convergence in Theorem 3.7, is that the Hessian $j''(\mathbf{U}^*)$ is of full possible rank $\dim \overline{\mathcal{U}} - \sum_{i=1}^{d-1} r_i^2$, meaning that $j''(\mathbf{U}^*)$ is positive definite on any complementary subspace of $T\mathcal{M}_{\mathbf{U}^*}$. The problem in showing that (MA) holds is that the convexity of the functional J is not necessarily inherited by the functional j defined on the parameter space, even if the redundancy on $T\mathcal{M}_{\mathbf{U}^*}$ is factorized out. The properties of j rather depend both on the functional under consideration and on the properties of the manifold $\mathcal{T}_{\leq \mathbf{r}}$ used; therefore, condition (MA) is usually not trivial to verify. For the case of approximation in the Tucker format, necessary conditions have recently been given in [8]. For the canonical format see [36]. We will give two sufficient conditions for the TT-case in this section: one simple, more generic result for convex functionals, and one more elaborate result for the concrete example of least-squares approximation problems, where a condition on the singular values of the tensor to be approximated guarantees that (MA) holds. To verify (MA), we will in both cases use the following equivalent criterion, based on the idea of the introduction of gauge conditions for the parameter space already used in [19]. Recall the left unfoldings defined in (2.6).

Criterion for (MA). *For $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d) \in \mathcal{U}$, define*

$$W_{\mathbf{U}} := \{\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_d) \in \overline{\mathcal{U}} \mid (\mathbf{W}_i^L)^T \mathbf{U}_i^L = 0 \text{ for } i = 1, 2, \dots, d-1\}, \quad (4.1)$$

the so-called gauge space at $\mathbf{U}$. This is a subspace of dimension $\dim \overline{\mathcal{U}} - \sum_{i=1}^{d-1} r_i^2$ of the component space $\overline{\mathcal{U}}$. Therefore, if

$$j''(\mathbf{U}^*) \text{ is positive definite on } W_{\mathbf{U}^*}, \quad (4.2)$$

then (MA) *is fulfilled. In fact, it holds $\overline{\mathcal{U}} = T\mathcal{M}_{\mathbf{U}^*} \oplus W_{\mathbf{U}^*}$ so that* (4.2) *and* (MA) *are equivalent.*

One could use any other space complementary to $T\mathcal{M}_{\mathbf{U}^*}$, but the above choice of $W_{\mathbf{U}^*}$ will turn out to be useful. Since $j = J \circ \tau$, we have to show that

$$j''(\mathbf{U}^*)[\mathbf{W},\mathbf{W}] = J''(\tau(\mathbf{U}^*))[\tau'(\mathbf{U}^*)\mathbf{W}, \tau'(\mathbf{U}^*)\mathbf{W}] + J'(\tau(\mathbf{U}^*))(\tau''(\mathbf{U}^*)[\mathbf{W},\mathbf{W}]) \quad (4.3)$$

is positive for all nonzero $\mathbf{W} \in W_{\mathbf{U}^*}$. We will assume that the first summand of the right side is always positive. Since $\tau'(\mathbf{U}^*)$ maps $W_{\mathbf{U}^*}$ onto the tangent space of $\mathcal{T}_{\mathbf{r}}$

at $\tau(\mathbf{U}^*)$ (Proposition 2.4), this means that J'' is positive definite on that tangent space – a natural condition and for instance always satisfied for the case of best approximation. However, this condition is not sufficient, one also has to control the second term.

4.1. A generic condition for positive definiteness. The following result shows that at least for local minimizers $\mathbf{U}^* \in \mathcal{U}$ of j, which are sufficiently close to the unique minimizer $\mathbf{Y}$ of J on $\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$, assumption (MA) can be expected to hold. For this, it is necessary that the approximation manifold $\mathcal{T}_{\mathbf{r}}$ is good enough, that is, close enough to $\mathbf{Y}$. As before, we denote by $\|\cdot\|$ arbitrary norms on $\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}$ and $\mathbb{R}^{r_{i-1}\times n_i\times r_i}$.

