DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation

A. Uschmajew

Preprint 103

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 the Alternating Least Squares Algorithm For Canonical Tensor Approximation

A. Uschmajew

Preprint 103

LOCAL CONVERGENCE OF THE ALTERNATING LEAST SQUARES ALGORITHM FOR CANONICAL TENSOR APPROXIMATION

ANDRÉ USCHMAJEW*

Abstract. A local convergence theorem for calculating canonical low-rank tensor approximations (PARAFAC, CANDECOMP) by the ALS algorithm is established. The main assumption is that the Hessian matrix of the problem is positive definite modulo the scaling indeterminacy. A discussion, whether this is realistic, and numerical illustrations are included. Also regularization is discussed.

Key words. ALS, low-rank approximation, nonlinear Gauss-Seidel, PARAFAC

AMS subject classifications. 15A69, 65D15, 65F30

1. Introduction. According to the review article of Kolda and Bader [9], the alternating least squares algorithm (ALS) is still the "workhorse" in computing low-rank approximations and decompositions of high-order tensors. It has been widely used in such fields as psychometrics, chemometrics and signal processing (see the references in [9]). The reason for the popularity of this algorithm lies in the fact that it is very simple, conceptually and numerically, while still delivering astonishing good results in many cases, if employed with care [17].

In the present paper we investigate the ALS algorithm for low-rank approximation by tensors in the canonical format (also known as CP, PARAFAC or CANDECOMP). The great majority of the literature focuses on global properties of the iteration like the existence of convergent subsequences and critical points, or the occurrence of swamps [5, 6, 11, 12, 15]. But any widely used algorithm should desirably also be backed by a local convergence theory. To prevent any misunderstandings: by a local convergence theory we mean a theory for the parameters (iterates) of an algorithm, not for the residuals (loss function).

Surprisingly, there are few works in this direction, an exception being the work of Zhang and Golub [19] on the rank-one approximation. For higher ranks most of the difficulties with the global behavior of the ALS iteration seem to be intimately related to the fact that the approximation problem itself can be ill-posed or ill-conditioned [3]. We will not enter into this discussion in the present paper but rather assume that a local minimum exists, since otherwise the question of local convergence does not make much sense.

There is a well-developed local theory for the so called nonlinear block Gauss-Seidel method of alternating optimization [1, 13, 16]. It can be shown that, up to higher-order terms, this method locally equals the linear block Gauss-Seidel iteration applied to the Hessian matrix at the solution. Thus it is locally linearly convergent (essentially at the same rate as the linear Gauss-Seidel) provided that this Hessian matrix is positive definite. The problem is that local minima in canonical low-rank tensor approximations do not have this property due to the nonuniqueness caused by the scaling indeterminacy. (The permutation indeterminacy is irrelevant in a local theory.) On the other hand, it is known that the Gauss-Seidel method for

*Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (uschmajew@math.tu-berlin.de). This research was supported in part by the DFG (Deutsche Forschungsgemeinschaft) through the Berlin Mathematical School and MATHEON.

semidefinite linear systems is convergent up to elements in the null space of the system matrix [7, 10]. Since an ALS algorithm usually implements some sort of normalization procedure to remove the scaling indeterminacy, both theories can be combined to obtain a local convergence result under very reasonable assumptions (Theorem 3.3 below), which, however, might be very difficult to verify a priori.

One convenient aspect of our approach is that it avoids the explicit use of Lagrange multipliers. It is therefore in principle applicable for more delicate types of redundancies as they for instance occur in the Tucker format or in the newly developed TT format [14]. This will be elaborated elsewhere.

The main ideas are not specifically related to the least squares error as the loss function to be minimized. We will consider arbitrary loss functions as well (Theorem 3.4). For them, the global minimization of each ALS direction should be replaced by a single Newton step, since a global minimization might not necessarily stay local.

Generalization to the complex case is not completely straightforward. The problem is that it is subtle to define a format which removes the scaling indeterminacy. We will comment on that issue at the related points of our exposition.

Notation. We use the notation $f'(\mathbf{x})$ for the derivative of a function f at $\mathbf{x}$ and $f''(\mathbf{x})$ for the Hessian at $\mathbf{x}$. By $(\cdot,\cdot)$ and $\|\cdot\|$ we denote the Euclidian (Frobenius) inner product and norm, respectively. However, we will write $f''(\mathbf{x})[\mathbf{h},\mathbf{h}]$ instead of $(\mathbf{h}, f''(\mathbf{x})\mathbf{h})$.

2. The ALS algorithm. For the sake of clarity, we restrict ourselves most of the time to third-order tensors. The reasoning for the higher-order case is completely analog. The matrix case is not adequately covered by our considerations since the crucial Assumption 1 below requires the local essential uniqueness of the CP decomposition, which is reasonable for high-order tensors, but only satisfied by rank-one matrices.

Let $n_1, n_2, n_3 \in \mathbb{N} \setminus \{1\}$ and $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{T} \neq 0$, be a real third-order tensor, treated here as a three-dimensional array. Given $r \in \mathbb{N}$, let

$$\mathcal{X} = \mathbb{R}^{n_1 \times r} \times \mathbb{R}^{n_2 \times r} \times \mathbb{R}^{n_3 \times r}.$$

The elements of $\mathcal{X}$ will be denoted by $\mathbf{x} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$.

Consider the function

$$f \colon \mathcal{X} \to \mathbb{R} \colon \mathbf{x} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \mapsto \frac{1}{2}\Big\|\mathcal{T} - \sum_{j=1}^{r} a_j \otimes b_j \otimes c_j\Big\|^2,$$

where a_j, b_j and c_j are supposed to be the columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, respectively. The matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are called factor matrices in the literature. We seek for a solution of

$$f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \min. \tag{2.1}$$

It is assumed that at least one local minimum of (2.1) exists. It will be denoted by $\mathbf{x}^* = (\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$.

