

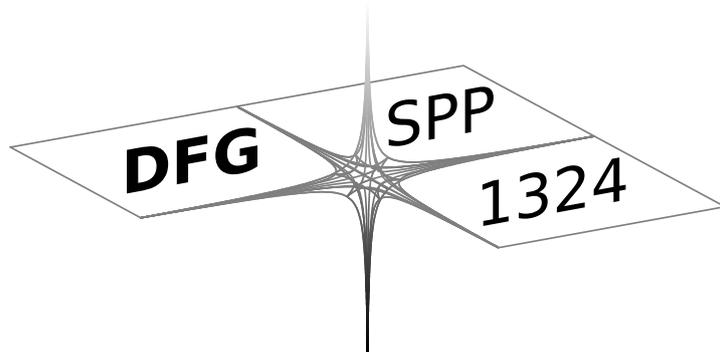
# DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen“

## Convergence of Alternating Least Squares Optimisation for Rank-One Approximation to High Order Tensors

M. Espig, A. Khachatryan

Preprint 173



Edited by

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

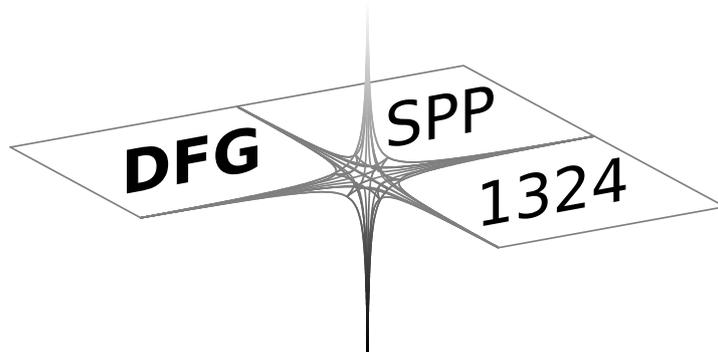
# DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen“

## Convergence of Alternating Least Squares Optimisation for Rank-One Approximation to High Order Tensors

M. Espig, A. Khachatryan

Preprint 173



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

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

# Convergence of Alternating Least Squares Optimisation for Rank-One Approximation to High Order Tensors

Mike Espig <sup>\*†</sup>

Aram Khachatryan <sup>\*</sup>

18th November 2014

## Abstract

The approximation of tensors has important applications in various disciplines, but it remains an extremely challenging task. It is well known that tensors of higher order can fail to have best low-rank approximations, but with an important exception that best rank-one approximations always exists. The most popular approach to low-rank approximation is the alternating least squares (ALS) method. The convergence of the alternating least squares algorithm for the rank-one approximation problem is analysed in this paper. In our analysis we are focusing on the global convergence and the rate of convergence of the ALS algorithm. It is shown that the ALS method can converge sublinearly, Q-linearly, and even Q-superlinearly. Our theoretical results are demonstrated on explicit examples.

**Keywords:** tensor format, tensor representation, tensor network, alternating least squares optimisation, orthogonal projection method.

## 1 Introduction

We consider a minimisation problem on the tensor space  $\mathcal{V} = \bigotimes_{\mu=1}^d \mathbb{R}^{n_\mu}$  equipped with the Euclidean inner product  $\langle \cdot, \cdot \rangle$ . The objective function  $f : \mathcal{V} \rightarrow \mathbb{R}$  of the optimisation task is quadratic

$$f(v) := \frac{1}{\|b\|^2} \left[ \frac{1}{2} \langle v, v \rangle - \langle b, v \rangle \right] \geq -\frac{1}{2}, \quad (1)$$

where  $b \in \mathcal{V}$ . In our analysis, a tensor  $u \in \mathcal{V}$  is represented as a rank-one tensor. The representation of rank-one tensors is described by the following multilinear map  $U$ :

$$U : P := \bigtimes_{\mu=1}^d \mathbb{R}^{n_\mu} \rightarrow \mathcal{V}$$
$$(p_1, \dots, p_d) \mapsto U(p_1, \dots, p_d) := \bigotimes_{\mu=1}^d p_\mu.$$

We call a  $d$ -tuple of vectors  $(p_1, \dots, p_d) \in P$  a representation system of  $u$  if  $u = U(p_1, \dots, p_d)$ . The tensor  $b$  is approximated with respect to rank-one tensors, i.e. we are looking for a representation system

---

<sup>\*</sup>RWTH Aachen University, Germany

<sup>†</sup>Address: RWTH Aachen University, Department of Mathematics, IGPM Templergraben 55, 52056 Aachen Germany. Phone: +49 (0)241 80 96343, E-mail address: mike.espig@alopax.de

$(p_1^*, \dots, p_d^*) \in P$  such that for

$$\begin{aligned} F &:= f \circ U : P \rightarrow \mathcal{V} \rightarrow \mathbb{R} \\ F(p_1, \dots, p_d) &= \frac{1}{\|b\|^2} \left[ \frac{1}{2} \langle U(p_1, \dots, p_d), U(p_1, \dots, p_d) \rangle - \langle b, U(p_1, \dots, p_d) \rangle \right] \end{aligned} \quad (2)$$

we have

$$F(p_1^*, \dots, p_d^*) = \min_{(p_1, \dots, p_d) \in P} F(p_1, \dots, p_d). \quad (3)$$

The range set  $U(P)$  is a closed in  $\mathcal{V}$ , see [6]. Therefore, the approximation problem is well defined. The set of best rank-one approximations of the tensor  $b$  is denoted by

$$\mathcal{M}_b := \{v \in U(P) : v \text{ is a best rank-one approximation of } b\}. \quad (4)$$

The alternating least squares (ALS) algorithm [2, 3, 4, 7, 8, 11, 12] is recursively defined. Suppose that the  $k$ -th iterate  $\underline{p}^k = (p_1^k, \dots, p_d^k)$  and the first  $\mu - 1$  components  $p_1^{k+1}, \dots, p_{\mu-1}^{k+1}$  of the  $(k + 1)$ -th iterate  $\underline{p}^{k+1}$  have been determined. The basic step of the ALS algorithm is to compute the minimum norm solution

$$p_\mu^{k+1} := \operatorname{argmin}_{q_\mu \in \mathbb{R}^{n_\mu}} F(p_1^{k+1}, \dots, p_{\mu-1}^{k+1}, q_\mu, p_{\mu+1}^k, \dots, p_d^k).$$

Thus, in order to obtain  $\underline{p}^{k+1}$  from  $\underline{p}^k$ , we have to solve successively  $L$  ordinary least squares problems.

The ALS algorithm is a nonlinear Gauss-Seidel method. The locale convergence of the nonlinear Gauss-Seidel method to a stationary point  $\underline{p}^* \in P$  follows from the convergence of the linear Gauss-Seidel method applied to the Hessian  $F''(\underline{p}^*)$  at the limit point  $\underline{p}^*$ . If the linear Gauss-Seidel method converges R-linear then there exists a neighbourhood  $B(\underline{p}^*)$  of  $\underline{p}^*$  such that for every initial guess  $\underline{p}^0 \in B(\underline{p}^*)$  the nonlinear Gauss-Seidel method converges R-linear with the same rate as the linear Gauss-Seidel method. We refer the reader to Ortega and Rheinboldt for a description of nonlinear Gauss-Seidel method [10, Section 7.4] and convergence analysis [10, Thm. 10.3.5, Thm. 10.3.4, and Thm. 10.1.3]. A representation system of a represented tensor is not unique, since the map  $U$  is multilinear. Consequently, the matrix  $F''(\underline{p}^*)$  is not positive definite. Therefore, convergence of the linear Gauss-Seidel method is in general not ensured. However, the convergence of the ALS method is discussed in [9, 13, 15, 16]. Recently, the convergence of the ALS method was analysed by means of Lojasiewicz gradient inequality, please see [14] for more details. The current analysis is not based on the mathematical techniques developed for the nonlinear Gauss-Seidel method neither on the theory of Lojasiewicz inequalities, but on the multilinearity of the map  $U$ .

**Notation 1.1** ( $\mathbb{N}_n$ ). The set  $\mathbb{N}_n$  of natural numbers smaller than  $n \in \mathbb{N}$  is denoted by

$$\mathbb{N}_n := \{j \in \mathbb{N} : 1 \leq j \leq n\}.$$

The precise analysis of the ALS method is a quite challenging task. Some of the difficulties of the theoretical understanding are explained following examples.

**Example 1.2.** The approximation of  $b \in \mathcal{V}$  by a tensor of rank one is considered, where

$$\begin{aligned} b &= \sum_{j=1}^r \lambda_j \underbrace{\bigotimes_{\mu=1}^d b_{j\mu}}_{b_j}, \quad \lambda_1 \geq \dots \geq \lambda_r > 0, \quad \|b_{j\mu}\| = 1, \\ B_\mu &:= (b_{j\mu} : 1 \leq j \leq r) \in \mathbb{R}^{m_\mu \times r} \quad (1 \leq \mu \leq d), \end{aligned} \quad (5)$$

and  $B_\mu^T B_\mu = \mathbf{Id}$ , see the example in [9, Section 4.3.5]. Let us further assume that  $v_k = p_1^k \otimes p_2^k \otimes \dots \otimes p_d^k$  is already determined. Corollary 2.4 leads to the recursion

$$p_1^{k+1} = \underbrace{\left[ \frac{1}{\|v_k\|^2} B_1 \text{diag} \left( \lambda_j^2 \prod_{\mu=2}^{d-1} \frac{\langle b_{j\mu}, p_\mu^k \rangle^2}{\|p_\mu^k\|^2} \right)_{j=1, \dots, r} B_1^T \right]}_{G_1(p_1^k, \dots, p_d^k) :=} p_1^k \quad (k \geq 2), \quad (6)$$

The linear map  $G_1(p_1^k, \dots, p_d^k) \in \mathbb{R}^{m_1 \times m_1}$  describes the first micro step  $p_1^k \otimes p_2^k \otimes \dots \otimes p_d^k \mapsto p_1^{k+1} \otimes p_2^k \otimes \dots \otimes p_d^k$  in the ALS algorithm. The iteration matrix  $G_1(p_1^k, \dots, p_d^k)$  is independent under rescaling of the representation system, i.e.  $G_1(\alpha_1 p_1, \dots, \alpha_d p_d) = G_1(p_1, \dots, p_d)$  for  $1 = \prod_{\mu=1}^d \alpha_\mu$ . Further, we can illustrate the difficulties of the ALS iteration in higher dimensions. For  $d = 2$ , the ALS method is given by the two power iterations

$$\begin{aligned} p_1^{k+1} &= \left[ \frac{1}{\|p_1^k\|^2 \|p_2^k\|^2} B_1 \text{diag} (\lambda_j^2)_{j=1, \dots, r} B_1^T \right] p_1^k, \\ p_2^{k+1} &= \left[ \frac{1}{\|p_1^{k+1}\|^2 \|p_2^k\|^2} B_2 \text{diag} (\lambda_j^2)_{j=1, \dots, r} B_2^T \right] p_2^k. \end{aligned}$$

Clearly, if the global minimum  $b_1$  is isolated, i.e.  $\lambda_1 > \lambda_2$ , then the ALS method converges to  $b_1$  provided that  $\langle v_0, b_1 \rangle \neq 0$ , where  $v_0 = p_1^0 \otimes p_2^0 \in \mathcal{V}$  is the initial guess. Further, we have linear convergence

$$\left| \tan \angle [b_{1\mu}, p_\mu^{k+1}] \right| \leq \left( \frac{\lambda_2}{\lambda_1} \right)^2 \left| \tan \angle [b_{1\mu}, p_\mu^k] \right| \quad (1 \leq \mu \leq 2).$$