THEOREM 4.1. *Let $\mathbf{Y}$ be the unique minimum of the strict convex functional J on $(\mathbb{R}^{n_1\times n_2\times\cdots\times n_d}, \|\cdot\|)$. Assume that $J''(Y)$ is positive definite. Then for every $\mathbf{U}^* \in \mathcal{U}$, there exists a $\Delta > 0$ (depending on J and $\mathbf{U}^*$) such that if $\mathbf{Y}$ is close enough to $\tau(\mathbf{U}^*)$,*

$$\|\mathbf{Y} - \tau(\mathbf{U}^*)\| < \Delta,$$

then the following holds:

(i) *Assumption* (MA) *is satisfied for $j''(\mathbf{U}^*)$.*

(ii) *If $\mathbf{U}^*$ solves $j'(\mathbf{U}^*) = 0$, there is an open neighbourhood $\mathcal{O}$ of $\tau(\mathbf{U}^*)$ such that $\tau(\mathbf{U}^*)$ is the unique minimizer of J on $\mathcal{O} \cap \mathcal{T}_{\leq \mathbf{r}}$.*

Proof. Let $\|\cdot\|$ also denote the norm on $\overline{\mathcal{U}}$. First, by assumption, there is an $\alpha > 0$ with

$$\|J''(\mathbf{Y})[\mathbf{W}, \mathbf{W}]\| \geq \alpha\|\mathbf{W}\|^2.$$

Second, on some fixed ball of radius ϵ around X, J' is Lipschitz continuous, so that

$$\|J'(\mathbf{X}) - J'(\tau(\mathbf{U}^*))\| \leq \beta\|\mathbf{X} - \tau(\mathbf{U}^*)\|$$

for some $\beta > 0$ and all $\mathbf{X}$ in that ball. Further, it is not difficult to show that $\tau'(\mathbf{U}^*)$ is injective on $W_{\mathbf{U}^*}$ (cf. Proposition 2.4). Hence there exists $\gamma > 0$ with

$$\|\tau'(\mathbf{U}^*)\mathbf{W}\| \geq \gamma\|\mathbf{W}\|$$

for all $\mathbf{W} \in W_{\mathbf{U}^*}$. Finally, we have

$$\|\tau''(\mathbf{U}^*)[\mathbf{W}, \mathbf{W}]\| \leq \delta\|\mathbf{W}\|^2$$

for some $\delta > 0$. Now if $\|\mathbf{Y} - \tau(\mathbf{U}^*)\| < \Delta = \min\{\frac{\alpha\gamma^2}{\beta\delta}, \epsilon\}$, then we can estimate (4.3) as follows:

$$\begin{aligned} j''(\mathbf{U}^*)[\mathbf{W}, \mathbf{W}] &\geq \alpha\|\tau'(\mathbf{U}^*)\mathbf{W}\|^2 - \|J'(\tau(\mathbf{U}^*)) - J'(\mathbf{Y})\| \, \|\tau''(\mathbf{U}^*)[\mathbf{W}, \mathbf{W}]\| \\ &\geq (\alpha\gamma - \beta\delta\Delta)\|\mathbf{W}\|^2 > 0 \end{aligned}$$

for all $\mathbf{W} \in W_{\mathbf{U}^*}$, where we used $J'(\mathbf{Y}) = 0$. By (4.2), this shows (i).

In particular, $\mathbf{U}^*$ is the unique solution of $j'(\mathbf{U}) = 0$ on some neighborhood of $\mathbf{U}^*$ in $\mathbf{U}^* + W_{\mathbf{U}^*}$. We again have to refer to Proposition 2.4 which implies that τ is a diffeomorphism between such a neighborhood and a neighborhood of $\tau(\mathbf{U}^*)$ in $\mathcal{T}_{\mathbf{r}}$. Hence $\tau(\mathbf{U}^*)$ is the unique minimizer of J in a neighborhood of itself in $\mathcal{T}_{\mathbf{r}}$. Since this set is open in $\mathcal{T}_{\leq \mathbf{r}}$, (ii) follows. ∎

4.2. A positive definiteness result for least squares approximation. In the previous theorem, Δ depended on J, but also on the constants γ, δ bounding the derivatives of τ at $\mathbf{U}^*$. It is natural to ask whether the condition on the latter also can be replaced by conditions on the minimizer $\mathbf{Y}$. In the case of least-square approximation, the answer is positive in certain cases: We can impose an (unfortunately rather pessimistic) separation condition on the singular values of the unfoldings of the tensor $\mathbf{Y}$ to be approximated. The exact statement is subject of the next theorem in this section. We will now consequently use the notation $\|A\|_F$ and $\langle A, B\rangle = \operatorname{tr}(A^T B)$ for the Frobenius (Euclidian) norm and inner product, and $\|A\|_2$ for the spectral norm of matrices.