The alternating least squares algorithm (ALS) is a simple method for (hopefully) solving (2.1). Given a starting point $\mathbf{x}^{(0)}$ (which is supposed to be close to $\mathbf{x}^*$), it consists in iterating the cycle

$$\begin{aligned}\mathbf{A}^{(n+1)} &= \operatorname*{argmin}_{\mathbf{A}\in\mathbb{R}^{n_1\times r}} f(\mathbf{A},\mathbf{B}^{(n)},\mathbf{C}^{(n)}),\\ \mathbf{B}^{(n+1)} &= \operatorname*{argmin}_{\mathbf{B}\in\mathbb{R}^{n_2\times r}} f(\mathbf{A}^{(n+1)},\mathbf{B},\mathbf{C}^{(n)}),\\ \mathbf{C}^{(n+1)} &= \operatorname*{argmin}_{\mathbf{C}\in\mathbb{R}^{n_3\times r}} f(\mathbf{A}^{(n+1)},\mathbf{B}^{(n+1)},\mathbf{C}).\end{aligned} \tag{2.2}$$

This algorithm is a particular example of the nonlinear block Gauss-Seidel (relaxation) method [13, 16]. The name ALS stems from the fact that each micro-iteration step in (2.2) is a linear least squares problem. If every such step possesses a unique solution (if not, this is usually enforced by applying a pseudo inverse), then one cycle (2.2) defines an operator S, that is,

$$(\mathbf{A}^{(n+1)},\mathbf{B}^{(n+1)},\mathbf{C}^{(n+1)}) = \mathbf{x}^{(n+1)} = S(\mathbf{x}^{(n)}) = S(\mathbf{A}^{(n)},\mathbf{B}^{(n)},\mathbf{C}^{(n)}). \tag{2.3}$$

From now on we will only consider local minima of (2.1) in the open subset

$$\hat{\mathcal{X}} = \{(\mathbf{A},\mathbf{B},\mathbf{C}) \in \mathcal{X} \mid a_j \neq 0,\ b_j \neq 0,\ c_j \neq 0 \text{ for } j = 1,2,\dots,r\}.$$

We assume such minima to exist. Restricting to $\hat{\mathcal{X}}$ is reasonable to avoid pseudo inverses, since only if $\mathbf{x}^* \in \hat{\mathcal{X}}$ we can hope (2.2) to have unique solutions (take for instance $c_1^{(n)} = 0$ in the first line of (2.2)). We consider other local minima as too degenerate for our framework. For them, at least one rank-one term vanishs, so the rank parameter r should be adjusted. However, it seems to be a difficult question whether such local minima can really exist if, say, the canonical rank of the target tensor $\mathcal{T}$ is larger than or equal to r.

The major difficulty in the analysis of algorithm (2.2) lies in the fact that $\mathbf{x}^*$ cannot be an isolated local minimum of (2.1), since every rank-one term $a_j^* \otimes b_j^* \otimes c_j^*$ may be replaced by $(\alpha_j a_j^*) \otimes (\beta_j b_j^*) \otimes (\gamma_j c_j^*)$ as long as $\alpha_j\beta_j\gamma_j = 1$. We will call this operation a rescaling of $\mathbf{x}^*$ if α, β and γ are positive. In fact, every such rescaled solution itself is a local minimum of (2.1) and a fixed point of the iteration (2.3). So there is no reason why the iteration should, if at all, converge to a particular prescribed solution $\mathbf{x}^*$.

Additionally, when applying the ALS algorithm in the naive form (2.2) it can happen that a component, say $a_1^{(n)}$, tends to infinity while another, say $b_1^{(n)}$, compensates this by tending to zero, such that $a_1^{(n)} \otimes b_1^{(n)} \otimes c_1^{(n)}$ remains bounded. This deteriorates the condition of each micro-step.

For both reasons a normalization strategy has to be invoked. The usual way to do this is to represent tensors in the normalized form

$$\sum_{j=1}^{r} \sigma_j a_j \otimes b_j \otimes c_j \quad \text{with} \quad \|a_j\| = \|b_j\| = \|c_j\| = 1,\ \sigma_j \in \mathbb{R} \quad \text{for } j = 1,2,\dots,r. \tag{2.4}$$

To avoid additional parameters we will instead consider tensors in the equilibrated format which fixes the representation up to change of signs[1]:

$$\sum_{j=1}^{r} a_j \otimes b_j \otimes c_j \quad \text{with} \quad \|a_j\| = \|b_j\| = \|c_j\| \quad \text{for } j = 1,2,\dots,r. \tag{2.5}$$

[1] In the complex case, up to change of angles.

The rescaling of a tensor into the equilibrated format (2.5) without changing the signs of the vectors uniquely defines an operator $R(\mathbf{x}) = R(\mathbf{A}, \mathbf{B}, \mathbf{C})$ for the corresponding parametrization via

$$(a_j, b_j, c_j) \mapsto \Big(\frac{\delta_j a_j}{\|a_j\|}, \frac{\delta_j b_j}{\|b_j\|}, \frac{\delta_j c_j}{\|c_j\|} \Big), \quad \delta_j = (\|a_j\| \|b_j\| \|c_j\|)^{1/3}, \quad j = 1, 2, \ldots, r.$$

Note that R can be defined on the whole space $\mathscr{X}$ by continuous extension (on $\mathscr{X} \setminus \hat{\mathscr{X}}$ it is zero), but is smooth only on $\hat{\mathscr{X}}$. We will call a representation $\mathbf{x} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ equilibrated if $R(\mathbf{x}) = \mathbf{x}$. This can only happen in $\hat{\mathscr{X}}$ or at the origin.

The ALS algorithm for calculating an equilibrated local solution of (2.1) reads as follows.

Algorithm 1 (ALS with equilibration).

```
Input: $\mathbf{x}^{(0)} = (\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{C}^{(0)})$
For $n = 0, 1, 2, \ldots$:
    1. Perform one ALS cycle:
        $\tilde{\mathbf{x}}^{(n+1)} = (\tilde{\mathbf{A}}^{(n+1)}, \tilde{\mathbf{B}}^{(n+1)}, \tilde{\mathbf{C}}^{(n+1)}) = S(\mathbf{x}^{(n)})$.
    2. Equilibrate factor matrices:
        $\mathbf{x}^{(n+1)} = R(\tilde{\mathbf{x}}^{(n+1)})$.
```

Only equilibrated local minima of (2.1) can be fixed points of $R \circ S$. Consequently, side conditions are not necessary.