Note that in this example the angle  $\angle [b_{1\mu}, p_\mu^k]$  is a more natural measure of the error than the usual distance  $\|b_{1\mu} - p_\mu^k\|$ . For  $d \geq 3$ , the factor  $\prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^k \rangle^2 / \|p_\mu^k\|^2$  from Eq. (6) describes the behaviour of the ALS iteration. Let  $1 \leq j^* \leq r$ . We say that a term  $b_{j^*}$  from Eq. (5) dominates at  $v_k = p_1^k \otimes \dots \otimes p_d^k$  if

$$d^{-2} \sqrt{\lambda_{j^*}^2} \langle b_{j^*\mu}, p_\mu^k \rangle^2 > d^{-2} \sqrt{\lambda_j^2} \langle b_{j\mu}, p_\mu^k \rangle^2 \quad (7)$$

for all  $j \in N_{j^*} := \{j \in \mathbb{N} : 1 \leq j \leq r \text{ and } j \neq j^*\}$  and all  $\mu \in \mathbb{N}_d$ . If  $b_{j^*}$  dominates at  $v_k$ , then the recursion formula (6) leads to

$$\left| \tan \angle [b_{j^*1}, p_1^{k+1}] \right| \leq \underbrace{\frac{\max_{j \in N_{j^*}} \left( \lambda_j \prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^k \rangle \right)^2}{\left( \lambda_{j^*} \prod_{\mu=2}^{d-1} \langle b_{j^*\mu}, p_\mu^k \rangle \right)^2}}_{< 1} \left| \tan \angle [b_{j^*1}, p_1^k] \right|, \quad (8)$$

i.e. the first component of the representation system  $p_1^{k+1}$  is turned towards the direction of  $b_{j^*1}$ . Note that for  $r = 2$  the bound for the convergence rate is sharp, i.e.

$$\left| \tan \angle [b_{j^*1}, p_1^{k+1}] \right| = \frac{\max_{j \in N_{j^*}} \left( \lambda_j \prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^k \rangle \right)^2}{\left( \lambda_{j^*} \prod_{\mu=2}^{d-1} \langle b_{j^*\mu}, p_\mu^k \rangle \right)^2} \left| \tan \angle [b_{j^*1}, p_1^k] \right| \quad (r = 2). \quad (9)$$

The inequality

$$\begin{aligned} d^{-2} \sqrt{\lambda_{j^*}^2} \langle b_{j^*1}, p_1^{k+1} \rangle^2 &= \frac{1}{\|v_k\|^4} \lambda_{j^*}^4 \prod_{\mu=2}^{d-1} \frac{\langle b_{j^*\mu}, p_\mu^k \rangle^4}{\|p_\mu^k\|^4} d^{-2} \sqrt{\lambda_{j^*}^2} \langle b_{j^*1}, p_1^k \rangle^2 \\ &> \frac{1}{\|v_k\|^4} \lambda_{j^*}^4 \prod_{\mu=2}^{d-1} \frac{\langle b_{j\mu}, p_\mu^k \rangle^4}{\|p_\mu^k\|^4} d^{-2} \sqrt{\lambda_j^2} \langle b_{j1}, p_1^k \rangle^2 = d^{-2} \sqrt{\lambda_j^2} \langle b_{j1}, p_1^{k+1} \rangle^2 \end{aligned}$$

shows that  $b_{j^*}$  also dominates at the successor  $p_1^{k+1} \otimes p_2^k \otimes \dots \otimes p_d^k$ . Further, we have for all  $j \in N_{j^*}$

$$\frac{d^{-2}\sqrt{\lambda_j^2} \langle b_{j,1}, p_1^{k+1} \rangle^2}{d^{-2}\sqrt{\lambda_{j^*}^2} \langle b_{j^*,1}, p_1^{k+1} \rangle^2} = \prod_{\mu=2}^{d-1} \left( \underbrace{\frac{d^{-2}\sqrt{\lambda_j^2} \langle b_{j\mu}, p_\mu^k \rangle^2}{d^{-2}\sqrt{\lambda_{j^*}^2} \langle b_{j^*\mu}, p_\mu^k \rangle^2}}_{<1} \right)^2 \frac{d^{-2}\sqrt{\lambda_j^2} \langle b_{j,1}, p_1^k \rangle^2}{d^{-2}\sqrt{\lambda_{j^*}^2} \langle b_{j^*,1}, p_1^k \rangle^2} < \frac{d^{-2}\sqrt{\lambda_j^2} \langle b_{j,1}, p_1^k \rangle^2}{d^{-2}\sqrt{\lambda_{j^*}^2} \langle b_{j^*,1}, p_1^k \rangle^2}.$$

By analogy for the following micro steps, we have

$$\frac{\max_{j \in N_{j^*}} \left( \lambda_j \prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^{k+1} \rangle \right)^2}{\left( \lambda_{j^*} \prod_{\mu=2}^{d-1} \langle b_{j^*\mu}, p_\mu^{k+1} \rangle \right)^2} < \frac{\max_{j \in N_{j^*}} \left( \lambda_j \prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^k \rangle \right)^2}{\left( \lambda_{j^*} \prod_{\mu=2}^{d-1} \langle b_{j^*\mu}, p_\mu^k \rangle \right)^2}.$$

Hence, the ALS iteration converges to  $b_{j^*}$ . Now it is easy to see that

$$\limsup_{k \rightarrow \infty} \left( \frac{\max_{j \in N_{j^*}} \left( \lambda_j \prod_{\mu=2}^{d-1} \langle b_{j\mu}, p_\mu^k \rangle \right)^2}{\left( \lambda_{j^*} \prod_{\mu=2}^{d-1} \langle b_{j^*\mu}, p_\mu^k \rangle \right)^2} \right) = 0.$$

Therefore, the tangent  $\tan \angle [b_{j^*\mu}, p_\mu^k]$  converges  $Q$ -superlinearly, i.e.

$$\left| \tan \angle [b_{j^*\mu}, p_\mu^k] \right| \xrightarrow[k \rightarrow \infty]{} 0 \quad (Q\text{-superlinearly}).$$

Furthermore, the ALS iteration converges faster for large  $d$ . Unfortunately, there is no guarantee that the global minimum  $b_1$  dominates at  $v_k$ . However, in this example it is more likely that a chosen initial guess dominates at the global minimum. For simplicity let us assume that  $r = 2$  and  $\lambda_1 > \lambda_2$ . see Eq. (5). Since the Tucker ranks of  $b$  are all equal to 2 and the condition from Eq. (7) does not depend on the norm of the vectors from the representation system, assume without loss of generality that for  $\mu \in \mathbb{N}_d$  the representation system of every initial guess has the following form:

$$p_\mu(\varphi_\mu) = \sin(\varphi_\mu) b_{\mu,2} + \cos(\varphi_\mu) b_{\mu,1}, \quad \left( \varphi_\mu \in \left[0, \frac{\pi}{2}\right], \|p_\mu(\varphi_\mu)\| = 1 \right).$$

If the global minimum dominates at the initial guess, we have for all  $\mu \in \mathbb{N}_d$

$$\begin{aligned} d^{-2}\sqrt{\lambda_1^2} \langle b_{1\mu}, p_\mu(\varphi_\mu) \rangle^2 &> d^{-2}\sqrt{\lambda_2^2} \langle b_{2\mu}, p_\mu(\varphi_\mu) \rangle^2 \\ \Leftrightarrow \tan(\varphi_\mu) &< \sqrt{\frac{\lambda_1}{\lambda_2}}. \end{aligned}$$

If we define the angle  $\varphi_{d,\mu}^* \in \left[0, \frac{\pi}{2}\right]$  such that

$$\tan(\varphi_{d,\mu}^*) = \sqrt{\frac{\lambda_1}{\lambda_2}},$$

then every initial guess with  $\varphi_\mu \in [0, \varphi_{d,\mu}^*)$  converges to the global minimum. Furthermore, we have

$$\tan(\varphi_{d,\mu}^*) > 1 \quad \Leftrightarrow \quad \varphi_{d,\mu}^* > \frac{\pi}{4},$$

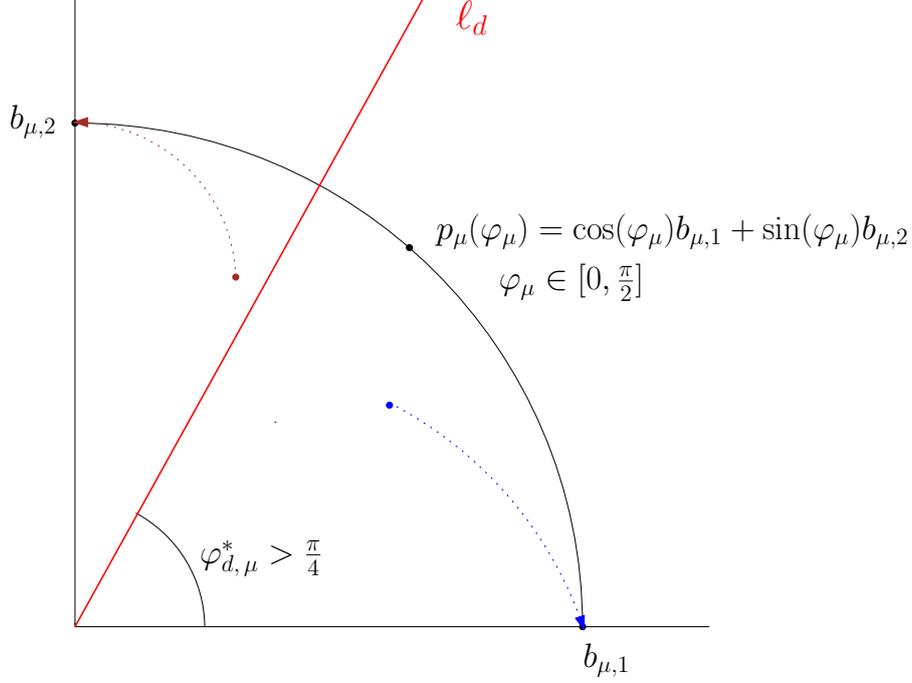


Figure 1: The angle  $\varphi_{d,\mu}^*$  describes the slice where the global minimum is a point of attraction. Every initial guess located under the red line  $\ell_d$  will converge to the global minimum. Note that the angle  $\varphi_{d,\mu}^*$  is larger than  $\frac{\pi}{4}$ , but interestingly enough  $\varphi_{d,\mu}^* \xrightarrow{d \rightarrow \infty} \frac{\pi}{4}$ .

*i.e. the slice where the global minimum is a point of attraction is more potent than the slice where the local minimum  $\lambda_2 b_2$  is a point of attraction, see Figure 1 for illustration. But we have for the asymptotic behavior*

$$\tan(\varphi_{d,\mu}^*) = \sqrt{\frac{\lambda_1}{\lambda_2}} \xrightarrow{d \rightarrow \infty} 1, \quad \Leftrightarrow \quad \varphi_{d,\mu}^* \xrightarrow{d \rightarrow \infty} \frac{\pi}{4},$$

*i.e. for sufficiently large  $d$  the slices are practically equal potent.*

**Example 1.3.** *In the following example a sublinear convergence of ALS procedure for rank-one approximation is shown. We will consider the tensor  $b_\lambda \in \mathcal{V}$  given by*

$$b_\lambda = \bigotimes_{\mu=1}^3 p + \lambda (p \otimes q \otimes q + q \otimes p \otimes q + q \otimes q \otimes p)$$

*for some  $\lambda \geq 0$  and  $p, q \in \mathbb{R}^n$  with  $\|p\| = \|q\| = 1$  and  $\langle p, q \rangle = 0$ . Let us first prove the following statement.*

**Proposition 1.4.** *Define  $v^* := \bigotimes_{\mu=1}^3 p$ . Then*

- a)  $\mathcal{M}_b = \{v^*\}$ , if  $\lambda \leq \frac{1}{2}$
- b)  $|\mathcal{M}_b| = 2$  and  $v^* \notin \mathcal{M}_b$ , if  $\lambda > \frac{1}{2}$