THEOREM 4.2. *Let $\mathbf{r} = (r_1, r_2 \ldots, r_{d-1})$ be fixed, and let $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be such that for each of its unfoldings $\mathbf{Y}^{(i)}$ defined in* (2.4)*, the r_i-th singular value $\overline{\sigma}_i := \sigma^{(i)}_{r_i}$ is well separated from the next lower eigenvalue $\underline{\sigma}_i := \sigma^{(i)}_{r_{i+1}}$, so that*

$$(d-1)\underline{\sigma}_i < \overline{\sigma}_i.$$

Let $\mathbf{Y}_{r_i}$ denote the (unique) best rank-r_i approximation of $\mathbf{Y}^{(i)}$. Then, if a TT tensor $\mathbf{X}^* = \tau(\mathbf{U}^*) \in \mathcal{T}_{\mathbf{r}}$ fulfills

$$\|\mathbf{Y}_{r_i} - (\mathbf{X}^*)^{(i)}\|_2 =: \delta_i < \frac{\overline{\sigma}_i - (d-1)\underline{\sigma}_i}{d}, \tag{4.4}$$

for all $i = 1, 2, \ldots, d-1$, the Hessian $j''(\mathbf{U}^*)$ of the functional

$$j(\mathbf{U}) = \frac{1}{2}\|\mathbf{Y} - \tau(\mathbf{U})\|_F^2 \tag{4.5}$$

at $\mathbf{U}^*$ is positive definite on $W_{\mathbf{U}^*}$ and hence (MA) *holds.*

Before we approach the proof, we highlight the following special cases.

1. For reconstruction problems, in which the rank of the tensor $\mathbf{Y}$ to be approximated is known to equal $\mathbf{r}$, we have $\underline{\sigma}_i = 0$; therefore, $j''(\mathbf{U}^*)$ is positive definite for all $\mathbf{U}^*$ with $\tau(\mathbf{U}^*) = \mathbf{Y}$. This can also be seen from (4.3), since the second term vanishes in that case. However, recall from Proposition 3.11 that Algorithm 1 is usually finite and exact in that case.

2. For the matrix case $d = 2$, the gap condition reads $\underline{\sigma}_1/\overline{\sigma}_1 < 1$, which is equivalent to the uniqueness of a best rank-r_1 approximation $Y_{r_1} = (\mathbf{U}_1^*)^L(\mathbf{U}_2^*)^R$. It is not too difficult to see that this condition is also neccessary for (MA) to hold. Namely, if it is not satisfied, we can find a curve on the level set of best approximations which cannot be obtained by transformations $(\mathbf{U}_1^*)^L A^{-1} A (\mathbf{U}_2^*)^R$.

The proof of Theorem 4.2 is is a bit more involved. The technical details are provided Lemma 4.3. We will make heavy use of the matrices

$$\mathbf{U}^{\leq i} = [U_1(x_1)U_2(x_2)\ldots U_i(x_i)] \in \mathbb{R}^{n_1 n_2 \ldots n_i \times r_i}$$

and

$$\mathbf{U}^{\geq i} = [U_i(x_i)U_{i+1}(x_{i+1})\ldots U_d(x_d)] \in \mathbb{R}^{r_{i-1} \times n_i n_{i+1} \ldots n_d}$$

introduced in (2.2) and (2.3), and additionally define similar matrices by letting

$$\mathbf{U}^{\leq i}(\mathbf{V}_j) := [U_1(x_1)U_2(x_2)\ldots V_j(x_j)\ldots U_i(x_i)] \in \mathbb{R}^{n_1 n_2 \ldots n_i \times r_i},$$

for $\mathbf{V}_j \in \mathbb{R}^{r_{j-1}\times n_j \times r_j}$. That is, $\mathbf{U}^{\leq i}(\mathbf{V}_j)$ is obtained by replacing in $\mathbf{U}^{\leq i}$ the component $\mathbf{U}_j$ by $\mathbf{V}_j$. In the same way, we define

$$\mathbf{U}^{\geq i}(\mathbf{V}_j) := [U_i(x_i)U_{i+1}(x_{i+1})\dots V_j(x_j)\dots U_d(x_d)] \in \mathbb{R}^{r_{i-1}\times n_i n_{i+1}\dots n_d},$$

