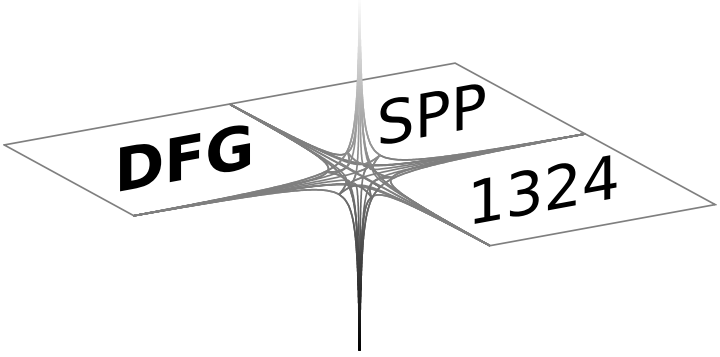
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors

L. Grasedyck, M. Kluge, S. Krämer

Preprint 149



Edited by

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

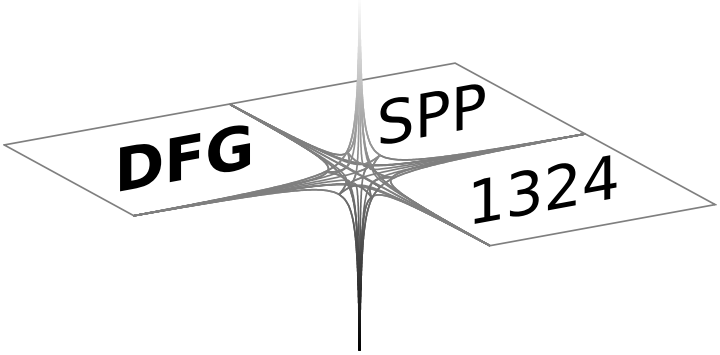
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors

L. Grasedyck, M. Kluge, S. Krämer

Preprint 149



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.

ALTERNATING DIRECTIONS FITTING (ADF) OF HIERARCHICAL LOW RANK TENSORS

LARS GRASEDYCK*, MELANIE KLUGE* , AND SEBASTIAN KRÄMER*

Abstract. We consider the problem of fitting a low rank tensor $A \in \mathbb{R}^{\mathcal{I}}$, $\mathcal{I} = \{1, \dots, n\}^d$, to a given set of data points $\{M_i \in \mathbb{R} \mid i \in P\}$, $P \subset \mathcal{I}$. The low rank format under consideration is the hierarchical or TT or MPS format. It is characterized by rank bounds r on certain matricizations of the tensor. The number of degrees of freedom is in $\mathcal{O}(dnr^2)$. For a fixed rank and mode size n we observe that it is possible to approximate a tensor from a number of samples that is in $\mathcal{O}(\log N)$ for a tensor having $N = n^d$ entries. Our approach is an alternating directions fitting (ADF) inspired by the LMaFit method for matrix completion, but generalized to tensor completion. We aim at finding a tensor A that fulfils the first order optimality conditions by a nonlinear SOR-type solver that consists of an alternating fit cycling through the directions $\mu = 1, \dots, d$. In the numerical experiments we observe robustness of the completion algorithm with respect to noise and good reconstruction capability. Our tests provide evidence that the algorithm is suitable in higher dimension (>10) as well as for moderate ranks.

Keywords: MPS, Tensor Completion, Tensor Train, TT, Hierarchical Tucker, HT, ALS.
MSC: 15A69, 65F99

1. Introduction.

We consider the problem of fitting a low rank tensor

$$A \in \mathbb{R}^{\mathcal{I}}, \quad \mathcal{I} := \mathcal{I}_1 \times \cdots \times \mathcal{I}_d, \quad \mathcal{I}_\mu := \{1, \dots, n_\mu\}, \quad \mu \in D := \{1, \dots, d\},$$

to given data points

$$\{M_i \in \mathbb{R} \mid i \in P\}, \quad P \subset \mathcal{I}, \qquad \#P \geq \sum_{\mu=1}^{d} n_\mu,$$

by minimizing the distance between the given values $(M_i)_{i \in P}$ and approximations $(A_i)_{i \in P}$:

$$A = \operatorname*{argmin}_{\tilde{A} \in T} \sum_{i \in P} (M_i - \tilde{A}_i)^2 \qquad (T \text{ being a certain tensor class})$$