Other variants of the ALS algorithm are possible which for instance include normalization of the iterates in the sense of (2.4) instead of equilibration, see, e.g., [9]. To increase the numerical stability it may also be reasonable to equilibrate after each micro-step (2.2). All such variants of the algorithm are equivalent in the sense that they, given the same starting point, produce the same iterates up to rescaling. The presented version is favorable for the convergence analysis of the algorithm, but the convergence result as stated in Theorem 3.3 trivially transfers to different scaling strategies. In fact, in our numerical experiments we used normalized iterates of the form (2.4).

3. Convergence analysis.

From now on $\mathbf{x}^*$ will always denote a nonzero equilibrated local solution of (2.1). Recall that then $\mathbf{x}^* \in \hat{\mathscr{X}}$. In this section we establish a local convergence theorem for Algorithm 1 in a neighborhood of $\mathbf{x}^*$.

3.1. The positive definiteness assumption.

We first return our attention to the scaling indeterminacy again. The function f is constant on the $2r$-dimensional (in the real case not connected) submanifold[2]

$$\mathscr{M}^* = \{(\mathbf{A}^* \Delta_1, \mathbf{B}^* \Delta_2, \mathbf{C}^* \Delta_1^{-1} \Delta_2^{-1}) \in \mathscr{X} \mid \Delta_1, \Delta_2 \text{ regular diagonal matrices}\},$$

which contains all equivalent representations of $\sum_{j=1}^r a_j^* \otimes b_j^* \otimes c_j^*$ that can be obtained by rescaling (including sign changing). Since every point of $\mathscr{M}^*$ is a local minimum, the derivative f' vanishes on $\mathscr{M}^*$. Consequently, the Hessian $f''(\mathbf{x}^*)$ has at most rank $\dim \mathscr{X} - 2r = r(n_1 + n_2 + n_3) - 2r$. More precisely, let $T\mathscr{M}^*_{\mathbf{x}^*}$ denote the tangent space on $\mathscr{M}^*$ at $\mathbf{x}^*$, then $f''(\mathbf{x}^*)[\mathbf{h}, \mathbf{h}] = 0$ for all $\mathbf{h} \in T\mathscr{M}^*_{\mathbf{x}^*}$. It is, by the way, easy to see that

$$T\mathscr{M}^*_{\mathbf{x}^*} = \{(\mathbf{A}^* \Delta_1, \mathbf{B}^* \Delta_2, -\mathbf{C}^*(\Delta_1 + \Delta_2)) \in \mathscr{X} \mid \Delta_1, \Delta_2 \text{ diagonal matrices}\}. \quad (3.1)$$

[2] The main reason that this is a submanifold of the specified dimension is that the matrices $\mathbf{A}^*$, $\mathbf{B}^*$ and $\mathbf{C}^*$ contain no zero columns. In the order-d case the corresponding submanifold is of dimension $(d-1)r$.

We now make

ASSUMPTION 1. *The rank of $f''(\mathbf{x}^*)$ equals $r(n_1+n_2+n_3)-2r$, that is, the null space of $f''(\mathbf{x}^*)$ is $T\mathcal{M}^*_{\mathbf{x}^*}$.*

In other words, $f''(\mathbf{x}^*)$ shall be positive definite in every direction except those tangent to the rescalings. This implies that the parametrization $\mathbf{x}^*$ is locally essentially unique (the converse may not always be true). We will discuss in Sect. 3.4 whether Assumption 1 is realistic. In any case, it seems unavoidable for a standard convergence proof of Algorithm 1. The reason for this is the following observation.

LEMMA 3.1. *$R'(\mathbf{x}^*)$ is a projector whose null space is precisely $T\mathcal{M}^*_{\mathbf{x}^*}$.*

Proof. First of all, $R = R \circ R$ shows that

$$R'(\mathbf{x}^*) = R'(R(\mathbf{x}^*))R'(\mathbf{x}^*) = R'(\mathbf{x}^*)R'(\mathbf{x}^*)$$

is a projector. Since R is constant on the connected components[3] of $\mathcal{M}^*$, it also follows that $R'(\mathbf{x}^*)\mathbf{h} = 0$ for all $\mathbf{h} \in T\mathcal{M}^*_{\mathbf{x}^*}$.

On the other hand, R is the identity on the set of all equilibrated $\mathbf{x}$, in particular on the submanifold of all $(\mathbf{A},\mathbf{B},\mathbf{C})$ whose columns have the same norm as the corresponding columns of $(\mathbf{A}^*,\mathbf{B}^*,\mathbf{C}^*)$. Hence $R'(\mathbf{x}^*)\mathbf{h} = \mathbf{h}$ for all $\mathbf{h}$ from the tangent space of that submanifold at $\mathbf{x}^*$. Since the latter can be regarded as a Cartesian product of spheres (on which the columns of the matrices are located) it is clear that its tangent space at $\mathbf{x}^*$ is

$$U^* = \{(\mathbf{A},\mathbf{B},\mathbf{C}) \in \mathcal{X} \mid a_j \perp a_j^*,\ b_j \perp b_j^*,\ c_j \perp c_j^* \text{ for } j = 1,2,\dots,r\},$$

where $\perp$ means Euclidian orthogonality.

Finally, it is easy to see that, for instance, $R'(\mathbf{x}^*)\mathbf{h} \neq 0$ for all $\mathbf{h} \neq 0$ from

$$V^* = \{(\mathbf{A}^*\Delta, 0, 0) \in \mathcal{X} \mid \Delta \text{ diagonal matrix}\}.$$

This finishes the proof, since $\dim T\mathcal{M}^*_{\mathbf{x}^*} + \dim(U^* \oplus V^*) = \dim \mathcal{X}$. □

Interestingly, Assumption 1 already ensures that Algorithm 1 is well-defined in a neighborhood of $\mathbf{x}^*$.

LEMMA 3.2. *Assume that Assumption 1 holds. Then the ALS operator S in (2.3) is well-defined in some neighborhood of $\mathbf{x}^*$ and continuously differentiable. Moreover, $\mathbf{x}^*$ is a fixed point of S and we have $S'(\mathbf{x}^*) = I - M^{-1}f''(\mathbf{x}^*)$, where M is the lower block triangular (including the block diagonal) part of $f''(\mathbf{x}^*)$ corresponding to the partition $\mathbf{x} = (\mathbf{A},\mathbf{B},\mathbf{C})$.*

In a possibly smaller neighborhood the composition $R \circ S$ is well-defined and continuously differentiable, that is, one execution of Algorithm 1 is feasible if the current iterate $\mathbf{x}^{(n)}$ is close enough to $\mathbf{x}^*$.