*Proof.* Let  $v_\lambda^* \in \mathcal{M}_b$ . Since tensor  $b$  is symmetric,  $v_\lambda^*$  also has to be symmetric. Write  $v_\lambda^* = C_\lambda \bigotimes_{\mu=1}^3 p_\lambda$ , where  $p_\lambda = p + \alpha_\lambda q$  (this is possible, since  $\langle b, q \otimes q \otimes q \rangle = 0$ ). Now the tuple  $(C_\lambda p_\lambda, p_\lambda, p_\lambda)$  is a stationary

point of  $F$ , therefore

$$(\mathbf{Id}_{\mathbb{R}^n} \otimes p_\lambda \otimes p_\lambda)^T b = Cp_\lambda$$

for some  $C \in \mathbb{R}$ . But

$$(\mathbf{Id}_{\mathbb{R}^n} \otimes p_\lambda \otimes p_\lambda)^T b = (1 + \lambda\alpha_\lambda^2)p + 2\lambda\alpha_\lambda q,$$

hence

$$\frac{2\lambda\alpha_\lambda}{1 + \lambda\alpha_\lambda^2} = \alpha_\lambda. \quad (10)$$

The solutions of (10) are

$$\alpha_\lambda = \begin{cases} 0, & \text{if } \lambda \leq \frac{1}{2}, \\ 0, \sqrt{\frac{2\lambda-1}{\lambda}} \text{ or } -\sqrt{\frac{2\lambda-1}{\lambda}}, & \text{if } \lambda > \frac{1}{2}. \end{cases}$$

Straightforward calculations show that for  $\lambda > \frac{1}{2}$  the solutions  $\alpha_\lambda = \pm\sqrt{\frac{2\lambda-1}{\lambda}}$  lead to the same value of  $F$  which is smaller than  $f(v^*)$ .  $\blacksquare$

Now let  $\lambda \leq \frac{1}{2}$  and  $v_k = C^k p_1^k \otimes p_2^k \otimes \dots \otimes p_d^k$ , with  $p_\mu^k = c_{\mu,k}^k p + s_{\mu,k}^k q$ ,  $c_{\mu,k}^2 + s_{\mu,k}^2 = 1$  and some  $C^k \in \mathbb{R}$ . Define  $\gamma_{\mu,k} := \begin{pmatrix} c_{\mu,k} \\ s_{\mu,k} \end{pmatrix}$ . Applying Corollary 2.4, one gets after short calculations the recursion formula

$$\gamma_{1,k+1} = C_{1,k} M_{1,k} M_{1,k}^T \gamma_{1,k}$$

with some  $C_{1,k} \in \mathbb{R}$  and

$$M_{1,k} = \begin{pmatrix} c_{2,k} & \lambda s_{2,k} \\ \lambda s_{2,k} & \lambda c_{2,k} \end{pmatrix}.$$

Then for  $t_{1,k} := \frac{s_{1,k}}{c_{1,k}}$  it holds

$$t_{1,k+1} = \frac{\lambda(\lambda+1)c_{2,k}c_{1,k} \frac{s_{2,k}}{s_{1,k}} + \lambda^2}{c_{2,k}^2 + \lambda^2 s_{2,k}^2 + \lambda(\lambda+1) \frac{c_{2,k}}{c_{1,k}} c_{2,k} s_{1,k}} \frac{s_{1,k}}{c_{1,k}}. \quad (11)$$

Thanks to Corollary 3.16 and Proposition 1.4 we know, that  $\lim_{k \rightarrow \infty} v^k = v^*$  for  $v^* = \bigotimes_{\mu=1}^3 p$ , therefore

$$\lim_{k \rightarrow \infty} c_{\mu,k} = 1 \quad (12)$$

$$\lim_{k \rightarrow \infty} s_{\mu,k} = 0 \quad (13)$$

for  $\mu \in \mathbb{N}_3$ . From Eq. 12 and 11 one gets

$$\limsup_{k \rightarrow \infty} \frac{t_{1,k+1}}{t_{1,k}} = \lambda^2 + \lambda(\lambda+1) \limsup_{k \rightarrow \infty} \frac{s_{2,k}}{s_{1,k}}. \quad (14)$$

The same way

$$\limsup_{k \rightarrow \infty} \frac{t_{2,k+1}}{t_{2,k}} = \lambda^2 + \lambda(\lambda+1) \limsup_{k \rightarrow \infty} \frac{s_{3,k}}{s_{2,k}} \quad (15)$$

$$\limsup_{k \rightarrow \infty} \frac{t_{3,k+1}}{t_{3,k}} = \lambda^2 + \lambda(\lambda+1) \limsup_{k \rightarrow \infty} \frac{s_{1,k+1}}{s_{3,k}} \quad (16)$$

Furthermore, from Eq. (21) we know that

$$p_{2,k+1} = C_{2,k+1} M_{2,k} p_{1,k+1}$$

with some  $C_{2,k+1} \in \mathbb{R}$  and

$$M_{2,k} = \begin{pmatrix} c_{3,k} & \lambda s_{3,k} \\ \lambda s_{3,k} & \lambda c_{3,k} \end{pmatrix}.$$

Simple calculations result in the relation

$$\frac{s_{2,k+1}}{s_{1,k+1}} = \lambda \frac{s_{3,k}}{s_{1,k+1}} c_{1,k+1} + \lambda c_{3,k},$$

and hence

$$\limsup_{k \rightarrow \infty} \frac{s_{2,k}}{s_{1,k}} = \lambda + \lambda \limsup_{k \rightarrow \infty} \frac{s_{3,k}}{s_{1,k+1}} \quad (17)$$

Now let  $\lambda = \frac{1}{2}$ . If  $\limsup_{k \rightarrow \infty} \frac{s_{2,k}}{s_{1,k}} \geq 1$ , then from Eq. (14) follows  $\limsup_{k \rightarrow \infty} \frac{t_{1,k+1}}{t_{1,k}} \geq 1$ , hence the convergence of  $p_{1,k}$  to  $p$  can not be  $Q$ -linearly. If  $\limsup_{k \rightarrow \infty} \frac{s_{2,k}}{s_{1,k}} < 1$ , then from Eq. (17)  $\limsup_{k \rightarrow \infty} \frac{s_{1,k+1}}{s_{3,k}} \geq 1$ , so from Eq. (15)  $\limsup_{k \rightarrow \infty} \frac{t_{3,k+1}}{t_{3,k}} \geq 1$ .

**Remark 1.5.**

a) In fact for  $\lambda = \frac{1}{2}$  it holds

$$\limsup_{k \rightarrow \infty} \frac{s_{2,k}}{s_{1,k}} = \limsup_{k \rightarrow \infty} \frac{s_{3,k}}{s_{2,k}} = \limsup_{k \rightarrow \infty} \frac{s_{1,k+1}}{s_{3,k}} = 1.$$

b) For  $\lambda < \frac{1}{2}$  ALS converges  $q$ -linearly with the convergence rate

$$\rho = \frac{\lambda}{2} \left( 3\lambda + \lambda^2 + \sqrt{(3\lambda + \lambda^2)^2 + 4\lambda} \right).$$

c) The example can be extended to higher dimensions in the following way. Let

$$b_\lambda = \bigotimes_{\mu=1}^d p + \lambda \sum_{\mu=1}^d \left( \bigotimes_{\nu=1}^{\mu-1} q \otimes p \otimes \bigotimes_{\nu=\mu+1}^d q \right)$$

with  $\|p\| = \|q\|$  and  $\langle p, q \rangle = 0$ . Then  $v^* = \bigotimes_{\mu=1}^d p$  is the unique best rank-one approximation of  $b_\lambda$  if and only if  $\lambda \leq \frac{1}{d-1}$ . Furthermore, ALS converges sublinear for  $\lambda = \frac{1}{d-1}$  and  $Q$ -linear for  $\lambda < \frac{1}{d-1}$ .

Our new convergence results are not obtained by using conventional technics like for the analysis of nonlinear Gauss-Seidel method or the theory of Lojasiewicz inequalities. Therefore, a detailed convergence approach is necessary.

## 2 The Alternating Least Squares Algorithm

In the following section, we recall the ALS algorithm. Where the algorithmic description of the ALS method is given in Algorithm 1.

**Notation 2.1** ( $L(A, B)$ ,  $P_{\nu, \mu}$ ). Let  $A, B$  be two arbitrary vector spaces. The vector space of linear maps from  $A$  to  $B$  is denoted by

$$L(A, B) := \{M : A \rightarrow B : M \text{ is linear}\}.$$

Let  $\mu, \nu \in \mathbb{N}_d$  with  $\nu \neq \mu$ . We define

$$P_{\nu, \mu} := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_{\nu-1}} \times \mathbb{R}^{n_{\nu+1}} \times \cdots \times \mathbb{R}^{n_{\mu-1}} \times \mathbb{R}^{n_{\mu+1}} \times \cdots \times \mathbb{R}^{n_d}.$$

---

**Algorithm 1** Alternating Least Squares (ALS) Algorithm
 

---

- 1: Set  $k := 1$  and choose an initial guess  $\underline{p}_1 = (p_1^1, \dots, p_d^1) \in P$ ,  $\underline{p}_{1,0} := \underline{p}_1$ , and  $v_1 := U(\underline{p}_1) \neq 0$ .
- 2: **while** Stop Condition **do**
- 3:    $v_{k,0} := v_k$
- 4:   **for**  $1 \leq \mu \leq d$  **do**
- 5:

$$\begin{aligned}
 p_\mu^{k+1} &:= \left( \frac{p_1^{k+1}}{\|p_1^{k+1}\|^2} \otimes \dots \otimes \frac{p_{\mu-1}^{k+1}}{\|p_{\mu-1}^{k+1}\|^2} \otimes \mathbf{Id}_{\mathbb{R}^{n_\mu}} \otimes \frac{p_{\mu+1}^k}{\|p_{\mu+1}^k\|^2} \otimes \dots \otimes \frac{p_d^k}{\|p_d^k\|^2} \right)^T b \quad (18) \\
 \underline{p}_{k,\mu+1} &:= (p_1^{k+1}, \dots, p_{\mu-1}^{k+1}, p_\mu^{k+1}, p_{\mu+1}^k, \dots, p_d^k) \\
 v_{k,\mu+1} &:= U(\underline{p}_{k,\mu+1})
 \end{aligned}$$

- 6:   **end for**
  - 7:    $\underline{p}_{k+1} := \underline{p}_{k,L}$  and  $v_{k+1} := U(\underline{p}_{k+1})$
  - 8:    $k \mapsto k + 1$
  - 9: **end while**
- 

The following map  $M_{\mu,\nu}$  from Lemma 2.2 is important for the analytical understanding of the ALS algorithm. As Corollary 2.4 shows, the map  $M_{\mu,\mu-1}$  describes an micro step of the ALS algorithm. Furthermore, there is an interesting relation between the map  $M_{\mu,\nu}$  and rank-one best approximations of the tensor  $b$ , see Theorem 2.10.

**Lemma 2.2.** *Let  $\mu, \nu \in \mathbb{N}_d$ ,  $\nu \neq \mu$ , and  $\underline{p}_{\nu,\mu} = (p_1, \dots, p_{\nu-1}, p_{\nu+1}, \dots, p_{\mu-1}, p_{\mu+1}, \dots, p_d) \in P_{\nu,\mu}$ . There exists a multilinear map  $M_{\nu,\mu} : P_{\nu,\mu} \times \mathcal{V} \rightarrow L(\mathbb{R}^{n_\nu}, \mathbb{R}^{n_\mu})$  such that*

$$M_{\nu,\mu}(\underline{p}_{\nu,\mu}, b)\mathbf{g}_\nu = (p_1 \otimes \dots \otimes p_{\nu-1} \otimes \mathbf{g}_\nu \otimes p_{\nu+1} \otimes \dots \otimes p_{\mu-1} \otimes \mathbf{Id}_{\mathbb{R}^{n_\mu}} \otimes p_{\mu+1} \otimes \dots \otimes p_d)^T b \quad (19)$$

for all  $\mathbf{g}_\nu \in \mathbb{R}^{n_\nu}$ . Further, we have  $M_{\mu,\nu}(\underline{p}_{\nu,\mu}, b) = M_{\nu,\mu}^T(\underline{p}_{\nu,\mu}, b)$ .