In the class of general dense tensors this is trivial, because the entries of the tensor are all independent. For sparse tensors this reduces to a simple Knapsack problem. Our target tensor class is the set of low rank tensors, i.e., we assume that the implicitly given tensor $M \in \mathbb{R}^{\mathcal{I}}$ allows for a low rank approximation

$$\|M - \tilde{M}\| \leq \varepsilon, \qquad \varepsilon \in \mathbb{R}_{\geq 0},$$

where the unknown approximant $\tilde{M} \in \mathbb{R}^{\mathcal{I}}$ fulfils certain rank bounds that will be introduced later. In particular we allow $\varepsilon = 0$ so that the task is to reconstruct the whole tensor $M = \tilde{M}$ in the low rank format. This particular case is considered, e.g. in [11, 3].

1.1. Completion versus Sampling.

A tensor fitting problem might arise as follows: the entries $(M_i)_{i \in P}$ could be measurements of a multiparameter model such that each index $i \in P$ represents a specific choice of d parameters. If the measurements are incomplete or in parts known to be incorrect, then the goal is to reconstruct all values of M for all parameter combinations $i \in \mathcal{I}$ from the known values $(M_i)_{i \in P}$ (prior to the assumption that M allows for an approximation in the low rank format). It is crucial that the points P

*Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Germany. Email: lgr@igpm.rwth-aachen.de. All three authors gratefully acknowledge support by the DFG priority programme 1324 under grant GR3179/2-2.

are given and we are not free to choose them. In case that the points can be chosen freely one after another, the problem simplifies drastically and can be approached as in [15, 2] by an adaptive sampling strategy. Sometimes one can propose rules on how the entries from P should be chosen, as it is done in quasi Monte Carlo methods. This approach is persued in [7] and defines sampling rules that allow an efficient approximation scheme. Again, this is different and possibly a simpler task than the tensor completion considered here.

1.2. Low Rank Tensor Formats. The class of tensors in which we aim for a completion of the given tensor entries is a low rank format. In the case $d = 2$ the rank of a tensor coincides with the usual matrix rank, but in dimension $d > 2$ there are several possibilities to define the rank of a tensor and thus there are several data-sparse low rank formats available.

In the CP(k)[1] format or representation

$$A = \sum_{\ell=1}^{k} \bigotimes_{\mu=1}^{d} g_{\mu,\ell}, \qquad A_{i_1,\dots,i_d} = \sum_{\ell=1}^{k} \prod_{\mu=1}^{d} g_{\mu,\ell}(i_\mu), \qquad g_{\mu,\ell}(i_\mu) \in \mathbb{R}$$

the tensor completion has been considered in [19, 1, 10]. The minimal number of summands k by which the tensor A can be represented is the *tensor rank* of A, but minimality of k is often not relevant. The CP(k) format is data sparse in the sense that storing the factors $g_{\mu,\ell}$ amounts to $\mathcal{O}(dnk)$ units (real numbers) of storage, as opposed to the n^d units of the full dense and unstructured tensor A. This is the reason for the attractivity of the format despite many theoretical and practical difficulties [8].

In the Tucker format

$$A_{i_1,\dots,i_d} = \sum_{\ell_1=1}^{k_1} \cdots \sum_{\ell_d=1}^{k_d} C_{\ell_1,\dots,\ell_d} \prod_{\mu=1}^{d} g_{\mu,\ell_\mu}(i_\mu), \qquad g_{\mu,\ell}(i_\mu) \in \mathbb{R}, \quad C \in \mathbb{R}^{k_1 \times \cdots \times k_d},$$

tensor completion has been considered in [17, 9, 12, 16]. This format is limited to small dimensions d since the so-called core tensor C requires $\prod_{\mu=1}^{d} k_\mu$ units of storage. The advantage on the other hand is that standard matrix approximation techniques can be used by matricizing the tensor.