Proof. The main argument is that the diagonal blocks of $f''(\mathbf{x}^*)$ are positive definite. To see this, consider for instance $\mathbf{h} \neq 0$ of the form $(\mathbf{h}_1, 0, 0)$ with $\mathbf{h}_1 \in \mathbb{R}^{n_1 \times r}$. Since, by (3.1), $\mathbf{h} \notin T\mathcal{M}^*_{\mathbf{x}^*}$ then, Assumption 1 guarantees $f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] > 0$, which shows that the first diagonal block of $f''(\mathbf{x}^*)$, corresponding to the block variable $\mathbf{A}$, is positive definite. The same reasoning works for all blocks.

It follows that the diagonal blocks of $f''(\mathbf{x})$ are positive definite for all $\mathbf{x}$ sufficiently close to $\mathbf{x}^*$. Every micro-step in (2.2) is a quadratic minimization problem whose

[3] In the complex case, $\mathcal{M}^*$ consists of only one component. One suggestion to obtain a constant equilibration operator R is to fix certain positions in the vectors a_j and b_j, and rotate the corresponding entries to the positive real axis, assuming that they are not zero in a neighborhood of $\mathbf{x}^*$. Although such positions surely exist, this is a little bit unsatisfactory.

system matrix is the corresponding diagonal block of the Hessian $f''(\mathbf{x})$ at the current point. Hence if $\mathbf{x}^{(n)}$ is close enough to $\mathbf{x}^*$, then every micro-step, taken by itself, possesses a unique (global) minimum which depends smoothly on the input. Clearly, $\mathbf{x}^*$ is a fixed point of (2.2). Therefore, we can even choose a neighborhood such small that all three micro-steps in (2.2) can be executed consecutively with unique solutions, that is, S is well-defined and smooth in this neighborhood.

In particular, we have shown that the lower block triangular part M of $f''(\mathbf{x}^*)$ is nonsingular. A proof that $S'(\mathbf{x}^*) = I - M^{-1}f''(\mathbf{x}^*)$ can be found in [1, Lemma 2].

Since S is smooth in a neighborhood of its fixed point $\mathbf{x}^* \in \hat{\mathcal{X}}$, and since $\hat{\mathcal{X}}$ is open, we also have $S(\mathbf{x}^{(n)}) \in \hat{\mathcal{X}}$ if $\mathbf{x}^{(n)}$ is close enough to $\mathbf{x}^*$, that is, $(R \circ S)(\mathbf{x}^{(n)})$ is well-defined then. □

As we have seen in the proof, Lemma 3.2 in principle holds under the weaker assumption that the diagonal blocks of $f''(\mathbf{x}^*)$ are positive definite. This is equivalent to the unique solvability of each micro-step (2.2) with input $\mathbf{x}^*$. For completeness we remark that this in turn is equivalent to the linear independence of the sets of complementary tensors

$$\{b_j^* \otimes c_j^* \mid j = 1, 2, \dots, r\}, \quad \{a_j^* \otimes c_j^* \mid j = 1, 2, \dots, r\}, \quad \{a_j^* \otimes b_j^* \mid j = 1, 2, \dots, r\}.$$

This is for instance the case, if the solution tensor $\sum_{j=1}^r a_j^* \otimes b_j^* \otimes c_j^*$ has canonical rank r [4].

3.2. Convergence theorem. Let us outline the idea of the following convergence theorem. The ALS iterator S can be locally regarded as the standard linear block Gauss-Seidel iteration applied to the semidefinite Hessian $f''(\mathbf{x}^*)$. That method is known to converge in the energy seminorm

$$|\mathbf{x}|_E = (f''(\mathbf{x}^*)[\mathbf{x}, \mathbf{x}])^{1/2}$$

of $f''(\mathbf{x}^*)$. If Assumption 1 holds, then the null space of $f''(\mathbf{x}^*)$, which is an undamped invariant subspace of the linear Gauss-Seidel method, is the tangent space $T\mathcal{M}^*_{\mathbf{x}^*}$ of the manifold $\mathcal{M}^*$. Elements in this space are essentially removed by the equilibration operator R. In particular, by Lemma 3.1,

$$|\mathbf{x}|_*^2 = \|(I - R'(\mathbf{x}^*))\mathbf{x}\|^2 + |\mathbf{x}|_E^2$$

defines a norm on $\mathcal{X}$. One can regard this norm as the energy norm of $f''(\mathbf{x}^*)$ with the zero eigenvalues replaced by 1.

Theorem 3.3. *Let $\mathbf{x}^* = (\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*)$ be an equilibrated local minimum of (2.1) for which Assumption 1 holds. Then for every $\epsilon > 0$ there exists a neighborhood of $\mathbf{x}^*$, such that for any starting point $\mathbf{x}^{(0)}$ in this neighborhood the iterates of Algorithm 1 converge linearly to $\mathbf{x}^*$ and particularly satisfy*

$$|\mathbf{x}^{(n+1)} - \mathbf{x}^*|_* \leq (q + \epsilon)|\mathbf{x}^{(n)} - \mathbf{x}^*|_*,$$

where $q = |S'(\mathbf{x}^*)|_E < 1$.

Proof. By Lemma 3.2, the operator $R \circ S$ is well-defined and continuously differentiable in a neighborhood of its fixed point $\mathbf{x}^*$. If we show that $|(R \circ S)'(\mathbf{x}^*)|_* \leq q < 1$, the assertion of the theorem will follow from the contraction principle.