*Proof.* Follows directly from the multilinearity of the tensor product and elementary calculations. ■

**Example 2.3.** *Let  $\mu, \nu \in \mathbb{N}_d$ ,  $\nu \neq \mu$ ,  $\underline{p}_{\nu,\mu} = (p_1, \dots, p_{\nu-1}, p_{\nu+1}, \dots, p_{\mu-1}, p_{\mu+1}, \dots, p_d) \in P_{\nu,\mu}$ , and  $b$  be given in a subspace decomposition, i.e.*

$$b = \sum_{i_1=1}^{t_1} \dots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \bigotimes_{\mu=1}^d b_{\mu, i_\mu} \quad (t_\mu \in \mathbb{N}_{n_\mu})$$

A matrix representation of the linear map  $M_{\nu,\mu}$  is given by

$$\begin{aligned}
 M_{\nu,\mu}(\underline{p}_{\nu,\mu}, b) &= \sum_{i_1=1}^{t_1} \dots \sum_{i_\nu=1}^{t_\nu} \dots \sum_{i_\mu=1}^{t_\mu} \dots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \prod_{\xi \in \mathbb{N}_d \setminus \{\mu, \nu\}} \langle b_{\xi, i_\xi}, p_\xi \rangle b_{\mu, i_\mu} b_{\nu, i_\nu}^T \\
 &= B_\mu \Gamma(\underline{p}_{\nu,\mu}) B_\nu^T,
 \end{aligned}$$

where  $B_\xi = (b_{\xi,1}, \dots, b_{\xi, t_\xi}) \in \mathbb{R}^{n_\xi \times t_\xi}$  for all  $\xi \in \{\mu, \nu\}$  and the entries of the matrix  $\Gamma(\underline{p}_{\nu,\mu})$  are defined by

$$[\Gamma(\underline{p}_{\nu,\mu})]_{(i_\nu, i_\mu)} = \sum_{i_1=1}^{t_1} \dots \sum_{i_{\nu-1}=1}^{t_{\nu-1}} \dots \sum_{i_{\nu+1}=1}^{t_{\nu+1}} \dots \sum_{i_{\mu-1}=1}^{t_{\mu-1}} \dots \sum_{i_{\mu+1}=1}^{t_{\mu+1}} \dots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \prod_{\xi \in \mathbb{N}_d \setminus \{\mu, \nu\}} \langle b_{\xi, i_\xi}, p_\xi \rangle.$$

**Corollary 2.4.** Let  $\mu \in \mathbb{N}_d$ ,  $k \geq 2$ , and  $\underline{p}_{k,\mu} = (p_1^{k+1}, \dots, p_{\mu-1}^{k+1}, p_\mu^k, p_{\mu+1}^k, \dots, p_d^k) \in P$  from Algorithm 1. With the matrix from Lemma 2.2, the following recursion formula holds:

$$p_\mu^{k+1} = \frac{1}{G_{k,\mu} G_{k,\mu-1}} M_{\mu,k} M_{\mu,k}^T p_\mu^k, \quad (20)$$

where

$$\begin{aligned} G_{k,\mu} &:= \prod_{\nu=1}^{\mu-1} \|p_\nu^{k+1}\|^2 \prod_{\nu=\mu+1}^d \|p_\nu^k\|^2 \\ G_{k,\mu-1} &:= \prod_{\nu=1}^{\mu-2} \|p_\nu^{k+1}\|^2 \prod_{\nu=\mu}^d \|p_\nu^k\|^2, \\ M_{\mu,k} &:= M_{\mu,\mu-1}(p_1^{k+1}, \dots, p_{\mu-2}^{k+1}, p_{\mu+1}^k, \dots, p_d^k, b). \end{aligned}$$

*Proof.* We have with Eq. (18) and Lemma 2.2

$$p_\mu^{k+1} = \frac{1}{G_{k,\mu}} M_{\mu,\mu-1}(p_1^{k+1}, \dots, p_{\mu-2}^{k+1}, p_{\mu+1}^k, \dots, p_d^k, b) p_{\mu-1}^{k+1}, \quad (21)$$

$$p_{\mu-1}^{k+1} = \frac{1}{G_{k,\mu-1}} M_{\mu,\mu-1}^T(p_1^{k+1}, \dots, p_{\mu-2}^{k+1}, p_{\mu+1}^k, \dots, p_d^k, b) p_\mu^k. \quad (22)$$

■

**Example 2.5.** Let  $v_k = p_1^k \otimes p_2^k \otimes \dots \otimes p_d^k$  and

$$b = \sum_{i_1=1}^{t_1} \dots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \bigotimes_{\mu=1}^d b_{\mu, i_\mu},$$

i.e. the tensor  $b$  is given in the Tucker decomposition. From Eq. (18) it follows

$$\begin{aligned} p_1^{k+1} &= \frac{1}{\prod_{\mu=2}^d \|p_\mu^k\|^2} \sum_{i_1=1}^{t_1} \dots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \prod_{\mu=2}^d \langle b_{\mu, i_\mu}, p_\mu^k \rangle b_{1, i_1} \\ &= \frac{1}{\prod_{\mu=2}^{d-1} \|p_\mu^k\| \|p_d^k\|^2} \left[ \sum_{i_1=1}^{t_1} \sum_{i_d=1}^{t_d} b_{1, i_1} \sum_{i_2=1}^{t_2} \dots \sum_{i_{d-1}=1}^{t_{d-1}} \beta_{(i_1, \dots, i_d)} \prod_{\mu=2}^{d-1} \frac{\langle b_{\mu, i_\mu}, p_\mu^k \rangle}{\|p_\mu^k\|} b_{d, i_d}^T \right] p_d^k \\ &= \frac{1}{\prod_{\mu=2}^{d-1} \|p_\mu^k\| \|p_d^k\|^2} B_1 \Gamma_{1,k} B_d^T p_d^k, \end{aligned}$$

where  $B_\mu = (b_{\mu, i_\mu} : 1 \leq i_\mu \leq t_\mu) \in \mathbb{R}^{n_\mu \times t_\mu}$ ,  $B_\mu^T B_\mu = \mathbf{I}_{\mathbb{R}^{t_\mu}}$ , and the entries of the matrix  $\Gamma_{1,k} \in \mathbb{R}^{t_1 \times t_d}$  are defined by

$$[\Gamma_{1,k}]_{i_1, i_d} = \sum_{i_2=1}^{t_2} \dots \sum_{i_{d-1}=1}^{t_{d-1}} \beta_{(i_1, \dots, i_d)} \prod_{\mu=2}^{d-1} \frac{\langle b_{\mu, i_\mu}, p_\mu^k \rangle}{\|p_\mu^k\|} \quad (1 \leq i_1 \leq t_1, 1 \leq i_d \leq t_d).$$

Note that  $\Gamma_{1,k}$  is a diagonal matrix if the coefficient tensor  $\beta \in \bigotimes_{\mu=1}^d \mathbb{R}^{t_\mu}$  is super-diagonal, see Eq. (6). For  $p_d^k$  it follows further

$$p_d^k = \frac{1}{\prod_{\mu=1}^{d-1} \|p_\mu^k\|^2} \sum_{i_1=1}^{t_1} \cdots \sum_{i_d=1}^{t_d} \beta_{(i_1, \dots, i_d)} \prod_{\mu=1}^{d-1} \langle b_{\mu, i_\mu}, p_\mu^k \rangle b_{d, i_d} = \frac{1}{\|p_1^k\|^2 \prod_{\mu=2}^{d-1} \|p_\mu^k\|} B_d \Gamma_{1,k}^T B_1^T p_1^k$$

and finally

$$p_1^{k+1} = \frac{1}{\prod_{\mu=1}^d \|p_\mu^k\|^2} B_1 \Gamma_{1,k} \Gamma_{1,k}^T B_1^T p_1^k.$$

Let  $v^* = \lambda p_1 \otimes \dots \otimes p_d \in \mathcal{M}_b$  be a rank-one best approximation of  $b$ . Without loss of generality we can assume that

$$\|p_1\| = \|p_2\| = \dots = \|p_d\| = 1 \text{ and } \|v^*\| = \lambda.$$

Further, let  $\mu, \nu \in \mathbb{N}_d$  and

$$\underline{p}_{\nu, \mu} := (p_1, \dots, p_{\nu-1}, p_{\nu+1}, \dots, p_{\mu-1}, p_{\mu+1}, \dots, p_d) \in P_{\nu, \mu}.$$

The following two maps are of interest for our analysis:

$$\bar{V} : S^{n_\nu-1} \times S^{n_\mu-1} \rightarrow \mathcal{V}$$

$$(g_\nu, g_\mu) \mapsto \bar{V}(g_\nu, g_\mu) := p_1 \otimes \dots \otimes p_{\nu-1} \otimes g_\nu \otimes p_{\nu+1} \otimes \dots \otimes p_{\mu-1} \otimes g_\mu \otimes p_{\mu+1} \otimes \dots \otimes p_d$$

and

$$\bar{U} : S^{n_\nu-1} \times S^{n_\mu-1} \rightarrow \mathcal{V}$$

$$(g_\nu, g_\mu) \mapsto \bar{U}(g_\nu, g_\mu) := \langle \bar{V}(g_\nu, g_\mu), b \rangle \bar{V}(g_\nu, g_\mu),$$

where  $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$  denotes the sphere in  $\mathbb{R}^n$ .

**Lemma 2.6.** Let  $\mu, \nu \in \mathbb{N}_d$ ,  $g_\nu \in S^{n_\nu-1}$  and  $g_\mu \in S^{n_\mu-1}$ . We have

$$-2f(\bar{U}(g_\nu, g_\mu)) = \left\langle \underbrace{\left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right)}_{\in L(\mathbb{R}^{n_\nu}, \mathbb{R}^{n_\mu})} g_\nu, g_\mu \right\rangle^2 = \langle \bar{U}(g_\nu, g_\mu), b \rangle = \|\bar{U}(g_\nu, g_\mu)\|^2.$$

*Proof.* Let  $g_\nu \in S^{n_\nu-1}$ ,  $g_\mu \in S^{n_\mu-1}$ , and define  $\pi(g_\nu, g_\mu) := \bar{V}(g_\nu, g_\mu)(\bar{V}(g_\nu, g_\mu))^T$ . It holds  $\bar{U}(g_\nu, g_\mu) = \pi(g_\nu, g_\mu)b$  and

$$\begin{aligned} f(\bar{U}(g_\nu, g_\mu)) &= \frac{1}{2} \langle \bar{U}(g_\nu, g_\mu), \bar{U}(g_\nu, g_\mu) \rangle - \langle \bar{U}(g_\nu, g_\mu), b \rangle = \frac{1}{2} \langle \pi^2(g_\nu, g_\mu)b, b \rangle - \langle \pi(g_\nu, g_\mu)b, b \rangle \\ &= \frac{1}{2} \langle \pi(g_\nu, g_\mu)b, b \rangle - \langle \pi(g_\nu, g_\mu)b, b \rangle = -\frac{1}{2} \langle \pi(g_\nu, g_\mu)b, b \rangle = -\frac{1}{2} \langle \bar{U}(g_\nu, g_\mu), b \rangle \\ &= -\frac{1}{2} \langle \pi^2(g_\nu, g_\mu)b, b \rangle = -\frac{1}{2} \langle \pi(g_\nu, g_\mu)b, \pi(g_\nu, g_\mu)b \rangle = -\frac{1}{2} \|\bar{U}(g_\nu, g_\mu)\|^2 \\ &= -\frac{1}{2} \langle \bar{V}(g_\nu, g_\mu), b \rangle^2 = -\frac{1}{2} \left\langle \left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right) g_\nu, g_\mu \right\rangle^2. \end{aligned}$$

■

**Remark 2.7.** Obviously, the minimisation problem from Eg. (3) is equivalent to the following constrained maximisation problem: Find  $\tilde{v} = \bigotimes_{\mu=1}^d p_\mu$  such that for all  $\mu \in \mathbb{N}_d$  it holds

$$\langle \tilde{v}, b \rangle = \max_{v \in U(P)} \langle v, b \rangle \quad \text{subject to } \|p_\mu\| = 1.$$