The low rank format that we consider lies in between these two, combining the benefits of both: the number of degrees of freedom scales linearly with the dimension d and the format is based on matricizations such that standard linear algebra tools are applicable.

Here, we put no special assumptions on the data points P, except that they are reasonably distributed:

DEFINITION 1.1 (Slices and slice density). *We define the density c_{ov} (or oversampling factor) of a point set $\{M_i \in \mathbb{R} \mid i \in P\}, P \subset \mathcal{I}$, in direction $\mu \in D$ and index $\boldsymbol{j_\mu} \in \mathcal{I}_\mu$ by*

$$c_{ov}(\boldsymbol{j_\mu}) := \#\{i \in P \mid i_\mu = \boldsymbol{j_\mu}\}$$

The corresponding slice of a tensor $A \in \mathbb{R}^{\mathcal{I}}$ is defined by

$$A_{i_\mu = \boldsymbol{j_\mu}} := \hat{A} \in \mathbb{R}^{\mathcal{I}_1 \times \cdots \times \mathcal{I}_{\mu-1} \times \mathcal{I}_{\mu+1} \times \cdots \times \mathcal{I}_d}, \qquad \hat{A}_{i_1,\dots,i_{\mu-1},i_{\mu+1},\dots,i_d} := A_{i_1,\dots,i_{\mu-1},\boldsymbol{j_\mu},i_{\mu+1},\dots,i_d}$$

Depending on the rank parameters of A (which in turn depend on the target accuracy of the approximation) the slice densities of the set P have to be high enough, i.e.

$$c_{ov}(\boldsymbol{j_\mu}) > C_{ov}, \qquad \boldsymbol{j_\mu} \in \mathcal{I}_\mu, \quad \mu \in D.$$

[1] CP stands for canonical polyadic, in the literature also called CANDECOMP and PARAFAC

If one of the values $c_{ov}(\boldsymbol{j_\mu})$ were zero, then this simply means that the slice $A_{i_\mu=\boldsymbol{j_\mu}}$ is undetermined and not observable for any of the low rank formats mentioned above and in the following.

The low rank format under consideration is the hierarchical [5, 4] or TT [14, 13] or MPS [22, 20] format.

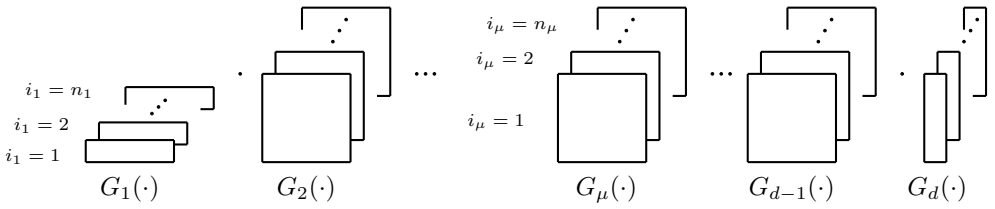


FIG. 1.1. *The TT representation of a tensor in* $TT(r_1, \dots, r_{d-1})$ *with* $G_\mu(i_\mu) \in \mathbb{R}^{r_{\mu-1} \times r_\mu}$.

DEFINITION 1.2 (TT tensor format). *Let* $r_0, \dots, r_d \in \mathbb{N}$ *and* $r_0 = r_d = 1$. *A tensor* $A \in \mathbb{R}^{\mathcal{I}}$ *of the form or representation*

$$A_{i_1,\dots,i_d} = G_1(i_1) \cdots G_d(i_d), \qquad G_\mu(i_\mu) \in \mathbb{R}^{r_{\mu-1} \times r_\mu} \tag{1.1}$$

for all $i \in \mathcal{I}$ *and* $G_\mu : \mathcal{I}_\mu \to \mathbb{R}^{r_{\mu-1} \times r_\mu}$ *is said to be of MPS (matrix product states) format or TT (tensor train) format or hierarchical format, cf. Figure 1.1. We define the set of tensors in TT format by*

$$TT(r_1, \dots, r_{d-1}) := \{A \in \mathbb{R}^{\mathcal{I}} \mid A \text{ is of the form } (1.1)\}.$$

The parameters r_μ *are called representation ranks and combined to the rank vector* $\boldsymbol{r} := (r_1, \dots, r_{d-1})$. *For the matrix blocks* $(G_\mu)_{\mu=1}^d$ *we use the short notation* G. G *is called a representation system of* A, *and if we want to indicate that* A *is represented by* G *we write* A^G. The minimal ranks r_μ for the representation of a tensor A in TT format are the ranks of certain matricizations of A [4, 15].