The reason for this notation is that $\tau(\mathbf{U})$ and its derivatives may be expressed in a short, versatile way: Using the mapping

$$\text{ten}\colon \mathbb{R}^{(n_1\dots n_i)\times(n_{i+1}\dots n_d)} \to \mathbb{R}^{n_1\times n_2\times\dots\times n_d}$$

that takes an unfolding of a tensor to the tensor itself, we have the identities

$$\tau(\mathbf{U}) = \text{ten}(\mathbf{U}^{\leq i}(\mathbf{U}_i)\mathbf{U}^{\geq i+1}) = \text{ten}(\mathbf{U}^{\leq i-1}\mathbf{U}^{\geq i}(\mathbf{U}_i))$$

for $i = 1, 2, \dots, d$, as well as the representations

$$\tau'(\mathbf{U})[\mathbf{W}] = \sum_{i=1}^{d} \tau(\mathbf{U}_1, \dots, \mathbf{W}_i, \dots, \mathbf{U}_d) = \sum_{i=1}^{d} \text{ten}(\mathbf{U}^{\leq i-1}\mathbf{U}^{\geq i}(\mathbf{W}_i)) \tag{4.6}$$

and

$$\begin{aligned}\tau''(\mathbf{U})[\mathbf{W},\mathbf{W}] &= 2\sum_{i=1}^{d}\sum_{j>i}^{d} \tau(\mathbf{U}_1, \dots, \mathbf{W}_i, \dots, \mathbf{W}_j, \dots, \mathbf{U}_d) \\ &= 2\sum_{i=1}^{d}\sum_{j>i}^{d} \text{ten}\left(\mathbf{U}^{\leq i}(\mathbf{W}_i)\mathbf{U}^{\geq i+1}(\mathbf{W}_j)\right).\end{aligned} \tag{4.7}$$

The derivatives of τ will be estimated based on the following lemma.

LEMMA 4.3. *Let* $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_d) \in \mathcal{U}$ *be a left orthonormal TT component vector, that is,* $(\mathbf{U}_i^L)^T\mathbf{U}_i^L = I_{r_i}$ *for* $i = 1, 2, \dots, d-1$.*, see* (2.14). *Further, let* $\mathbf{W} \in W_{\mathbf{U}}$.

(i) *For* $1 \leq j \leq i \leq d$ *there holds*

$$(\mathbf{U}^{\leq i})^T\mathbf{U}^{\leq i} = I_{r_i}, \quad \|\mathbf{U}^{\leq i}(\mathbf{W}_i)\|_F = \|\mathbf{W}_i\|_F, \tag{4.8}$$

(ii) *For* $i \neq j$ *the matrices* $\mathbf{U}^{\leq i-1}\mathbf{U}^{\geq i}(\mathbf{W}_i)$ *and* $\mathbf{U}^{\leq j-1}\mathbf{U}^{\geq j}(\mathbf{W}_j)$ *are orthogonal with respect to the Frobenius inner product.*

(iii) *It holds*

$$\|\tau'(\mathbf{U})\mathbf{W}\|_F^2 = \sum_{i=1}^{d} \|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F^2. \tag{4.9}$$

(iv) *Let* $\mathbf{X} = \tau(\mathbf{U})$. *Denote by* γ_i, Γ_i *the smallest resp. largest singular value of the unfolding* $\mathbf{X}^{(i)}$. *Then for* $1 \leq j \leq i \leq d$ *the estimate*

$$\gamma_i\|\mathbf{W}_i\|_F \leq \|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F = \|\mathbf{U}^{\geq j}(\mathbf{W}_i)\|_F \leq \Gamma_i\|\mathbf{W}_i\|_F \tag{4.10}$$

holds.