Lagrangian method for constrained optimisation leads us to the following Lagrangian

$$L_{\underline{\lambda}}(q_1, \dots, q_d) = \langle U(q_1, \dots, q_d), b \rangle + \frac{1}{2} \sum_{\mu=1}^d \lambda_\mu (1 - \|q_\mu\|^2),$$

where  $q_\mu \in \mathbb{R}^{n_\mu}$  and  $\underline{\lambda} = (\lambda_1, \dots, \lambda_d)^T \in \mathbb{R}^d$  is the vector of Lagrange multipliers. A rank-one best approximation  $v^* = \lambda p_1 \otimes \dots \otimes p_d \in \mathcal{M}_b$  with  $\lambda \in \mathbb{R}$  and  $\|p_\mu\| = 1$  satisfies

$$\begin{aligned} \frac{\partial}{\partial p_\mu} L_{\underline{\lambda}}(p_1, \dots, p_d) &= (p_1 \otimes \dots \otimes p_{\mu-1} \otimes \mathbf{Id}_{\mathbb{R}^{n_\mu}} \otimes p_{\mu+1} \otimes \dots \otimes p_d)^T b - \lambda_\mu^* p_\mu = 0, \\ \frac{\partial}{\partial \lambda_\mu^*} L_{\underline{\lambda}}(p_1, \dots, p_d) &= \frac{1}{2} (1 - \|p_\mu\|^2) = 0. \end{aligned}$$

For  $\nu \in \mathbb{N}_d \setminus \{\mu\}$  it follows that

$$\lambda = \langle p_1 \otimes \dots \otimes p_d, b \rangle, \quad \lambda p_\mu = M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) p_\nu, \quad \lambda p_\nu = M_{\nu, \mu}^T(\underline{p}_{\nu, \mu}, b) p_\mu,$$

where  $\underline{p}_{\nu, \mu} \in P_{\nu, \mu}$  is like in Lemma 2.2. Therefore,  $\lambda$  is a singular value of the matrix  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$  and  $p_\nu, p_\mu$  are the associated singular vectors.

**Proposition 2.8.** Let  $v^* = \lambda p_1 \otimes \dots \otimes p_d \in \mathcal{M}_b$  a best approximation of  $b$  with  $\|p_1\| = \dots = \|p_d\| = 1$ . We have

$$f(v^*) = -\frac{1}{2 \|b\|^2} \|v^*\|^2 = -\frac{1}{2 \|b\|^2} \langle b, v^* \rangle.$$

*Proof.* Since  $v^* \in \mathcal{M}_b$  we have that  $v^* = \Pi b$ , where  $\Pi := \frac{v^* v^{*T}}{\|v^*\|^2}$ . Furthermore, it holds

$$\langle v^*, v^* \rangle = \langle \Pi b, v^* \rangle = \langle b, \Pi v^* \rangle = \langle b, v^* \rangle.$$

The rest follows from the definition of  $f$ , see Eq. (1). ■

**Remark 2.9.** From Proposition 2.8 it follows instantly that the global minimum of the best approximation problem from Eq. (3) has the largest norm among all other  $\tilde{v} \in \mathcal{M}_b$ .

**Theorem 2.10.** Let  $\mu, \nu \in \mathbb{N}_d$  and  $v^* = \|v^*\| p_1 \otimes \dots \otimes p_d \in \mathcal{M}_b$  be a rank-one best approximation of  $b$  with  $\|p_1\| = \dots = \|p_d\| = 1$ . Then  $\|v^*\|$  is the largest singular value of  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$  and  $p_\nu, p_\mu$  are the associated singular vectors. Furthermore, if  $v^*$  is isolated, then  $\|v^*\|$  is a simple singular value of  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$ .

*Proof.* Let  $\mu, \nu \in \mathbb{N}_d$ . From Lemma 2.6 and Remark 2.7 it follows that  $\|v^*\|$  is a singular value of  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$  and  $p_\nu, p_\mu$  are associated singular vectors. Assume that there is a singular value  $\tilde{\lambda}$  of  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$  and associated singular vectors  $q_\nu \in \mathbb{R}^{n_\nu}, q_\mu \in \mathbb{R}^{n_\mu}$  with  $\tilde{\lambda} > \|v^*\|$ . Let  $\alpha \in [0, 1]$  and  $\beta \in (0, 1]$  with  $\alpha^2 + \beta^2 = 1$ . Define further  $g_\nu(\alpha, \beta) := g_\nu := \alpha p_\nu + \beta q_\nu \in \mathbb{R}^{n_\nu}$  and  $g_\mu(\alpha, \beta) := g_\mu := \alpha p_\mu + \beta q_\mu \in \mathbb{R}^{n_\mu}$ . We have  $\|g_\nu\|^2 = \|g_\mu\|^2 = \alpha^2 + \beta^2 = 1$  and with Lemma 2.6 it follows then

$$\begin{aligned} -2f(\bar{U}(g_\nu, g_\mu)) &= \left\langle \left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right) g_\nu, g_\mu \right\rangle^2 = \left\langle \left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right) \alpha p_\nu + \beta q_\nu, \alpha p_\mu + \beta q_\mu \right\rangle^2 \\ &= \left\langle \alpha \|v^*\| p_\nu + \beta \tilde{\lambda} q_\nu, \alpha p_\mu + \beta q_\mu \right\rangle^2 = \left( \alpha^2 \|v^*\| + \beta^2 \tilde{\lambda} \right)^2 \\ &\stackrel{(\beta \neq 0)}{>} \left( \alpha^2 \|v^*\| + \beta^2 \|v^*\| \right)^2 = \|v^*\|^2 = \left\langle \left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right) p_\nu, p_\mu \right\rangle^2 = -2f(v^*). \end{aligned}$$

Consequently, it is

$$f(\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))) < f(v^*) \quad \text{for all } \alpha \in [0, 1] \text{ and } \beta \in (0, 1] \text{ with } \alpha^2 + \beta^2 = 1,$$

i.e. we can find a better approximation  $\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))$  of  $b$  which is arbitrary close to  $v^*$ . This contradicts the fact that  $v^* \in \mathcal{M}_b$ .

Additionally, let  $v^*$  be a isolated rank-one best approximation of  $b$ . Assume that there is a singular value  $\lambda$  of  $M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b)$  and associated singular vectors  $q_\nu \in \mathbb{R}^{n_\nu}$ ,  $q_\mu \in \mathbb{R}^{n_\mu}$  with  $\lambda = \|v^*\|$ ,  $p_\nu \perp q_\nu$ , and  $p_\mu \perp q_\mu$ . Almost like above, let  $\alpha, \beta \in [0, 1]$  with  $\alpha^2 + \beta^2 = 1$  and consider again  $g_\nu(\alpha, \beta) = \alpha p_\nu + \beta q_\nu \in \mathbb{R}^{n_\nu}$ ,  $g_\mu(\alpha, \beta) = \alpha p_\mu + \beta q_\mu \in \mathbb{R}^{n_\mu}$ . With Lemma 2.6 it follows

$$-2f(\bar{U}(g_\nu, g_\mu)) = (\alpha^2 \|v^*\| + \beta^2 \lambda)^2 = \|v^*\|^2 = \left\langle \left( M_{\nu, \mu}(\underline{p}_{\nu, \mu}, b) \right) p_\nu, p_\mu \right\rangle^2 = -2f(v^*),$$

i.e. we have

$$f(\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))) = f(v^*) \quad \text{for all } \alpha, \beta \in [0, 1] \text{ with } \alpha^2 + \beta^2 = 1.$$

Therefore, we can find a approximation  $\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))$  of  $b$  which is arbitrary close to  $v^*$  and  $f(\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))) = f(v^*)$ . This contradicts the fact that  $v^*$  is isolated.  $\blacksquare$

**Remark 2.11.** *The proof of Theorem 2.10 shows that if we have two different best approximations of  $b$  which differ only in two arbitrary components of the representation systems and  $f(v^*) = f(v^{**})$ , then there is a complete path between  $v^*$  and  $v^{**}$  described by  $\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta))$  such that  $f(v^*) = f(\bar{U}(g_\nu(\alpha, \beta), g_\mu(\alpha, \beta)))$ .*

### 3 Convergence Analysis

In the following, we are using the notations and definitions from Section 2. Our convergence analysis is mainly based on the recursion introduced in Corollary 2.4 and the following Lemma 3.1.

**Lemma 3.1.** *Let  $k \in \mathbb{N}$ ,  $\mu \in \mathbb{N}$ , and  $v_{k, \mu} = p_1^{k+1} \otimes \dots \otimes p_{\mu-1}^{k+1} \otimes p_\mu^k \otimes \dots \otimes p_d^k$  from Algorithm 1. Then*

$$\Pi_{k, \mu} := \frac{p_1^{k+1} (p_1^{k+1})^T}{\|p_1^{k+1}\|^2} \otimes \dots \otimes \frac{p_{\mu-1}^{k+1} (p_{\mu-1}^{k+1})^T}{\|p_{\mu-1}^{k+1}\|^2} \otimes \mathbf{Id}_{\mathbb{R}^{n_\mu}} \otimes \frac{p_{\mu+1}^k (p_{\mu+1}^k)^T}{\|p_{\mu+1}^k\|^2} \otimes \dots \otimes \frac{p_d^{k+1} (p_d^{k+1})^T}{\|p_d^{k+1}\|^2}$$

is a orthogonal projection and

$$v_{k, \mu+1} = v_{k, \mu} + \Pi_{k, \mu} r_{k, \mu},$$

where  $r_{k, \mu} := b - v_{k, \mu}$ .

*Proof.* Obviously,  $\Pi_{k, \mu}$  is a orthogonal projection. Straightforward calculations show that  $v_{k, \mu} = \Pi_{k, \mu} v_{k, \mu}$  and  $v_{k, \mu+1} = \Pi_{k, \mu} b$ . Hence we have  $v_{k, \mu} + \Pi_{k, \mu} r_{k, \mu} = \Pi_{k, \mu} b = v_{k, \mu+1}$ .  $\blacksquare$

**Lemma 3.2.** *Let  $k \in \mathbb{N}$ ,  $\mu \in \mathbb{N}_L$ . We have*

$$f(v_{k, \mu}) - f(v_{k, \mu+1}) = \frac{1}{2} \frac{\langle \Pi_{k, \mu} r_{k, \mu}, r_{k, \mu} \rangle}{\|b\|^2} \quad (23)$$

*Proof.* It follows with Lemma 3.1 that

$$\begin{aligned}
f(v_{k,\mu+1}) &= \frac{1}{\|b\|^2} \left[ \frac{1}{2} \langle v_{k,\mu} + \Pi_{k,\mu} r_{k,\mu}, v_{k,\mu} + \Pi_{k,\mu} r_{k,\mu} \rangle - \langle b, v_{k,\mu} + \Pi_{k,\mu} r_{k,\mu} \rangle \right] \\
&= f(v_{k,\mu}) + \frac{1}{\|b\|^2} \left[ \frac{1}{2} \langle \Pi_{k,\mu} r_{k,\mu}, \Pi_{k,\mu} r_{k,\mu} \rangle + \langle v_{k,\mu}, \Pi_{k,\mu} r_{k,\mu} \rangle - \langle b, \Pi_{k,\mu} r_{k,\mu} \rangle \right] \\
&= f(v_{k,\mu}) + \frac{1}{\|b\|^2} \left[ \frac{1}{2} \langle r_{k,\mu}, \Pi_{k,\mu} r_{k,\mu} \rangle - \langle r_{k,\mu}, \Pi_{k,\mu} r_{k,\mu} \rangle \right] \\
&= f(v_{k,\mu}) - \frac{1}{2} \frac{\langle \Pi_{k,\mu} r_{k,\mu}, r_{k,\mu} \rangle}{\|b\|^2},
\end{aligned}$$

i.e.  $f(v_{k,\mu}) - f(v_{k,\mu+1}) = \frac{1}{2} \frac{\langle \Pi_{k,\mu} r_{k,\mu}, r_{k,\mu} \rangle}{\|b\|^2}$ . ■

**Corollary 3.3.** *There exists  $\alpha \in \mathbb{R}$  such that  $f(v_k) \xrightarrow[k \rightarrow \infty]{} \alpha$ .*

*Proof.* Let  $k \in \mathbb{N}$  and  $\mu \in \mathbb{N}_L$ . From Lemma 3.2 and Lemma 3.1 it follows that

$$\begin{aligned}
f(v_{k+1}) - f(v_k) &= f(v_{k,d}) - f(v_{k,0}) = \sum_{\mu=1}^d f(v_{k,\mu}) - f(v_{k,\mu-1}) \\
&= -\frac{1}{2\|b\|^2} \sum_{\mu=0}^{d-1} \|\Pi_{k,\mu} r_{k,\mu}\|^2 \leq 0,
\end{aligned}$$

This shows that  $(f(v_k))_{k \in \mathbb{N}} \subset \mathbb{R}$  is a descending sequence. The sequence of function values  $(f(v_k))_{k \in \mathbb{N}}$  is bounded from below. Therefore, there exist an  $\alpha \in \mathbb{R}$  such that  $f(v_k) \xrightarrow[k \rightarrow \infty]{} \alpha$ . ■

**Remark 3.4.** *From the definition of the ALS method it is already clear that  $(f(v_{k,\mu}))_{\mu \in \mathbb{N}_d, k \in \mathbb{N}}$  is a descending sequence.*

**Lemma 3.5.** *Let  $(v_{k,\mu})_{k \in \mathbb{N}, \mu \in \mathbb{N}_d} \subset \mathcal{V}$  be the sequence from Algorithm 1. We have*

$$f(v_{k,\mu}) = -\frac{1}{2\|b\|^2} \langle v_{k,\mu}, b \rangle = -\frac{1}{2\|b\|^2} \|v_{k,\mu}\|^2 \quad (24)$$

for all  $k \in \mathbb{N}, \mu \in \mathbb{N}_d$ .