The number of degrees of freedom or parameters in the MPS representation is in

$$\mathcal{O}\left(\sum_{\mu=1}^{d} r_{\mu-1} r_\mu n_\mu\right) \sim \mathcal{O}(dr^2 n), \qquad r := \max_{\mu \in D} r_\mu, \quad n := \max_{\mu \in D} n_\mu.$$

It could thus in principle be possible to reconstruct the tensor from a number of samples that is in $\mathcal{O}(\log N)$ for a tensor having $N = \prod_{i=1}^d n_i$ entries, cf. Section 4.3.

1.3. Statement of the Main Approximation Problem. The full approximation problem can be stated as follows. For $S \subset \mathcal{I}$ let

$$\|X\|_F := \sqrt{\sum_{i \in \mathcal{I}} X_i^2}, \qquad (X|_S)_i := \begin{cases} X_i & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}, \qquad \|X\|_S := \|X|_S\|_F.$$

PROBLEM 1.3 (Main problem). *Given a tensor* $M \in \mathbb{R}^{\mathcal{I}}$ *known only at points* $P \subset \mathcal{I}$, *and given representation ranks* $r_1, \dots, r_{d-1}$, *find a representation (1.1) with representation system* G *such that* $A = A^G$ *fulfils*

$$A = \operatorname*{argmin}_{\tilde{A} \in TT(r_1,\dots,r_{d-1})} \|M - \tilde{A}\|_P.$$

A related approach for tensor completion is presented in [18] where the authors use a steepest descent iteration on the tensor manifold. Our approach, the alternating directions fitting (ADF) algorithm, is based on ideas from LMaFit for matrix completion [21]; and we generalize these to tensor completion.

1.4. First Order Optimality. For a representation system $(G_\mu)_{\mu=1}^d$ such that $A = A^G$ one can write the main problem in the form

$$G = \operatorname*{argmin}_{\tilde{G}} \|M - A^{\tilde{G}}\|_P$$

and by introducing an additional tensor $Z \in \mathbb{R}^{\mathcal{I}}$, G can be found via the task to

$$\text{minimize } f(G,Z) := \|Z - A^G\|_F \quad s.t. \quad Z|_P = M|_P, \quad A^G \in TT(r_1,\dots,r_{d-1}).$$

The latter function f yields first order optimality conditions

$$Z|_{\mathcal{I}\setminus P} = A^G|_{\mathcal{I}\setminus P} \qquad \text{and} \quad G_\mu = \operatorname*{argmin}_{G_\mu} \|Z - A^G\|_P.$$

Solving this nonlinear system of equations simultaneously for $G_1,\dots,G_d,Z$ is not trivial. In a hard or soft thresholding iteration one would have to find a best approximation A^G to a given tensor Z, and in the matrix case $d = 2$ this is expensive but possible. For tensors in $d > 2$ such a best approximation is not available. A common technique for finding something close to a best approximation is an alternating fit cycling through the unknowns G_μ. But since our final goal is not the approximation of Z but the minimization of f, it makes sense to directly solve the nonlinear system by an alternating fit. We approach this nonlinear system by a nonlinear block Gauß-Seidel iteration where the blocks of unknowns are $G_1,\dots,G_d,Z$:

Require: Initial guess A^G
  for i=1,... **do**
    For all $i \in \mathcal{I} \setminus P$ set $Z_i := A^G_i$ and for all $i \in P$ set $Z_i := M_i$
    For all $\mu \in D$ minimize $\|Z - A^G\|_F$ with respect to G_μ
  end for

Finally, we use (partial) successive overrelaxation in order to speed up the convergence, and we change the order of optimization of blocks to be $Z, G_1, \dots, G_h, Z, G_d, \dots, G_h$ for some fixed $h \in D$. Altogether this defines the basic proceeding of the alternating directions fit (ADF) Algorithm 1 by which we solve the tensor completion problem.