Proof. The first statement in (i) follows immediately from the left orthonormality of the component functions and (2.7). For the second, we observe that $\mathbf{U}^{\leq i}(\mathbf{W}_i) =$

$\mathbf{U}^{\le i-1}R(\mathbf{W}_i)$ with the right-unfolding defined in (2.2), by which we obtain using the previous result that

$$\langle \mathbf{U}^{\le i}(\mathbf{W}_i), \mathbf{U}^{\le i}(\mathbf{W}_i)\rangle = \langle (\mathbf{U}^{\le i-1})^T\mathbf{U}^{\le i-1}\mathbf{W}_i^R, \mathbf{W}_i^R\rangle = \|\mathbf{W}_i^R\|_F = \|\mathbf{W}_i\|_F.$$

For $j < i$,

$$\|\mathbf{U}^{\ge j}(\mathbf{W}_i)\|_F = \|\mathbf{U}^{\ge i}(\mathbf{W}_i)\|_F$$

again holds as a consequence of left orthonormality, which one can use to sum up the components left of W_i to obtain the above identiy. To show (ii), let $j < i$, then it follows from (i) and (2.7) that

$$\langle \mathbf{U}^{\le j-1}\mathbf{U}^{\ge j}(\mathbf{W}_j), \mathbf{U}^{\le i-1}\mathbf{U}^{\ge i}(\mathbf{W}_i)\rangle = \langle \mathbf{U}^{\ge j}(\mathbf{W}_j), \mathbf{U}^{\ge j}(\mathbf{W}_i)\rangle$$

is zero, since $(\mathbf{W}_j^L)^T\mathbf{U}_i^L = 0$. Now (iii) follows from (i),(ii) and (4.6). Finally, we see that

$$\mathbf{X}^{(i)} = \mathbf{U}^{\le i}\mathbf{U}^{\ge i+1}$$

is a full rank decomposition with $\mathbf{U}^{\le i}$ being orthogonal by (a). Thus, the singular values of $\mathbf{U}^{\ge i+1}$ coincide with those of $\mathbf{X}^{(i)}$. Using that $\mathbf{U}^{\ge i}(\mathbf{W}_i) = \mathbf{U}^{\ge i+1}\mathbf{W}_i^L$, the inequality

$$\gamma_i^2\|\mathbf{W}_i^L\|_F^2 \le \langle (\mathbf{U}^{\ge i+1})^T\mathbf{U}^{\ge i+1}\mathbf{W}_i^L, \mathbf{W}_i^L\rangle \le \Gamma_i^2\|\mathbf{W}_i^L\|_F^2$$

shows (iv), since the middle term equals $\|\mathbf{U}^{\ge i+1}\mathbf{W}_i^L\|_F^2$. ☐

We are now in the position to give the proof of Theorem 4.2.

Proof of Theorem 4.2. We will write $\mathbf{U}$ instead of $\mathbf{U}^*$ and $\mathbf{X} = \tau(\mathbf{U})$ for brevity. At first, we note that by Proposition 3.5, the condition (MA) transfers from any $\mathbf{U} \in \mathcal{U}$ to its orbit $\mathcal{M}_{\mathbf{U}}$. It therefore suffices to verify the condition (4.2) for left orthonormal $\mathbf{U} \in \mathcal{U}$ (Also note that (4.4) is a condition on the orbit of $\mathbf{U}$). Let therefore $\mathbf{W} \in W_{\mathbf{U}}$ be nonzero. Differentiating j given by (4.5) we obtain

$$j''(\mathbf{U})[\mathbf{W},\mathbf{W}] = \langle \tau'(\mathbf{U})\mathbf{W}, \tau'(\mathbf{U})\mathbf{W}\rangle + \langle \mathbf{Y} - \mathbf{U}, \tau''(\mathbf{U})[\mathbf{W},\mathbf{W}]\rangle. \tag{4.11}$$

We start by estimating the second term on the right hand side. We denote the i-th unfolding of $\mathbf{Y} - \mathbf{X}$ by $\mathbf{Z}^{(i)}$. Using the expression (4.7), we have

$$|\langle \mathbf{Z}^{(i)}, \tau''(\mathbf{U})[\mathbf{W},\mathbf{W}]\rangle| \le 2\sum_{i=1}^{d}\sum_{j>i}^{d} |\langle \mathbf{Z}^{(i)}, \mathbf{U}^{\le i}(\mathbf{W}_i)\mathbf{U}^{\ge i+1}(\mathbf{W}_j)\rangle|.$$