*Proof.* Let  $k \in \mathbb{N}$  and  $\mu \in \mathbb{N}_d$ . With Lemma 3.1 it follows

$$\langle v_{k,\mu}, v_{k,\mu} \rangle = \langle \Pi_{k,\mu-1} b, \Pi_{k,\mu-1} b \rangle = \langle \Pi_{k,\mu-1}^2 b, b \rangle = \langle \Pi_{k,\mu-1} b, b \rangle = \langle v_{k,\mu}, b \rangle.$$

The rest follows from the definition of  $f$ , see Eq. (1). ■

**Corollary 3.6.** *Let  $(v_{k,\mu})_{k \in \mathbb{N}, \mu \in \mathbb{N}_d} \subset \mathcal{V}$  be the sequence of represented tensors from the ALS algorithm. Further, let  $\mu \in \mathbb{N}_d$  and  $k \in \mathbb{N}$ . The following statements are equivalent:*

- (a)  $f(v_{k,\mu+1}) \leq f(v_{k,\mu})$
- (b)  $\|v_{k,\mu+1}\|^2 \geq \|v_{k,\mu}\|^2$
- (c)  $\|p_\mu^{k+1}\|^2 \geq \|p_\mu^k\|^2$
- (d)  $\cos^2(\varphi_{k,\mu+1}) \geq \cos^2(\varphi_{k,\mu})$ , where  $\cos^2(\varphi_{k,\mu}) := \frac{\langle \Pi_{k,\mu} b, b \rangle}{\|b\|^2}$ .

*Proof.* Follows direct from Lemma 3.5 and

$$\|v_{k,\mu+1}\|^2 \geq \|v_{k,\mu}\|^2 \Leftrightarrow G_{k,\mu} \|p_\mu^{k+1}\|^2 \geq G_{k,\mu} \|p_\mu^k\|^2$$

where  $G_{k,\mu} > 0$  is defined in Corollary 2.4. ■

**Lemma 3.7.** *Let  $(v_k)_{k \in \mathbb{N}} \subset \mathcal{V}$  be the sequence of represented tensors from the ALS method. It holds*

$$\|v_{k+1} - v_k\| \xrightarrow[k \rightarrow \infty]{} 0.$$

*Proof.* Let  $k \in \mathbb{N}$ . We have

$$\|v_{k+1} - v_k\|^2 = \left\| \sum_{\mu=1}^d v_{k,\mu} - v_{k,\mu-1} \right\|^2 \leq \left( \sum_{\mu=1}^d \|v_{k,\mu} - v_{k,\mu-1}\| \right)^2 \leq d \sum_{\mu=0}^{d-1} \|v_{k,\mu+1} - v_{k,\mu}\|^2. \quad (25)$$

Since  $v_{k,\mu+1} - v_{k,\mu} = \Pi_{k,\mu} r_{k,\mu}$ , see Lemma 3.1, it follows further with Eq. (23) and (25) that

$$\|v_{k+1} - v_k\|^2 \leq 2d \|b\|^2 \sum_{\mu=0}^{d-1} (f(v_{k,\mu+1}) - f(v_{k,\mu})).$$

With Corollary 3.3 we have  $(f(v_{k,\mu+1}) - f(v_{k,\mu})) \xrightarrow[k \rightarrow \infty]{} 0$ , hence  $\|v_{k+1} - v_k\| \xrightarrow[k \rightarrow \infty]{} 0$ . ■

**Definition 3.8** ( $\mathcal{A}(v_k)$ , critical points). *Let  $(v_k)_{k \in \mathbb{N}} \subset \mathcal{V}$  be the sequence of represented tensors from Algorithm 1. The set of accumulation points of  $(v_k)_{k \in \mathbb{N}}$  is denoted by  $\mathcal{A}(v_k)$ , i.e.*

$$\mathcal{A}(v_k) := \{v \in \mathcal{V} : v \text{ is an accumulation point of } (v_k)_{k \in \mathbb{N}}\}. \quad (26)$$

The set  $\mathfrak{M}$  of critical points of the optimisation problem from Eq. (2) is defined as follows:

$$\mathfrak{M} := \{v \in \mathcal{V} : \exists \underline{p} \in P : v = U(\underline{p}) \wedge F'(\underline{p}) = 0\}. \quad (27)$$

**Proposition 3.9.** *The sequence of parameter  $(p_{\mu,k})_{\mu \in \mathbb{N}_d, k \in \mathbb{N}}$  from the ALS algorithm is bounded.*

*Proof.* From the definition of  $f$  and Lemma 3.5 it follows that

$$-\frac{1}{2} \leq f(v_{k,\mu}) = -\frac{1}{2} \frac{\|v_{k,\mu}\|^2}{\|b\|^2} \Leftrightarrow \|v_{k,\mu}\| \leq \|b\|,$$

i.e. the sequence  $(\|v_{\mu,k}\|)_{\mu \in \mathbb{N}_d, k \in \mathbb{N}} \subset \text{Range}(U)$  is bounded. The sequence  $(\|v_{\mu,k}\|)_{\mu \in \mathbb{N}_d, k \in \mathbb{N}}$  is the product of the following  $d$  sequences  $(\|p_\mu^k\|)_{k \in \mathbb{N}} \subset \mathbb{R}^{n_\mu}$ . According to Corollary 3.6 the sequences  $(\|p_\mu^k\|)_{k \in \mathbb{N}}$  are monotonically increasing. Since the product  $\|v_{\mu,k}\|$  is bounded and all sequences  $(\|p_\mu^k\|)_{k \in \mathbb{N}}$  are monotonically increasing, it follows that all  $(p_\mu^k)_{k \in \mathbb{N}}$  are bounded. This means the sequence  $(p_{\mu,k})_{\mu \in \mathbb{N}_d, k \in \mathbb{N}}$  is bounded. ■

The following statements are proved in a corresponding article about the convergence of alternating least squares optimisation in general tensor format representations, please see [5] for more informations regarding the proofs.

**Lemma 3.10** ([5]). *We have*

$$\max_{0 \leq \mu \leq L-1} \left\| F'_\mu(p_\mu^k) \right\| \xrightarrow[k \rightarrow \infty]{} 0.$$

**Corollary 3.11** ([5]). Let  $(\underline{p}_k)_{k \in \mathbb{N}}$  be the sequence from Algorithm 1 and  $F : P \rightarrow \mathbb{R}$  from Eq. (2). We have

$$\lim_{k \rightarrow \infty} F'(\underline{p}_k) = 0.$$

**Theorem 3.12** ([5]). Let  $(v_k)_{k \in \mathbb{N}}$  be the sequence of represented tensors from the ALS method. Every accumulation point of  $(v_k)_{k \in \mathbb{N}}$  is a critical point, i.e.  $\mathcal{A}(v_k) \subseteq \mathfrak{M}$ . Further, we have

$$\text{dist}(v_k, \mathfrak{M}) \xrightarrow[k \rightarrow \infty]{} 0.$$

Let  $\bar{v} \in \mathfrak{M}$  be a critical point and  $N := \prod_{\mu=1}^d n_\mu \in \mathbb{N}$ . Further, let  $(\underline{p}_{k,\mu})_{k \in \mathbb{N}, \mu \in \mathbb{N}_d} \subset P$  be the sequence of parameter from the ALS algorithm and  $R \in \mathbb{R}^{N \times N-1}$  be a matrix with  $R^T R = \mathbf{Id}_{\mathbb{R}^{N-1}}$  and  $\text{span}(\bar{v})^\perp = \text{Range}(R)$ , i.e. the column vectors of  $R$  build an orthonormal basis of the linear space  $\text{span}(\bar{v})^\perp$ . Then the block matrix

$$V := \begin{bmatrix} \underline{v} & R \end{bmatrix} \in \mathbb{R}^{N \times N}, \quad (\underline{v} := \bar{v}/\|\bar{v}\|). \quad (28)$$

is orthogonal, i.e. the columns of the matrix  $V$  build an orthonormal basis of the tensor space  $\mathcal{V}$ . The following matrix  $N_{k,\mu} \in \mathbb{R}^{N \times N}$  is imported in order to describe the rate of convergence for the ALS method:

$$N_{k,\mu} := \bigotimes_{\nu=1}^{\mu-1} \mathbf{Id} \otimes \left( \frac{1}{G_{k,\mu} G_{k,\mu-1}} M_{\mu,k} M_{\mu,k}^T \right) \otimes \bigotimes_{\nu=\mu+1}^d \mathbf{Id},$$

where the matrix  $\frac{1}{G_{k,\mu} G_{k,\mu-1}} M_{\mu,k} M_{\mu,k}^T$  is from Corollary 2.4. Further, it follows from Corollary 2.4 that for the ALS micro step the following equation:

$$v_{k,\mu+1} = N_{k,\mu} v_{k,\mu} \quad (29)$$

holds. The tensor  $v_{k,\mu}$  and the matrix  $N_{k,\mu}$  are represented with respect to the basis  $V$ , i.e

$$v_{k,\mu} = V V^T v_{k,\mu} = \begin{bmatrix} \underline{v} & R \end{bmatrix} \begin{pmatrix} \underbrace{\underline{v}^T v_{k,\mu}}_{c_{k,\mu} :=} \\ \underbrace{R^T v_{k,\mu}}_{s_{k,\mu} :=} \end{pmatrix} = \begin{bmatrix} \underline{v} & R \end{bmatrix} \begin{pmatrix} c_{k,\mu} \\ s_{k,\mu} \end{pmatrix}$$

and

$$N_{k,\mu} = V (V^T N_{k,\mu} V) V^T = \begin{bmatrix} \underline{v} & R \end{bmatrix} \begin{bmatrix} \underline{v}^T N_{k,\mu} \underline{v} & \underline{v}^T N_{k,\mu} R \\ R^T N_{k,\mu} \underline{v} & R^T N_{k,\mu} R \end{bmatrix} \begin{bmatrix} \underline{v} & R \end{bmatrix}^T.$$

The recursion formula (29) leads to the recursion of the coefficient vector

$$\begin{pmatrix} c_{k+1,\mu} \\ s_{k+1,\mu} \end{pmatrix} = \begin{bmatrix} \underline{v}^T N_{k,\mu} \underline{v} & \underline{v}^T N_{k,\mu} R \\ R^T N_{k,\mu} \underline{v} & R^T N_{k,\mu} R \end{bmatrix} \begin{pmatrix} c_{k,\mu} \\ s_{k,\mu} \end{pmatrix} = \begin{pmatrix} \underline{v}^T N_{k,\mu} \underline{v} c_{k,\mu} + \underline{v}^T N_{k,\mu} R s_{k,\mu} \\ R^T N_{k,\mu} \underline{v} c_{k,\mu} + R^T N_{k,\mu} R s_{k,\mu} \end{pmatrix}.$$