1.5. Organization of the Article. In Section 2 we introduce the necessary basic tools for the analysis and algorithmic treatment of the tensor approximation problem. Section 3 presents the ADF algorithm in detail and analyses the computational and storage complexity of one iterative step. Several practical issues like adaptive choice of the ranks, improved performance, and stopping criteria are considered. In the numerical examples Section 4 we apply our algorithm to two classes of examples: a) smooth function related tensors, and b) random low rank tensors with and without noise. We end with a conclusion in Section 5.

2. Optimization in TT-Format.

2.1. Matrix Fitting. We first present a central tool, by which a k-rank approximation of a sparsely known matrix can be obtained quite efficiently, cf. [21].

Algorithm 1 Basic proceeding of the ADF algorithm

Require: Initial guess A^G, overrelaxation parameter $\alpha \geq 1$
  while breaking condition not fulfilled **do**
    For all $i \in \mathcal{I} \setminus P$ set $Z_i := A_i^G$ and for all $i \in P$ set $Z_i := M_i$
    for $\mu = 1, \ldots, h$ **do**
      Determine $G_\mu^+ := \operatorname{argmin}_{G_\mu} \|Z - A^G\|_F$ and set $G_\mu := G_\mu + \alpha(G_\mu^+ - G_\mu)$
    end for
    For all $i \in \mathcal{I} \setminus P$ set $Z_i := A_i^G$ and for all $i \in P$ set $Z_i := M_i$
    for $\mu = d, \ldots, h$ **do**
      Determine $G_\mu^+ := \operatorname{argmin}_{G_\mu} \|Z - A^G\|_F$ and set $G_\mu := G_\mu + \alpha(G_\mu^+ - G_\mu)$
    end for
  end while

THEOREM 2.1 (LMaFit principle). *Let $W \in \mathbb{R}^{n\times m}$ and $Y \in \mathbb{R}^{k\times m}$ be given matrices and Y of full rank k. Define $X \in \mathbb{R}^{n\times k}$ as the minimizer*

$$X := \operatorname*{argmin}_{\tilde{X}} \|W - \tilde{X}Y\|_F$$

Then we gain

$$\operatorname{range}(WY^T) = \operatorname{range}(X).$$

Proof. ([21]) Using the SVD of $Y = U\Sigma V^T$ we have the pseudo inverse $Y^+ = V\Sigma^+U^T$. Hence $X = WV\Sigma^+U^T$, while $WY^T = WV\Sigma U^T$. Thereby $\operatorname{range}(WY^T) = \operatorname{range}(X)$. □

Fortunatly, the LMaFit principle also proves to be a very useful tool in the approximation of a sparsely known tensor of TT-ranks $r_1, \ldots, r_{d-1}$ with the help of the representation G. It takes, however, some preparation until its use can be justified.

2.2. Statement of the Lefthand and Righthand Approximation Problem. For practical reasons (based on the LMaFit principle) it is most efficient if in Algorithm 1 the blocks $G_1, \ldots, G_h$ or $G_d, \ldots G_h$ are optimized consecutively, where the direction h can be anything between 1 and d. Since we would like to update Z not only after all blocks G_μ are optimized but earlier, we choose

$$h := \lfloor (d+1)/2 \rfloor. \tag{2.1}$$

PROBLEM 2.2 (Lefthand problem). *Given a tensor $Z \in \mathbb{R}^{\mathcal{I}}$ (all entries known), ranks $\boldsymbol{r}$, and a starting representation G, replace for $\mu = 1, \ldots, h$ the entries of G_μ such that $G_\mu = \operatorname{argmin}_{G_\mu} \|Z - A^G\|_F$.* In order to avoid confusion we want to remark that the task for the lefthand problem is neither to find the minimizer $(G_1, \ldots, G_d)$ nor the partial minimizer $(G_1, \ldots, G_h)$. Instead, we seek consecutively the partial minimizers in the stated order

$$G_1 := \operatorname*{argmin}_{G_1} \|Z - A^G\|_F,\ G_2 := \operatorname*{argmin}_{G_2} \|Z - A^G\|_F,\ \ldots,\ G_h := \operatorname*{argmin}_{G_h} \|Z - A^G\|_F.$$