Utilising a singular value decomposition $\mathbf{Z}^{(i)} = P_i\Sigma_iQ_i$ and the estimate $\|AB\|_F \le \|A\|_2\|B\|_F$, we have for each term from this sum

$$\begin{aligned} |\langle \mathbf{Z}^{(i)}, \mathbf{U}^{\le i}(\mathbf{W}_i)\mathbf{U}^{\ge i+1}(\mathbf{W}_j)\rangle| &= |\langle Q_i(\mathbf{U}^{\ge i+1}(\mathbf{W}_j))^T, \Sigma_i^TP_i^T\mathbf{U}^{\le i}(\mathbf{W}_i)\rangle| \\ &\le \|Q_i(\mathbf{U}^{\ge i+1}(\mathbf{W}_j))^T\|_F\|\Sigma_i\|_2 \cdot \|P_i^T\mathbf{U}^{\le i}(\mathbf{W}_i)\|_F \\ &= \|\Sigma_i\|_2\|\mathbf{U}^{\ge i+1}(\mathbf{W}_j)\|_F\|\mathbf{W}_i\|_F, \end{aligned} \tag{4.12}$$

where we also used (4.8) in the last line. The singular values σ contained in Σ_i are bounded by

$$\sigma \le \|\mathbf{Z}^{(i)}\|_2 = \|\mathbf{Y}^{(i)} - \mathbf{Y}_{r_i}\|_2 + \|\mathbf{Y}_{r_i} - \mathbf{X}^{(i)}\|_2 < \underline{\sigma}_i + \delta_i.$$

Additionally, let again γ_i denote the smallest singular value of $\mathbf{X}^{(i)}$, then, since $\overline{\sigma}_i$ is the smallest singular value of $\mathbf{Y}_{r_i}$, it holds that $|\overline{\sigma}_i - \gamma_i| \leq \|\mathbf{Y}_{r_i} - \mathbf{X}^{(i)}\|_2 = \delta_i$, see [21, Eq. (3.5.32)]. Hence, by (4.10),

$$\|\mathbf{W}_i\|_F \leq \gamma_i^{-1}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F \leq (\overline{\sigma}_i - \delta_i)^{-1}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F.$$

Inserting these relations into (4.12), the assumption on δ_i implies

$$\begin{aligned}\langle \mathbf{Z}^{(i)}, \mathbf{U}^{\leq i}(\mathbf{W}_i)\mathbf{U}^{\geq i+1}(\mathbf{W}_j)\rangle &\leq (\underline{\sigma}_i + \delta_i)(\overline{\sigma}_i - \delta_i)^{-1}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F\|\mathbf{U}^{\geq j}(\mathbf{W}_j)\|_F \\ &< \frac{1}{d-1}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F\|\mathbf{U}^{\geq j}(\mathbf{W}_j)\|_F.\end{aligned}$$

for each i, so that we find

$$\tau''(\mathbf{U})[\mathbf{W}, \mathbf{W}] < \frac{2}{d-1}\sum_{i=1}^{d}\sum_{j>i}^{d}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F\|\mathbf{U}^{\geq j}(\mathbf{W}_j)\|_F.$$

Now, taking (4.9) into account, we can bound (4.11) from below by

$$\begin{aligned}j''(\mathbf{U})[\mathbf{W}, \mathbf{W}] &> \sum_{i=1}^{d}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F^2 - \frac{2}{d-1}\sum_{i=1}^{d}\sum_{j>i}^{d}\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F\|\mathbf{U}^{\geq j}(\mathbf{W}_j)\|_F \\ &= \frac{1}{d-1}\sum_{i=1}^{d}\sum_{j>i}^{d}\Big(\|\mathbf{U}^{\geq i}(\mathbf{W}_i)\|_F \;-\; \|\mathbf{U}^{\geq j}(\mathbf{W}_j)\|_F\Big)^2 \geq 0,\end{aligned}$$

which completes the proof. □

5. Conclusions and perspectives. We have shown local linear convergence of the alternating linear scheme for the TT format, supplementing the linear convergence behaviour of TT-ALS as observed in practice [20] by an according theoretical analysis. The proof bases on the convergence of the nonlinear Gauss-Seidel method, and shows that convergence does not depend on a specific component realization $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_d)$ of a tensor $\tau(\mathbf{U})$, but rather on the orbits $\mathcal{M}_{\mathbf{U}}$ of equivalent TT representations. As detailed in Section 3.3, an important consequence is that the proof therefore also holds for the ALS-QR algorithm as proposed in [20]. There, a QR orthogonalization step is performed after each component optimization step each to keep the resulting equations for the components $\mathbf{U}_i$ well-conditioned, and we found this step to be an essential ingredient in view of the practical applicability of the ALS algorithm.