Without loss of generality we can assume that  $\|s_{k,\mu}\| \neq 0$  and  $|c_{k,\mu}| \neq 0$ . Therefore, the following terms are well defined:

$$q_{k,\mu}^{(s)} := \frac{\|R^T N_{k,\mu} \underline{v} c_{k,\mu} + R^T N_{k,\mu} R s_{k,\mu}\|}{\|s_{k,\mu}\|},$$

$$q_{k,\mu}^{(c)} := \frac{|\underline{v}^T N_{k,\mu} \underline{v} c_{k,\mu} + \underline{v}^T N_{k,\mu} R s_{k,\mu}|}{|c_{k,\mu}|}.$$

This preconsideration gives a recursion formula for the tangent of the angle between  $\bar{v}$  and  $v_{k,\mu+1}$ . We have

$$\begin{aligned}\tan^2 \angle[\bar{v}, v_{k,\mu+1}] &= \frac{\langle RR^T v_{k,\mu+1}, v_{k,\mu+1} \rangle}{\langle \underline{v} \underline{v}^T v_{k,\mu+1}, v_{k,\mu+1} \rangle} = \frac{\|R^T v_{k,\mu+1}\|^2}{(\underline{v}^T v_{k,\mu+1})^2} = \frac{\|s_{k,\mu+1}\|^2}{(c_{k,\mu+1})^2} = \frac{\left(\frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}}\right)^2 \|s_{k,\mu}\|^2}{\left(\frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}}\right)^2 (c_{k,\mu})^2} \\ &= \left(\frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}}\right)^2 \frac{\|R^T v_{k,\mu}\|^2}{(\underline{v}^T v_{k,\mu})^2} = \left(\frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}}\right)^2 \tan^2 \angle[\bar{v}, v_{k,\mu}].\end{aligned}$$

**Remark 3.13.** Obviously, if the sequence of parameter  $(\underline{p}_k)_{k \in \mathbb{N}} \subset P$  is bounded, then the set of accumulation points of  $(\underline{p}_k)_{k \in \mathbb{N}}$  is not empty. Consequently, the set  $\mathcal{A}(v_k)$  is not empty, since the map  $U$  is continuous.

**Theorem 3.14** ([5]). If one accumulation point  $\bar{v} \in \mathcal{A}(v_k) \subseteq \mathfrak{M}$  is isolated, then we have

$$v_k \xrightarrow[k \rightarrow \infty]{} \bar{v}.$$

Furthermore, we have for the rate of convergence of an ALS micro step

$$|\tan \angle[\bar{v}, v_{k,\mu+1}]| \leq q_\mu |\tan \angle[\bar{v}, v_{k,\mu}]|,$$

where

$$q_\mu := \limsup_{k \rightarrow \infty} \left| \frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}} \right|.$$

If  $q_\mu = 0$ , then the sequence  $(|\tan \angle[\bar{v}, v_{k,\mu}]|)_{k \in \mathbb{N}}$  converges  $Q$ -superlinearly. If  $q_\mu < 1$ , then the sequence  $(|\tan \angle[\bar{v}, v_{k,\mu}]|)_{k \in \mathbb{N}}$  converges at least  $Q$ -linearly. If  $q_\mu \geq 1$ , then the sequence  $(|\tan \angle[\bar{v}, v_{k,\mu}]|)_{k \in \mathbb{N}}$  converges not  $Q$ -linearly.

**Remark 3.15.** The calculation from Example 1.2 shows that

$$\limsup_{k \rightarrow \infty} \left| \frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}} \right| = 0 \quad \text{for all } \mu \in \mathbb{N}_d.$$

Hence, the ALS algorithm converges here  $Q$ -superlinearly. Furthermore, in Example 1.3 we showed for  $\lambda < \frac{1}{2}$

$$\limsup_{k \rightarrow \infty} \left| \frac{q_{k,\mu}^{(s)}}{q_{k,\mu}^{(c)}} \right| = \frac{\lambda}{2} \left( 3\lambda + \lambda^2 + \sqrt{(3\lambda + \lambda^2)^2 + 4\lambda} \right) < 1 \quad \text{for all } \mu \in \mathbb{N}_d.$$

Hence, we have here  $Q$ -linear convergence.

**Corollary 3.16** ([5]). If the set of critical points  $\mathfrak{M}$  is discrete,<sup>1</sup> then the sequence of represented tensors  $(v_k)_{k \in \mathbb{N}}$  from the ALS method is convergent.

In the following example it will be shown, that the ordering of the indices may play an important role for the convergence of ALS procedure.

**Remark 3.17.** Let  $b = \bigotimes_{\mu=1}^3 b_{1\mu} + \lambda \bigotimes_{\mu=1}^3 b_{2\mu}$ , with  $0 < \lambda < 1$ ,  $\|b_{1\mu}\| = \|b_{2\mu}\| = 1$  and  $\langle b_{1\mu}, b_{2\mu} \rangle = 0$  for  $\mu \in \mathbb{N}_{\leq d}$ . Let further  $v^0 = C \bigotimes_{\mu=1}^d p_1^0$  for some  $C \in \mathbb{R}$  and

$$p_\mu^0 = b_{1\mu} + \alpha_\mu b_{2\mu} \tag{30}$$

<sup>1</sup>In topology, a set which is made up only of isolated points is called discrete.

for some  $\alpha_\mu \in \mathbb{R}$ . Assume after each ALS micro step the parameters  $p_\mu^k$  are rescaled to the form (30) (obviously, a scaling of parameters has no effect on the future behavior of the ALS method). After the first four micro steps one gets

$$\begin{aligned} p_1^1 &= b_{11} + \lambda \alpha_2 \alpha_3 b_{21} \\ p_2^1 &= b_{12} + \lambda^2 \alpha_2 \alpha_3^2 b_{22} \\ p_3^1 &= b_{13} + \lambda^4 \alpha_2^2 \alpha_3^3 b_{23} \\ p_1^2 &= b_{11} + \lambda^7 \alpha_2^3 \alpha_3^5 b_{21} \end{aligned}$$

So for  $v_1^2 := p_1^2 \otimes p_2^1 \otimes p_3^1$  one gets

$$v_1^2 = \hat{C}(b_{11} + \lambda^7 \alpha_2^3 \alpha_3^5 b_{21}) \otimes (b_{13} + \lambda^2 \alpha_2 \alpha_3^2 b_{23}) \otimes (b_{12} + \lambda^4 \alpha_2^2 \alpha_3^3 b_{22})$$

with some  $\hat{C} \in \mathbb{R}$ . Now assume the order of the directions for ALS optimization is changed from (1, 2, 3) to (1, 3, 2), i.e. after optimizing the first component  $p_1^1$  we optimize the third one (i.e.  $p_3^1$ ) and only then the second one (i.e.  $p_2^1$ ). The same number of micro steps will result in a tensor

$$\tilde{v}_1^2 = \tilde{C}(b_{11} + \lambda^7 \alpha_2^5 \alpha_3^2 b_{21}) \otimes (b_{13} + \lambda^4 \alpha_2^3 \alpha_3^2 b_{23}) \otimes (b_{12} + \lambda^2 \alpha_2^2 \alpha_3 b_{22})$$

with some  $\tilde{C} \in \mathbb{R}$ . Now if  $\alpha_2$  and  $\alpha_3$  satisfy

$$\begin{aligned} \alpha_2 &\geq 1 \geq \alpha_3, \\ \alpha_2^3 \alpha_3^2 &\geq \frac{1}{\lambda^5} \geq \alpha_2^2 \alpha_3^3, \end{aligned}$$

then it is not difficult to check, that  $v_1^2$  satisfies the dominance condition from Eq. (7) for  $j = 1$ , whereas  $\tilde{v}_1^2$  satisfies the dominance condition for  $j = 2$ . Thus, with the same starting point  $v^0$  ALS iteration will converge to the global minimum  $\bigotimes_{\mu=1}^d b_{1\mu}$  for one ordering of the indices and to local minimum  $\lambda \bigotimes_{\mu=1}^d b_{2\mu}$  for another ordering. Note that  $v_0$  did not fulfil the dominance conditions, but depending on the ordering of the ALS micro steps  $v_0$  leads to different dominance conditions.

## 4 Numerical Experiments

In this subsection, we observe the convergence behavior of the ALS method by using data from interesting examples and more importantly from real applications. In all cases, we focus particularly on the convergence rate.

### 4.1 Example 1

We consider an example introduced by Mohlenkamp in [9, Section 4.3.5]. Here we have

$$b = 2 \underbrace{\left( \begin{array}{c} 1 \\ 0 \end{array} \right) \otimes \left( \begin{array}{c} 1 \\ 0 \end{array} \right) \otimes \left( \begin{array}{c} 1 \\ 0 \end{array} \right)}_{b_1 :=} + \underbrace{\left( \begin{array}{c} 0 \\ 1 \end{array} \right) \otimes \left( \begin{array}{c} 0 \\ 1 \end{array} \right) \otimes \left( \begin{array}{c} 0 \\ 1 \end{array} \right)}_{b_2 :=},$$

see Eq. (1). The tensor  $b$  is orthogonally decomposable. Although the example is rather simple, it is of theoretical interest. Since the ALS method converges superlinear, cf. the discussion in Section 1. The tensor

$b$  has only two terms, therefore the upper bound for convergence rate from Eq. (8) is sharp, cf. Eq. (9). Let  $\tau \geq 0$ , we define the initial guess of the ALS algorithm by

$$v_0(\tau) := \begin{pmatrix} \tau \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \tau \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \tau \\ 1 \end{pmatrix}.$$

Since

$$4 \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau \\ 1 \end{pmatrix} \right\rangle^2 = 4\tau^2 \quad \text{and} \quad \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} \tau \\ 1 \end{pmatrix} \right\rangle^2 = 1,$$

we have for  $\tau < \frac{1}{2}$  that the initial guess  $v_0(\tau)$  dominates at  $b_2$ . Therefore, the ALS iteration converge to  $b_2$ . If  $\tau > \frac{1}{2}$ , then  $v_0(\tau)$  dominates at  $b_1$  and the sequence from the ALS method will converges to  $b_1$ . In the first test the tangents of the angle between the current iteration point and the corresponding parameter of the dominate term  $b_l$  ( $1 \leq l \leq 2$ ) is plotted, i.e.

$$\tan \varphi_{k,l} = \sqrt{\frac{1 - \cos^2 \varphi_{k,l}}{\cos^2 \varphi_{k,l}}}, \quad (31)$$

where  $\cos \varphi_{k,l} = \frac{\langle p_1^k, e_l \rangle}{\|p_1^k\|}$ . To illustrate the superlinear convergence of the ALS method, we present further plots for the quotient

$$q_{k,l} := \frac{\tan \varphi_{k+1,l}}{\tan \varphi_{k,l}}. \quad (32)$$

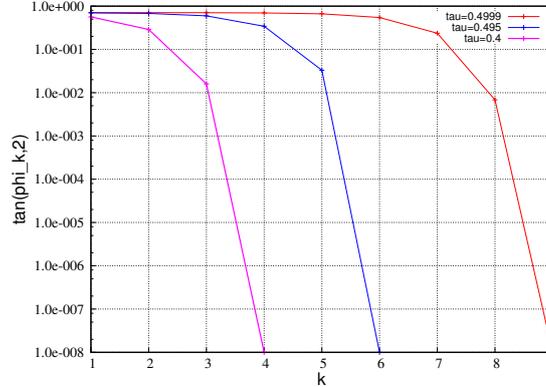
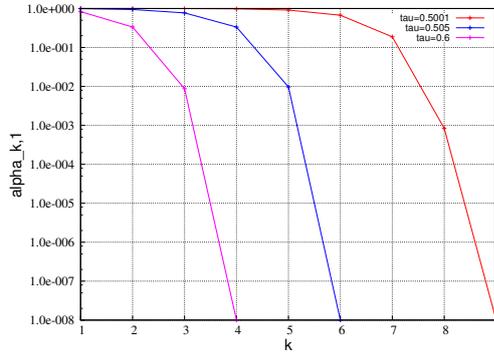


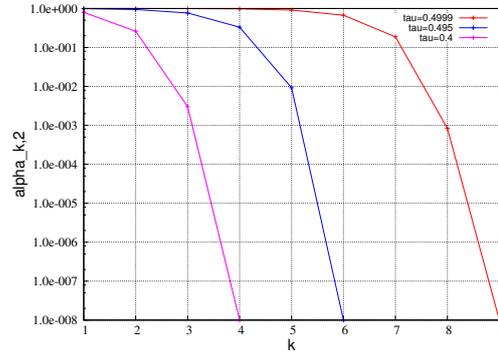
Figure 2: The tangents  $\tan \varphi_{k,2}$  from Eq. (31) is plotted for  $\tau \in \{0.4, 0.495, 0.4999\}$ .