PROBLEM 2.3 (Righthand problem). *Given a tensor $Z \in \mathbb{R}^{\mathcal{I}}$ (known at all points), ranks $\boldsymbol{r}$, and a starting representation G, replace for $\mu = d, \ldots, h$ the entries of G_μ such that $G_\mu = \operatorname{argmin}_{G_\mu} \|Z - A^G\|_F$.*

2.3. Tensor Calculus. In this section we introduce the neccessary tools to work with matrix blocks in order to construct and prove the core step of the ADF algorithm (Theorem 3.1).

DEFINITION 2.4 (Matrix block). *Let $k_1, k_2, n \in \mathbb{N}$. We define a matrix block $H \in (\mathbb{R}^{k_1 \times k_2})^n$ as a vector of matrices $H(1), \dots, H(n) \in \mathbb{R}^{k_1 \times k_2}$. We call $k_1 \times k_2$ the dimension and n the length of H.*

REMARK 2.5 (Algebraic properties). *Let $k_1, k_2 \in \mathbb{N}$ be fixed. Then the set of matrix blocks $H \in (\mathbb{R}^{k_1 \times k_2})^n$ forms an $\mathbb{R}$-vectorspace. Additionally, it can also be viewed as left-module over the non-abelian matrix ring $\mathbb{R}^{k_1 \times k_1}$ as well as right-module over $\mathbb{R}^{k_2 \times k_2}$.* Next we introduce a tool to combine multiple matrix blocks into one, which is ultimately used for the definition of the TT representation A^G.

DEFINITION 2.6 ((Kronecker) product between matrix blocks). *We define the (Kronecker) product $\otimes$ for matrix blocks H_1, H_2 of dimensions $k_1 \times k_m, k_m \times k_2$ and lengths n_1, n_2 as*

$$(H_1 \otimes H_2)((i,j)) := H_1(i) H_2(j)$$

where $(H_1 \otimes H_2)$ is a matrix block of dimension $k_1 \times k_2$ and length $n_1 n_2$. Note that this is analogous to the conventional Kronecker product, but $(H_1 \otimes H_2)^T$ does neither equal $H_1^T \otimes H_2^T$ nor $H_2^T \otimes H_1^T$. In order to simplify the notation we use the following convention:

- *We treat the product of a matrix and a matrix block as if the matrix was a block of length 1 and skip the $\otimes$. It is referred to as pointwise multiplication.*
- *We write $(H_1 \otimes \dots \otimes H_n)(i_1 \dots i_n)$ instead of $(H_1 \otimes \dots \otimes H_n)((i_1 \dots i_n))$.*
- *The empty Kronecker product is defined to be I, the identity matrix of suitable size.*

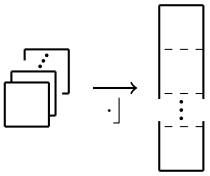
REMARK 2.7 (Generating A^G). *Using the Kronecker product, one can express A^G by*

$$A^G_{(i_1,\dots,i_d)} = (G_1 \otimes \dots \otimes G_d)(i_1, \dots, i_d), \quad A^G = G_1 \otimes \dots \otimes G_d.$$

Next, we define operations to switch between matrix blocks, matrices, and tensors.

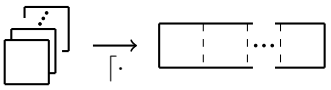
DEFINITION 2.8 (Lefthand block and righthand block). *Let $H \in (\mathbb{R}^{k_1 \times k_2})^n$ be a matrix block. We define the lefthand block $H\rfloor$ as*

$$H\rfloor := \begin{bmatrix} H(1) \\ H(2) \\ \vdots \\ H(n) \end{bmatrix} \in \mathbb{R}^{nk_1 \times k_2}, \qquad \xrightarrow{\cdot\rfloor} \tag{2.2}$$



and the righthand block $\lceil H$ as

$$\lceil H := \begin{bmatrix} H(1) & H(2) & \dots & H(n) \end{bmatrix} \in \mathbb{R}^{k_1 \times nk_2}. \qquad \xrightarrow{\lceil\cdot} \tag{2.3}$$



The transpose H^T of a matrix block is defined as $H^T(i) := H(i)^T$.