Our general idea of proof as pursued in this work extends to the Tucker format and, more generally, to those tensor networks [13] for which the redundancy of the respective parametrization can be characterized explicitly as exemplified in this work for the TT format. Also, it should be investigated, if on the basis of our present proof convergence of the promising DMRG algorithm [40, 38, 34], used for eigenvalue computations in quantum physics and investigated lately under the acronym MALS for more general optimization tasks in [20], also may be verified.

The major task that remains in the TT case as well as in others is probably to verify the main assumption, i.e. the rank condition (MA) on the Hessian j'' of the composed functional $j = J \circ \tau$. For the TT case, we showed here that this assumption holds true for two interesting special cases, but finding generic conditions under which the assumption holds appears to be a rather nontrivial task. Note that the condition

(MA) also is on the very basis of proofs of convergence for many other algorithms that are or might be used in tensor optimization, examples including for instance the nonlinear (parallelizable) Jacobi method for the treatment of optimization tasks, or various variants of Newton's method applied to the gauge space characterized in [19] (see also [8] for an analogous approach for the Tucker format). As well, (MA) implies positive answers to important theoretical questions as existence of local best approximations on manifolds, local uniqueness of solutions of more global convex optimization tasks etc. Therefore, further necessary or sufficient conditions giving a characterization of cases where (MA) holds are strongly desirable.

REFERENCES

[1] J. Ballani, L. Grasedyck, M. Kluge, Black Box Solution of SPDEs in H-Tucker, in preparation.
[2] J. Ballani, L. Grasedyck, A Projection Method to Solve Linear Systems in Tensor Format, preprint available at *http://www.dfg-spp1324.de/download/preprints/preprint100.pdf*
[3] J.C. Bezdek, R.J. Hathaway, Convergence of alternating optimization, Neural Parallel Sci. Comput., 11 (2003), pp. 351–368.
[4] D. Conte, C. Lubich, An error analysis of the multi-configuration time-dependent Hartree method of quantum dynamics, M2AN, 44 (2010), pp. 759–780.
[5] V. De Silva, L.-H. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[6] S.V. Dolgov, I.V. Oseledets, Solution of linear systems and matrix inversion in the TT-format, preprint, available at *http://www.mis.mpg.de/preprints/2011/preprint2011_19.pdf*.
[7] J. Dieudonné, Treatise on Analysis. Volume III, Academic Press, New York, 1972.
[8] L. Eldén, B. Savas, Perturbation theory and optimality conditions for the best multilinear rank approximation of a tensor, preprint, submitted to SIAM J. Matrix Anal. Appl., available at *http://www.cs.utexas.edu/ berkant/*.
[9] M. Espig, Effziente Bestapproximation mittels Summen von Elementartensoren in hohen Dimensionen, Ph.D. thesis, Leipzig, 2007.
[10] M. Espig, W. Hackbusch, T. Rohwedder, R. Schneider, Variational Calculus with Sums of Elementary Tensors of Fixed Rank, to appear in Num. Math. (2012).
[11] A. Falcó, W. Hackbusch, On minimal subspaces in tensor representations, to appear in Found. Comput. Math. (2012), preprint available at *http://www.mis.mpg.de/publications/preprints/2010/prepr2010-70.html*.
[12] A. Falcó, A. Nouy, Proper Generalized Decomposition for Nonlinear Convex Problems in tensor banach spaces, to appear in Numer. Math. (20112), preprint available at *http://arxiv.org/abs/1106.4424*.
[13] M. Espig, W. Hackbusch, S. Handschuh, R. Schneider, Optimization Problems in Contracted Tensor Networks, preprint, available at *http://www.dfg-spp1324.de/download/preprints/preprint108.pdf*
[14] L. Grasedyck, Hierarchical singular value decomposition of tensors, SIAM. J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.
[15] L. Grasedyck, W. Hackbusch, An Introduction to hierarchical ($\mathcal{H}$-) Rank and TT-Rank of Tensors with Examples, Comput. Methods Appl. Math, 3 (2011), pp. 291–304.
[16] W. Hackbusch, S. Kühn, A new scheme for the tensor representation, J. Fourier Anal. Appl., 15 (2009), pp. 706–722 (2009)
[17] E. Hairer, C. Lubich, G. Wanner, Geometrical Numerical Integration - Structure-Preserving Algorithms for Ordinary Differential Equations, 2nd edition, Springer, Berlin, 2006.
[18] C.J. Hillar, L.-H. Lim, Most tensor problems are NP hard, preprint, available at *http://www.msri.org/people/members/chillar/files/tensorNPhardApril8.pdf*.
[19] S. Holtz, T. Rohwedder, R. Schneider, On manifolds of tensors of fixed TT-rank, Numer. Math., DOI: 10.1007/s00211-011-0419-7 (2011).
[20] S. Holtz, T. Rohwedder, R. Schneider, The alternating linear scheme for tensor optimisation in the TT format, to appear in SIAM J. Sci. Comp. (2012).
[21] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambrifge University Press, Cambridge, 1994.
[22] T. Huckle, K. Waldherr, T. Schulte-Herbrüggen, Exploiting matrix symmetries and physical symmetries in matrix-product states and tensor trains, in preparation.
[23] H.B. Keller, On the solution of singular and semidefinite linear systems by iteration, J. Soc.