It holds $f(R(\mathbf{x}^* + \mathbf{h})) = f(\mathbf{x}^* + \mathbf{h})$ for all sufficiently small $\mathbf{h}$. Since $f'(\mathbf{x}^*) = f'(R(\mathbf{x}^*)) = 0$, it follows that

$$|R'(\mathbf{x}^*)\mathbf{h}|_E^2 = f''(\mathbf{x}^*)[R'(\mathbf{x}^*)\mathbf{h}, R'(\mathbf{x}^*)\mathbf{h}] = f''(\mathbf{x}^*)[\mathbf{h}, \mathbf{h}] = |\mathbf{h}|_E^2.$$

Additionally, by Lemma 3.1, $(I - R'(\mathbf{x}^*))R'(\mathbf{x}^*) = 0$. (Actually, $R'(\mathbf{x}^*)$ is an orthogonal projector with respect to the inner product which induces the norm $|\cdot|_*$.) Hence, for all $\mathbf{h} \in \mathcal{X}$ we have

$$\begin{aligned} |(R \circ S)'(\mathbf{x}^*)\mathbf{h}|_* &= |R'(\mathbf{x}^*)S'(\mathbf{x}^*)\mathbf{h}|_* \\ &= |S'(\mathbf{x}^*)\mathbf{h}|_E \leq |S'(\mathbf{x}^*)|_E |\mathbf{h}|_E \leq |S'(\mathbf{x}^*)|_E |\mathbf{h}|_*. \end{aligned} \tag{3.2}$$

As noted in Lemma 3.2, $S'(\mathbf{x}^*)$ equals the error iteration matrix $I - M^{-1}f''(\mathbf{x}^*)$ of the linear block Gauss-Seidel method for $f''(\mathbf{x}^*)$. It is known that for semidefinite symmetric systems this error iteration matrix is a contraction in the energy seminorm, that is, $|S'(\mathbf{x}^*)|_E < 1$, see [7, Eq. (9)] or [10, Theorem 3.2]. Due to (3.2), this proves the theorem. □

For (quite involved) estimates for $q = |S'(\mathbf{x}^*)|_E$ we refer to [18].

3.3. General target functions. The concrete form of the function f only entered in Lemma 3.2, where we used that every micro-step of the ALS iteration (2.2) is a quadratic minimization problem. Due to the multilinearity of the tensor product, this is the case for any quadratic cost function $J\colon \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}$, so our theorem applies without change to

$$f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = J\Big(\sum_{j=1}^{r} a_j \otimes b_j \otimes c_j\Big). \tag{3.3}$$

Important choices for J include energy norms of selfadjoint operators of eigenvalue problems or partial differential equations. There have been attempts to use a tensor calculus in the solution of such problems in very high dimensions, see for instance [2].

In the case of a general nonlinear functional J, the micro-steps in (2.2) do not need to have unique solutions, even if assumptions on the Hessian are made. Moreover, the global minima of a micro-step might lie far away from the considered local minimum $\mathbf{x}^*$ (see the discussion in [1]). To stay local, one should replace at each micro-step the function f in (3.3) by its second-oder expansion

$$\hat{f}(\mathbf{x}) = f(\mathbf{x}^{(n)}) + f'(\mathbf{x})(\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2}f''(\mathbf{x})[\mathbf{x} - \mathbf{x}^{(n)}, \mathbf{x} - \mathbf{x}^{(n)}]$$

at the current micro-iterate and minimize that one, which is nothing else than to perform one Newton step with respect to the current block variable. This is called approximate nonlinear relaxation in [16] and Newton-SOR method in [13]. For brevity, we do not write the formulas out in detail. If we denote by $\hat{S}$ the iteration operator of that method, the claim of Lemma 3.2 and its proof hold for $\hat{S}$ (also see [13, Theorem 10.3.3]). In fact, if J is quadratic as above, then obviously $\hat{S} = S$.

Theorem 3.4. *Let $J \in C^2(\mathbb{R}^{n_1 \times n_2 \times n_3}, \mathbb{R})$ and f be defined as in* (3.3). *Let $\mathbf{x}^*$ be an equilibrated local minimum of f for which Assumption 1 holds. Then the iteration*

$$\mathbf{x}^{(n+1)} = (R \circ \hat{S})(\mathbf{x}^{(n)}).$$

is locally linearly convergent to $\mathbf{x}^*$.

3.4. Discussion of Assumption 1. We return to the least squares approximation. Assumption 1 appears quite difficult to verify. We wish to formulate some sufficient criteria, which, however, seem far from being necessary in general cases.

Define $\tau(\mathbf{x}) = \tau(\mathbf{A},\mathbf{B},\mathbf{C}) = \sum_{j=1}^{r} a_j \otimes b_j \otimes c_j$. Then $f(\mathbf{x}) = \frac{1}{2}\|\mathcal{T} - \tau(\mathbf{x})\|^2$ and

$$f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] = \|\tau'(\mathbf{x}^*)\mathbf{h}\|^2 + (\tau(\mathbf{x}^*) - \mathcal{T}, \tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]). \tag{3.4}$$

Assumption 1 states that $f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] = 0$ only if $\mathbf{h} \in T\mathcal{M}^*_{\mathbf{x}^*}$. Note that for these $\mathbf{h}$ we necessarily have $\tau'(\mathbf{x}^*)\mathbf{h} = 0$ (since τ is constant on $\mathcal{M}^*$) and therefore also $(\tau(\mathbf{x}^*) - \mathcal{T}, \tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]) = 0$.

Let us give an example, taken from [11], for which Assumption 1 does not hold. Consider $r = 3$ and $\mathcal{T}$ given pointwise by

$$\mathcal{T}_{i_1 i_2 i_3} = \sin(i_1 + i_2 + i_3).$$

One can prove that

$$\sin(i_1 + i_2 + i_3) = \sum_{j=1}^{3} \sin(i_j + \beta_j) \prod_{\substack{k=1\\k\neq j}}^{3} \frac{\sin(i_j + \beta_j + \alpha_j - \alpha_k)}{\alpha_j - \alpha_k} \tag{3.5}$$

for all $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ with $\sin(\alpha_j - \alpha_k) \neq 0$ for $j \neq k$ and all $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$ for which $\beta_1 + \beta_2 + \beta_3 = 0$. Hence if $\tau(\mathbf{x}^*) = \mathcal{T}$ is any of the exact decompositions given by (3.5), the representation can be smoothly changed by manipulations beyond the scaling indeterminacy. Geometrically this means that τ and also f are constant on a submanifold of global minima of dimension more than $2r$. The null spaces of $\tau'(\mathbf{x}^*)$ and $f''(\mathbf{x}^*)$ are therefore larger than $T\mathcal{M}^*_{\mathbf{x}^*}$ and Assumption 1 cannot hold. A same kind of example can be given for higher order tensors.