REMARK 2.9 (Conjugacy of block operations). *The lefthand block operation is conjugate to the righthand block operation by means of*

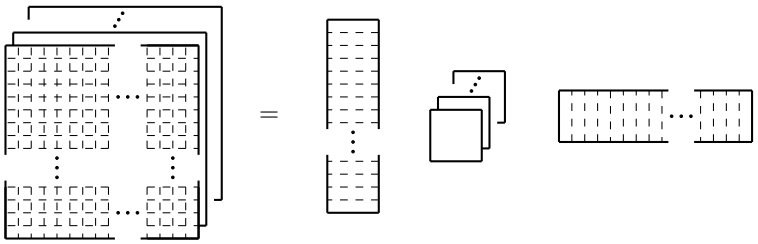
$$(\cdot)^T \circ \cdot\rfloor = \lceil\cdot \circ (\cdot)^T.$$

DEFINITION 2.10 (Block matricization). *Let $A \in \mathbb{R}^{\mathcal{I}}$ be a d-dimensional tensor. A block matricization with respect to $s \in \{1, \dots, d\}$, $A_{(s)}$, is defined as the matrix block of dimension $(n_1 \dots n_{s-1}) \times (n_{s+1} \dots n_d)$ and length n_s, given by*

$$(A_{(s)}(i_s))_{(i_1,\dots,i_{s-1}),(i_{s+1},\dots,i_d)} := A_{i_1,\dots,i_d}, \quad \forall i_s \in \mathcal{I}_s.$$

EXAMPLE 2.11 (Block matricization of A^G). *For a better understanding one may have a look at the following equation, which is valid for any s and representation G:*

$$A^G_{(s)} = (G_1 \otimes \dots \otimes G_{s-1})\rfloor \; G_s \; \lceil (G_{s+1} \otimes \dots \otimes G_d).$$



Note that both the multiplication with the lefthand as well as righthand block is elementwise, i.e. $A_{(s)}(i_s)_{(i_1,\dots,i_{s-1}),(i_{s+1},\dots,i_d)} = (G_1(i_1) \cdots G_{s-1}(i_{s-1}))\; G_s(i_s)\; (G_{s+1}(i_{s+1}) \cdots G_d(i_d))$.

For further calculations we give a scalar product for matrix blocks and derive a norm, which is essentially the Frobenius norm.

DEFINITION 2.12 ((Scalar) product of matrix blocks). *Let G and H be matrix blocks of the dimensions $k_1 \times k_m$,$k_m \times k_2$ and same length. Then we define the (scalar) product of two matrix blocks as*

$$\langle G, H \rangle := \sum_i G(i)H(i) = \lceil GH \rfloor \in \mathbb{R}^{k_1 \times k_2}.$$

For a matrix $J \in \mathbb{R}^{k_m \times k_m}$ we define

$$\langle G, J, H \rangle := \langle GJ, H \rangle = \langle G, JH \rangle.$$

Note that $\langle \cdot, \cdot \rangle$ is only a product with scalar output regarding its module properties.

DEFINITION 2.13 ($\mathbb{R}$-scalar product and matrix block norm). *Let $V := (\mathbb{R}^{k_1 \times k_2})^n$ be the $\mathbb{R}$-Vectorspace of matrix blocks of dimension $k_1 \times k_2$ and length n. Then $\langle \cdot, \cdot \rangle$ defines a scalar product $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ on V via*

$$\langle G, H \rangle_{\mathbb{R}} := \operatorname{trace}\langle G, H^T \rangle = \operatorname{trace}\langle G^T, H \rangle, \qquad G, H \in V.$$

The corresponding norm $|| \cdot ||$ on V is defined as $\|G\| := \sqrt{\langle G, G \rangle_{\mathbb{R}}}$.