## 4.2 Example 2

Most algorithms in ab initio electronic structure theory compute quantities in terms of one- and two-electron integrals. In [1] we considered the low-rank approximation of the two-electron integrals. In order to demonstrate the convergence of the ALS method on an example of practical interest, we use the order 4 tensor for the two-electron integrals of the so called AO basis for the  $\text{CH}_4$  molecule. We refer the reader to [1] for a detailed description our example. In this example the ALS method converges Q-linearly, see Figure 4.



(a)  $q_{k,1}$  is plotted for  $\tau \in \{0.5001, 0.505, 0.6\}$ . Here the term  $b_1$  dominates at every iteration point.



(b)  $q_{k,2}$  is plotted for  $\tau \in \{0.4999, 0.495, 0.4\}$ . Here the term  $b_2$  dominates at every iteration point.

Figure 3:  $q_{k,l}$  from Eq. (32) is plotted for  $l \in \{1, 2\}$  and different values for  $\tau$ .

### 4.3 Example 3

We consider the tensor

$$b_\lambda = \bigotimes_{\mu=1}^3 p + \lambda (p \otimes q \otimes q + q \otimes p \otimes q + q \otimes q \otimes p)$$

from Ex. 1.3. The vectors  $p$  and  $q$  are arbitrarily generated orthogonal vectors with norm 1. The values of  $\tan(\varphi_k^1)$  are plotted, where  $\varphi_k^1$  is the angle between  $p_k^1$  and the limit point  $p$  (i.e.  $\tan \varphi_k^1 = \frac{\langle p_k^1, q \rangle}{\langle p_k^1, p \rangle}$ , for  $k \geq 2$ ). For the case  $\lambda = 0.5$  the convergence is sublinearly, whereas for  $\lambda = 0.2$  it is Q-linearly.

## References

- [1] U. Benedikt, A. Auer, M. Espig, and W. Hackbusch. Tensor decomposition in post-hartree–fock methods. i. two-electron integrals and mp2. *The Journal of Chemical Physics*, 134(5):–, 2011.
- [2] G. Beylkin and M. J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proceedings of the National Academy of Sciences*, 99(16):10246–10251, 2002.
- [3] G. Beylkin and M. J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM Journal on Scientific Computing*, 26(6):2133–2159, 2005.
- [4] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization problems in contracted tensor networks. *Computing and Visualization in Science*, 14(6):271–285, 2011.
- [5] M. Espig, W. Hackbusch, and A. Khachatryan. On the convergence of alternating least squares optimisation in tensor format representations. *Preprint*, 2014.
- [6] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer, 2012.
- [7] S. Holtz, T. Rohwedder, and R. Schneider. The alternating linear scheme for tensor optimization in the tensor train format. *SIAM J. Sci. Comput.*, 34(2):683–713, March 2012.

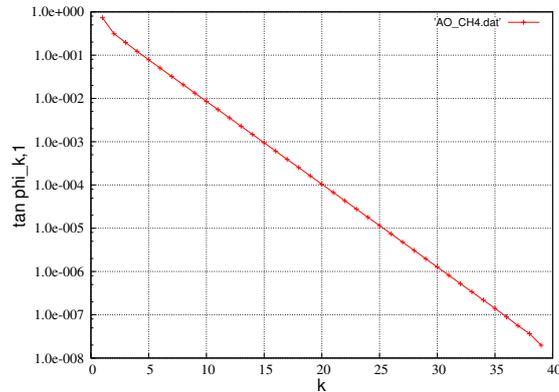
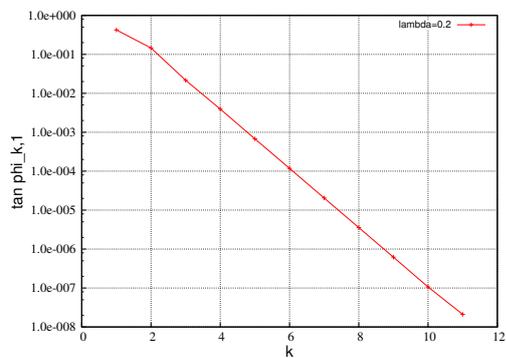
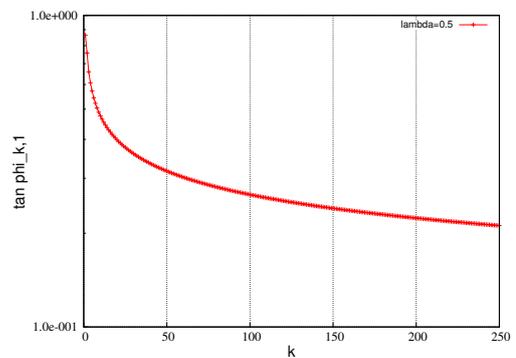


Figure 4: The approximation of two-electron integrals for methane is considered. The tangents of the angle between the current iteration point and the limit point with respect to the iteration number is plotted.

- [8] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM REVIEW*, 51(3):455–500, 2009.
- [9] M. J. Mohlenkamp. Musings on multilinear fitting. *Linear Algebra Appl.*, 438(2):834–852, 2013.
- [10] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial Mathematics, 1970.
- [11] I. V. Oseledets. Dmrg approach to fast linear algebra in the tt-format. *Comput. Meth. in Appl. Math.*, 11(3):382–393, 2011.
- [12] I. V. Oseledets and S. V. Dolgov. Solution of linear systems and matrix inversion in the tt-format. *SIAM J. Scientific Computing*, 34(5), 2012.
- [13] A. Uschmajew. Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM Journal on Matrix Analysis and Applications*, 33(2):639–652, 2012.
- [14] André Uschmajew. A new convergence proof for the high-order power method and generalizations. July 2014.
- [15] L. Wang and M. Chu. On the global convergence of the alternating least squares method for rank-one approximation to generic tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(3):1058–1072, 2014.
- [16] Tong Zhang and Gene H. Golub. Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.*, 23(2):534–550, February 2001.



(a) The tangents  $\tan \varphi_{k,1}$  for  $\lambda = 0.2$ .



(b) The tangents  $\tan \varphi_{k,1}$  for  $\lambda = 0.5$ .

Figure 5: The approximation of  $b$  from Example 1.3 is considered. The tangents of the angle between the current iteration point and the limit point with respect to the iteration number is plotted. For  $\lambda = 1/2$ , we have sublinear convergence. But for  $\lambda = 0.2 < 1/2$  the sequence converges Q-linearly.

# 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  $\mathbb{R}^N$ . 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. Kreuzler 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 Multi-level 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.
- [113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.
- [114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.
- [115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.
- [116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.
- [117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

- [118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.
- [119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.
- [120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.
- [121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.
- [122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.
- [123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.
- [124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.
- [125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.
- [126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.
- [127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.
- [128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.
- [129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.
- [130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the  $L_q(L_p)$ -regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

- [131] M. Hansen.  $n$ -term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.
- [132] R. E. Bank and H. Yserentant. On the  $H^1$ -stability of the  $L_2$ -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.
- [133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.
- [134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.
- [135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.
- [136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.
- [137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.
- [138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.
- [139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.
- [140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.
- [141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.
- [142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.
- [143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

- [144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.
- [145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.
- [146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.
- [147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.
- [148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.
- [149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.
- [150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.
- [151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.
- [152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.
- [153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.
- [154] M. Kovács, S. Larsson, and F. Lindgren. On the Backward Euler Approximation of the Stochastic Allen-Chan Equation. Preprint 154, DFG-SPP 1324, November 2013.
- [155] S. Dahlke, M. Fornasier, U. Friedrich, and T. Raasch. Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms. Preprint 155, DFG-SPP 1324, December 2013.
- [156] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. On the complexity of computing quadrature formulas for marginal distributions of SDEs. Preprint 156, DFG-SPP 1324, January 2014.

- [157] M. Giles, T. Nagapetyan, and K. Ritter. Multi-Level Monte Carlo Approximation of Distribution Functions and Densities. Preprint 157, DFG-SPP 1324, February 2014.
- [158] F. Dickmann and N. Schweizer. Faster Comparison of Stopping Times by Nested Conditional Monte Carlo . Preprint 158, DFG-SPP 1324, February 2014.
- [159] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate periodic functions by trigonometric polynomials based on sampling along rank-1 lattice with generating vector of Korobov form . Preprint 159, DFG-SPP 1324, February 2014.
- [160] S. Dereich and T. Müller-Gronbach. Quadrature for Self-Affine Distributions on  $\mathbb{R}^d$ . Preprint 160, DFG-SPP 1324, March 2014.
- [161] S. Dereich and S. Li. Multilevel Monte Carlo for Lévy-driven SDEs: central limit theorems for adaptive Euler schemes. Preprint 161, DFG-SPP 1324, March 2014.
- [162] M. Kluge. Sampling Rules for Tensor Reconstruction in Hierarchical Tucker Format. Preprint 162, DFG-SPP 1324, April 2014.
- [163] D. Rudolf and N. Schweizer. Error Bounds of MCMC for Functions with Unbounded Stationary Variance. Preprint 163, DFG-SPP 1324, April 2014.
- [164] E. Novak and D. Rudolf. Tractability of the approximation of high-dimensional rank one tensors. Preprint 164, DFG-SPP 1324, April 2014.
- [165] J. Dick and D. Rudolf. Discrepancy estimates for variance bounding Markov chain quasi-Monte Carlo. Preprint 165, DFG-SPP 1324, April 2014.
- [166] L. Kämmerer, S. Kunis, I. Melzer, D. Potts, and T. Volkmer. Computational Methods for the Fourier Analysis of Sparse High-Dimensional Functions. Preprint 166, DFG-SPP 1324, April 2014.
- [167] T. Müller-Gronbach and L. Yaroslavtseva. Deterministic quadrature formulas for SDEs based on simplified weak Ito-Taylor steps. Preprint 167, DFG-SPP 1324, June 2014.
- [168] D. Belomestny, T. Nagapetyan, and V. Shiryayev. Multilevel Path Simulation for Weak Approximation Schemes: Myth or Reality. Preprint 168, DFG-SPP 1324, June 2014.
- [169] S. Dahlke, N. Döhring, and S. Kinzel. A class of random functions in non-standard smoothness spaces. Preprint 169, DFG-SPP 1324, September 2014.

- [170] T. Patschkowski and A. Rohde. Adaptation to lowest density regions with application to support recovery. Preprint 170, DFG-SPP 1324, September 2014.
- [171] D. Potts and T. Volkmer. Sparse high-dimensional FFT based on rank-1 lattice sampling. Preprint 171, DFG-SPP 1324, October 2014.
- [172] M. Kovács, F. Lindner, and R. L. Schilling. Weak convergence of finite element approximations of linear stochastic evolution equations with additive Lévy noise. Preprint 172, DFG-SPP 1324, November 2014.
- [173] M. Espig and A. Khachatryan. Convergence of Alternating Least Squares Optimisation for Rank-One Approximation to High Order Tensors. Preprint 173, DFG-SPP 1324, November 2014.