Ind. Appl. Math., Ser. B, Numer. Anal., 2 (1965), pp. 281–290.
[24] B. N. Khoromskij, I. V. Oseledets, DMRG + QTT approach to high-dimensional quantum molecular dynamics, preprint, available at *http://www.mis.mpg.de/preprints/2010/preprint2010_69.pdf*.
[25] O. Koch, C. Lubich, Dynamical Low-rank Approximation of Tensors, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2360–2375.
[26] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, SIAM Rev., 51 (2009), pp. 455–500.
[27] D. Kressner, C. Tobler, Preconditioned low-rank methods for high-dimensional elliptic PDE eigenvalue problems, Comp. Methods. Appl. Math., 3 (2011), pp. 363–381.
[28] Y.-J. Lee, J. Wu, J. Xu, L. Zikatanov, On the convergence of iterative methods for semidefinite linear systems, SIAM J. Matrix. Anal. Appl., 28 (2006), pp. 634–641
[29] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.70
[30] I.V. Oseledets, Tensor-train decomposition, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.
[31] I.V. Oseledets, TT Toolbox 2.0: Fast multidimensional array operations in TT format, preprint, available online at *http://spring.inm.ras.ru/osel/wp-content/plugins/wp-publications-archive/openfile.php?action=open&file=27*.
[32] I.V. Oseledets, E.E. Tyrtyshnikov, Tensor tree decomposition does not need a tree, preprint, available at *http://pub.inm.ras.ru/pub/inmras2009-08.pdf*.
[33] S. Schechter, Iteration methods for nonlinear problems, Trans. Am. Math. Soc., 104 (1962), pp. 179–189.
[34] U. Schollwöck, The density-matrix renormalization group, Rev. Mod. Phys., 77 (2005), pp. 259–315.
[35] L. R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika, 31 (1966), pp. 279–311.
[36] A. Uschmajew, Local convergence of the alternating least squares algorithm for canonical tensor approximation, to appear in SIAM J. Matrix Anal. Appl. (2012).
[37] A. Uschmajew, B. Vandereyken *The geometry of algorithms using hierarchical tensors*, in preparation.
[38] F. Verstraete, J. I. Cirac, J. I. Latorre, E. Rico, and M. M. Wolf, Renormalization-Group Transformations on Quantum States, Phys. Rev. Lett., 94 (2005), 140601.
[39] G. Vidal, Efficient classical simulation of slightly entagled quantum computation, Phys. Rev. Lett., 91 (2003), 147902.
[40] S. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett., 69 (1992), pp. 2863–2866.
[41] J. Wu, Y.-J. Lee, J. Xu, L. Zikatanov, Convergence analysis on iterative methods for semidefinite systems, J. Comput. Math., 26 (2008), pp. 797–815.

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.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.