Let us formulate

ASSUMPTION 2. *We have $\tau'(\mathbf{x}^*)\mathbf{h} \neq 0$ for all $\mathbf{h} \neq T\mathcal{M}^*_{\mathbf{x}^*}$, that is, the null space of $\tau'(\mathbf{x}^*)$ is $T\mathcal{M}^*_{\mathbf{x}^*}$.*

This assumption implies that $\tau'(\mathbf{x}^*)$ is locally essentially unique and excludes examples like (3.5). If now again $\tau(\mathbf{x}^*) = \mathcal{T}$ is an exact *decomposition*, then, by (3.4), Assumption 2 is equivalent to Assumption 1, which proves the following:

THEOREM 3.5. *If $\tau(\mathbf{x}^*) = \mathcal{T}$ and Assumption 2 holds, then in a neighborhood of $\mathbf{x}^*$ the ALS algorithm is linearly convergent to $\mathbf{x}^*$.*

Let us now discuss the case of *approximation*, that is, the case $r < \operatorname{rank} \mathcal{T}$. According to (3.4), we can expect Assumption 1 to hold, if Assumption 2 holds and if $\tau(\mathbf{x}^*) - \mathcal{T}$ is sufficiently small, that is, if $\tau(\mathbf{x}^*)$ is a good approximation for $\mathcal{T}$. This argument however lacks rigor, since the norm of $\tau''(\mathbf{x}^*)$ would have to be estimated, which seems not a priori possible[4]. It is therefore very difficult to make the argument more precise in general. As so often, we have to content ourselves with an investigation of the rank-one approximation.

In that case, we have to show $\operatorname{rank} f''(\mathbf{x}^*) = n_1 + n_2 + n_3 - 2$. Since $\mathbf{x}^* \in \hat{\mathcal{X}}$, a space of this dimension is given by

$$W^* = \{\mathbf{h} = (\delta a, \delta b, \delta c) \in \mathcal{X} \mid \delta a \perp a^*,\ \delta b \perp b^*\}.$$

We have $\tau(\mathbf{x}) = a \otimes b \otimes c$,

$$\tau'(\mathbf{x}^*)\mathbf{h} = \delta a \otimes b^* \otimes c^* + a^* \otimes \delta b \otimes c^* + a^* \otimes b^* \otimes \delta c$$

[4] As far as we can see, this amounts in estimating $\sum_{j=1}^{r} \|a_j^* \otimes b_j^* \otimes c_j^*\|$. This quantity can hardly be controlled a priori, a fact closely related to the phenomenon that the canonical low-rank approximation can be an ill-posed problem ("diverging rank-one terms").

and

$$\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] = 2(\delta a \otimes \delta b \otimes c^* + \delta a \otimes b^* \otimes \delta c + a^* \otimes \delta b \otimes \delta c).$$

Let $\|\tau(\mathbf{x}^*)\| = \sigma$, that is, $\sigma^{1/3} = \|a^*\| = \|b^*\| = \|c^*\|$. Using $(a\otimes b\otimes c, \tilde{a}\otimes\tilde{b}\otimes\tilde{c}) = (a,\tilde{a})(b,\tilde{b})(c,\tilde{c})$, one verifies that for $\mathbf{h} \in W^*$ Eq. (3.4) reads

$$f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] = \sigma^{4/3}\|\mathbf{h}\|^2 - (\mathcal{T}, \tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]), \tag{3.6}$$

where of course $\|\mathbf{h}\|^2 = \|\delta a\|^2 + \|\delta b\|^2 + \|\delta c\|^2$. It is sufficient to consider $\|\mathbf{h}\| = 1$. Using that for $\mathbf{h} \in W^*$ the terms in $\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]$ are pairwise orthogonal, one may verify that, under the constraint $\|\mathbf{h}\| = 1$, the norm $\|\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]\|^2$ is maximal only if $\|\delta a\| = \|\delta b\| = \|\delta c\| = 1/\sqrt{3}$. This gives

$$\|\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]\| \le 2\sqrt{3\cdot\Big(\frac{\sigma^{1/3}}{3}\Big)^2} = \frac{2}{\sqrt{3}}\sigma^{1/3}.$$

Now notice that, if x^* locally minimizes f, we have $\sigma = \|\tau(\mathbf{x}^*)\| = (\mathcal{T}, \tau(\mathbf{x}^*)/\|\tau(\mathbf{x}^*)\|)$. Since moreover $\tau(\mathbf{x}^*)$ and $\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]$ are orthogonal for $\mathbf{h} \in W^*$, it follows that

$$(\mathcal{T}, \tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]) \le \|\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]\|\sqrt{\|\mathcal{T}\|^2 - \sigma^2} \le \frac{2}{\sqrt{3}}\sigma^{1/3}\sqrt{\|\mathcal{T}\|^2-\sigma^2}.$$

Inserting this into (3.6) leads to the estimate

$$f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] \ge \sigma^{4/3} - \frac{2}{\sqrt{3}}\sigma^{1/3}\sqrt{\|\mathcal{T}\|^2-\sigma^2}$$

for all $\mathbf{h} \in W^*$ with $\|\mathbf{h}\| = 1$. The right side is positive if

$$\sigma^2 > \frac{4}{7}\|\mathcal{T}\|^2,$$

or, equivalently,

$$\|\mathcal{T} - \tau(\mathbf{x}^*)\|^2 = \|\mathcal{T}\|^2 - \sigma^2 < \frac{3}{7}\|\mathcal{T}\|^2.$$

In the order-d case the same reasoning leads to the estimate

$$\|\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]\| \le 2\sqrt{\frac{d(d-1)}{2}\cdot\Big(\frac{\sigma^{(d-2)/d}}{d}\Big)^2} = \sqrt{\frac{2d-2}{d}}\sigma^{(d-2)/d}$$

(since $\frac{1}{2}\tau''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}]$ consists of $d(d-1)/2$ orthogonal terms then), and with that

$$f''(\mathbf{x}^*)[\mathbf{h},\mathbf{h}] \ge \sigma^{2(d-1)/d} - \sqrt{\frac{2d-2}{d}}\sigma^{(d-2)/d}\sqrt{\|\mathcal{T}\|^2-\sigma^2} \tag{3.7}$$

for all normalized $\mathbf{h}$ in an analogously defined space W^*. The right side is positive if

$$\sigma^2 > \frac{2d-2}{3d-2}\|\mathcal{T}\|^2,$$

or, equivalently,

$$\|\mathcal{T} - \tau(\mathbf{x}^*)\|^2 = \|\mathcal{T}\|^2 - \sigma^2 < \frac{1}{3-2/d}\|\mathcal{T}\|^2.$$

If $\tau(\mathbf{x}^*)$ is supposed to be the best rank-one approximation to $\mathcal{T} \neq 0$ (which always exists), we can formulate the following criterion:

Theorem 3.6. *If the Euclidian distance between an order-d tensor $\mathcal{T}$ and the set of rank-one tensor is strictly smaller than $\|\mathcal{T}\|/\sqrt{3-2/d}$, then Assumption 1 holds for any best rank-one approximation of $\mathcal{T}$. Consequently, the ALS algorithm then converges linearly in a neighborhood of a best rank-one approximation.*

Although this already is a surprisingly soft condition, we remark that it may not need to be necessary for Assumption 1 to hold. In particular, one expects estimate (3.7) to be too rough in many cases. It is however sharp and might be used to construct examples for which Assumption 1 does not hold. Consider for instance the matrix case $d = 2$ with $\mathcal{T}$ being the 2×2 identity matrix, which is the worst case when it comes to rank-one approximation. Its distance to the set of rank-one matrices is 1, which is not strictly smaller than $\|\mathcal{T}\|/\sqrt{3-2/d} = 1$. Indeed, every matrix

$$\begin{pmatrix} \sin^2 t & \sin t \cos t \\ \sin t \cos t & \cos^2 t \end{pmatrix} = \begin{pmatrix} \sin t \\ \cos t \end{pmatrix} \otimes \begin{pmatrix} \sin t \\ \cos t \end{pmatrix} \tag{3.8}$$

is a best rank-one approximation of the identity. Obviously, this set cannot be obtained by rescaling of a single rank-one matrix. Therefore this again is an example where Assumption 1 does not hold[5].

The rank-one case has also been studied in [19], but within a different framework. A quite general condition has been formulated there for the local convergence of ALS, namely the positive definiteness of a certain Lagrangian. As we believe, it is closely related to our Assumption 1.

3.5. A note on regularization.

There are several reasons why one should instead of (2.1) consider a Tikhonov regularized problem such as

$$g_\lambda(\mathbf{A},\mathbf{B},\mathbf{C}) = f(\mathbf{A},\mathbf{B},\mathbf{C}) + \lambda(\|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 + \|\mathbf{C}\|^2) = \text{ min}, \tag{3.9}$$

where $\lambda > 0$ is a regularization parameter and the norms are the corresponding Euclidian (Frobenius) matrix norms. (Note that $\|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 + \|\mathbf{C}\|^2 = \|\mathbf{x}\|^2$.)

The first reason is that this problem is always well-posed, that is, admits a global minimizer (it is coercive). This also has direct consequences to the behavior of an ALS algorithm applied to (3.9). It has been observed that swamps occur less often so that (global) convergence to a critical point is much faster [12, 15].

A second, equally important reason to consider (3.9) is that the scaling indeterminacy is completely removed. One can check that for a local minimum $\mathbf{x}^* \in \hat{\mathcal{X}}$ of (3.9) it necessarily holds $\|a_j^*\| = \|b_j^*\| = \|c_j^*\|$ for all $j = 1, 2, \dots, r$, that is, the solution is equilibrated[6]. It is hence reasonable to apply the standard convergence theory of the nonlinear block Gauss-Seidel method by assuming that $g_\lambda''(\mathbf{x}^*)$ is positive definite at a local solution. Since Lemma 3.2 holds, the local convergence of ALS applied to g_λ (without equilibration) then follows immediately from the above considerations, cf. [1, Theorem 2]. Indeed, we have

Theorem 3.7. *If λ is large enough, $g_\lambda''(\mathbf{x}^*)$ is positive definite at* global *minimizers $\mathbf{x}^*$. Consequently, the ALS algorithm (without equilibation) is locally linearly convergent at such points.*

[5] For higher ranks, Assumption 1 will never hold in the matrix case, since a low-rank decomposition UV^T might be replaced by $UAA^{-1}V^T$, which is more than just a scaling indeterminacy.

[6] Using the Lagrange multiplier rule one can rigorously show that among all rescalings $\alpha_j a_j^*$, $\beta_j b_j^*$, $\gamma_j c_j^*$ with $\alpha_j\beta_j\gamma_j = 1$ only the one that leads to an equilibrated solution minimizes the second term in (3.9).

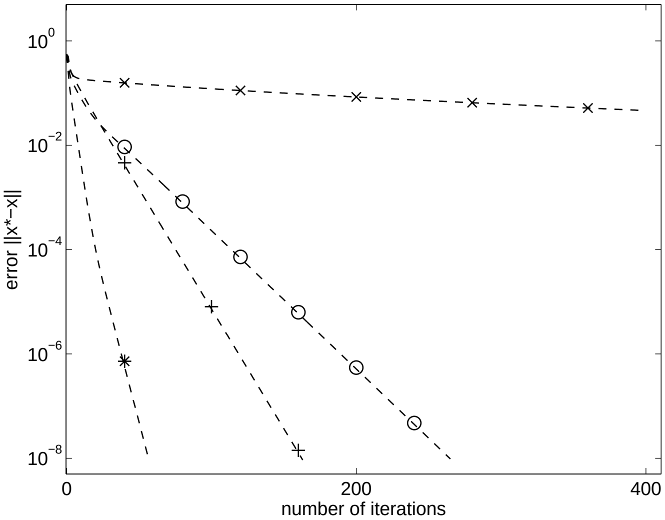

FIG. 4.1. *Depicted are the absolute errors* $\|\mathbf{x}^* - \mathbf{x}^{(n)}\|$ *between "exact" normalized factor matrices of a rank-r approximation of a random* $10 \times 10 \times 10$ *tensor and the iterates of the corresponding ALS algorithm. Plots are given for* $r = 2$ $(*)$, $r = 3$ $(\circ)$, $r = 4$ $(+)$ *and* $r = 5$ $(\times)$. *As one can see, the convergence rate is better for* $r = 4$ *than for* $r = 3$. *However, this behavior was rather exceptional in our experiments.*

Proof. Fix $\lambda_0 > 0$. All global minima $\mathbf{x}^*_{\lambda_0}$ of g_{λ_0} lie in a certain ball of radius depending on λ_0 (again since g_{λ_0} is coercive). Then for $\lambda > \lambda_0$ the global minima $\mathbf{x}^*_\lambda$ of g_λ also have to lie in that ball, since otherwise we would obtain the contradiction

$$g_\lambda(\mathbf{x}^*_{\lambda_0}) = g_{\lambda_0}(\mathbf{x}^*_{\lambda_0}) + (\lambda - \lambda_0)\|\mathbf{x}^*_{\lambda_0}\| < g_{\lambda_0}(\mathbf{x}^*_\lambda) + (\lambda - \lambda_0)\|\mathbf{x}^*_\lambda\| = g_\lambda(\mathbf{x}^*_\lambda).$$

Consequently, the Hessian $f''(\mathbf{x}^*_\lambda)$ can be bounded by a constant depending only on λ_0, so that

$$g''_\lambda(\mathbf{x}^*_\lambda) = f''(\mathbf{x}^*_\lambda) + 2\lambda I$$

will be positive definite if λ is large enough. □

Of course one does not want to make λ too large. At least, in contrast to the unregularized case, for every λ the heuristic mentioned in the previous section can be justified for global minima: since $\tau''(\mathbf{x}^*_\lambda)$ now only has to be bounded on the ball in which its global minimizers $\mathbf{x}^*_\lambda$ are located (which depends on λ), the Hessian

$$g''_\lambda(\mathbf{x}^*_\lambda)[\mathbf{h}, \mathbf{h}] = \|\tau'(\mathbf{x}^*_\lambda)\mathbf{h}\|^2 + (\tau(\mathbf{x}^*_\lambda) - \mathcal{T}, \tau''(\mathbf{x}^*_\lambda)[\mathbf{h}, \mathbf{h}]) + 2\lambda\|\mathbf{h}\|^2$$

will be positive definite, if $\tau(\mathbf{x}^*_\lambda) - \mathcal{T}$ is sufficiently small. In particular, for exact decompositions it is always positive definite.

4. Numerical experiments.

Our numerical experiments were quite simple and only meant to demonstrate the linear convergence rate and check the size of the convergence region. The calculations have neither been systematic nor exhaustive. Real applications can be found elsewhere. We emphasize again that we were interested in the local convergence of the factor matrices, that is, say in the order-3 case, we measured the Euclidian norm

$$\|\mathbf{x}^* - \mathbf{x}\| = \sqrt{\|\mathbf{A}^* - \mathbf{A}\|^2 + \|\mathbf{B}^* - \mathbf{B}\|^2 + \|\mathbf{C}^* - \mathbf{C}\|^2},$$

where the matrix norms are the Frobenius norms. To obtain a relative measure we normalized the columns of the factor matrices to one (instead of equilibrating them).

In our experiments we randomly generated tensors of different size and order and first calculated a “best” rank-r approximation using ALS with a random starting point. As a stopping criterion we imposed that the difference $\|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}\|$ between two subsequent normalized iterates should be sufficiently small ($\sim 10^{-14}$) in order “guarantee” that the solution is at least a critical point of the approximation problem. For large values of r we had to try many starting points to achieve this precision due to the appearance of swamps.

The normalized factor matrices of the calculated solutions then have been randomly perturbed by different orders of magnitude and then used as a starting point for the ALS algorithm again. We observed that the initial solution would be recovered if the perturbation was of magnitude at most 10^0. In fact, larger perturbations result in an almost random starting point, so the ALS iteration almost surely will converge to a different critical point. Usually, the convergence rate decreased for larger r, but not in all experiments. In Fig 4.1 we plotted the errors $\|\mathbf{x}^* - \mathbf{x}^{(n)}\|$ of the iteration for the rank $r = 2, 3, 4$ and 5 approximation of a random $10 \times 10 \times 10$ tensor. As one can see, the convergence is linear and the rate is better for $r = 4$ than for $r = 3$ in this example, but then again really slow for $r = 5$.

For special tensors, such as hyperdiagonal or, more generally, complete orthogonal tensors [8], one can exactly determine the best rank-r approximation by truncation of a complete orthogonal representation. Applying the same kind of experiment for such tensors we observed very fast convergence of ALS (at most 4 iterations independently of r), so we did not include plots for this case.

REFERENCES

[1] J.C. Bezdek, R.J. Hathaway, Convergence of alternating optimization, Neural Parallel Sci. Comput., 11 (2003), pp. 351–368.

[2] G. Beylkin, M. J. Mohlenkamp, Algorithms for numerical analysis in high dimensions, SIAM J. Sci. Comput. 26 (2005), pp. 2133–2159 (electronic).

[3] 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.

[4] W.H. Greub, Multilinear Algebra, Springer-Verlag, New York, 1967.

[5] L. Grippo, M. Sciandrone, Globally convergent block-coordinate techniques for unconstrained optimization, Optim. Methods Softw., 10 (1999), pp. 587–637.

[6] L. Grippo, M. Sciandrone, On the convergence of the block nonlinear Gauss-Seidel method under convex constraints, Oper. Res. Lett., 26 (2000), pp. 127–136.

[7] 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.

[8] T.G. Kolda, Orthogonal Tensor Decompositions, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243-255

[9] T.G. Kolda, B.W. Bader, Tensor Decompositions and Applications, SIAM Rev., 51 (2009), pp. 455–500.

[10] 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 .

[11] M.J. Mohlenkamp, Musings on Multilinear Fitting, to appear in Linear Algebra Appl.

[12] C. Navasca, L.D. Lathauwer and S. Kindermann, Swamp reducing technique for tensor decomposition, 16th Proceedings of the European Signal Processing Conference, Lausanne, 2008.

[13] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[14] I.V. Oseledets, A new tensor decomposition, Dokl. Math., 80 (2009), pp. 495–496; translation from Dokl. Akad. Nauk, Ross. Akad. Nauk, 427 (2009), pp. 168–169.

[15] P. Paatero, Construction and analysis of degenerate PARAFAC models, J. Chemometrics, 14 (2000), pp. 285–299.

[16] S. Schechter, Iteration Methods for Nonlinear Problems, Trans. Am. Math. Soc. 104 (1962), pp. 179–189.

[17] G. Tomasi, R. Bro, A comparison of algorithms for fitting the PARAFAC model, Comput. Statist. Data Anal., 50 (2006), pp. 1700–1734.

[18] 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.

[19] T. Zhang, G.H. Golub, Rank-one approximation to high order tensors, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.

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.