REMARK 2.14 (Properties of the matrix block norm). *For a matrix block G, tensor A and index $s \in D$ holds*

$$\|G\| = \sqrt{\sum_i ||G(i)||_F^2}, \qquad \|A\|_F = \|A_{(s)}\|.$$

We introduce the concept of orthogonality (cf. [6]) for matrix blocks, by which we can greatly simplify the minimization problem.

DEFINITION 2.15 (Orthogonality of matrix blocks). *For a matrix block H, we call H lefthand orthogonal if the columns of $H\rfloor$ are orthogonal (this being $\langle H^T, H \rangle = I$), and*

righthand orthogonal if the rows of $\lceil H$ *are orthogonal (this being* $\langle H, H^T\rangle = I$*).*

Let Q *be a matrix block of same dimensions as* H*. We then define the (non-unique) operation* $orth^\ell$ *such that for* $Q = orth^\ell(H)$*, the pair* $(Q\rfloor, R)$ *is a QR-decomposition of* $H\rfloor$*. Then* Q *is lefthand orthogonal and* $QR = H$*.*
Likewise $orth^r$ *is such that for* $Q = orth^r(H)$*, the pair* $(L, \lceil Q)$ *is an LQ-decomposition of* $\lceil H$*. Then* Q *is righthand orthogonal and* $LQ = H$*.* In the following we demonstrate how orthogonality, the scalar product and the Kronecker product act together and form a central result (Corollary 2.18) regarding the feasibility (Theorem 3.4) of the ADF core step (Theorem 3.1) for which we initially only need Corollary 2.17.

LEMMA 2.16 (Scalar products of Kronecker products). *Let* G_1, G_2 *and* H_1, H_2 *be matrix blocks of appropriate dimensions and lengths. Then*

$$\langle (G_1 \otimes G_2)^T, H_1 \otimes H_2\rangle = \langle G_2^T, \langle G_1^T, H_1\rangle, H_2\rangle,$$

respectively

$$\langle G_1 \otimes G_2, (H_1 \otimes H_2)^T\rangle = \langle G_1, \langle G_2, H_2^T\rangle, H_1^T\rangle.$$

Proof. Due to analogy we consider only the first case. By definition and reordering of summation we get

$$\begin{aligned}
&\langle (G_1 \otimes G_2)^T, H_1 \otimes H_2\rangle \quad = \sum_i ((G_1 \otimes G_2)(i))^T (H_1 \otimes H_2)(i) \\
&= \sum_{i_1,i_2} G_2(i_2)^T G_1(i_1)^T (H_1(i_1) H_2(i_2)) \quad = \sum_{i_2} G_2(i_2)^T \sum_{i_1} (G_1(i_1)^T H_1(i_1)) H_2(i_2) \\
&= \sum_{i_2} G_2(i_2)^T \langle G_1^T, H_1\rangle H_2(i_2) \quad = \langle G_2^T, \langle G_1^T, H_1\rangle, H_2\rangle.
\end{aligned}$$

□

COROLLARY 2.17 (Orthogonality of Kronecker products). *If two matrix blocks* G, H *are lefthand (righthand) orthogonal, then* $G \otimes H$ *is lefthand (righthand) orthogonal. This follows as for a matrix block* K *and a lefthand (righthand) orthogonal matrix block* Q *(*P*), we have*

$$\langle (Q \otimes K)^T, Q \otimes K\rangle = \langle K^T, K\rangle \quad \textit{and} \quad \langle K, K^T\rangle = \langle K \otimes P, (K \otimes P)^T\rangle, \textit{ respectively.}$$

Furthermore this implies

$$||Q \otimes K|| = ||K|| = ||K \otimes P||.$$

COROLLARY 2.18 (Scalar products of multiple Kronecker products). *Let* $G_1, \ldots, G_\ell$ *and* $H_1, \ldots, H_\ell$ *be matrix blocks of appropriate dimensions. Let* $G := G_1 \otimes \ldots \otimes G_\ell$ *and* $H := H_1 \otimes \ldots \otimes H_\ell$*. Then*

$$\langle G^T, H\rangle = \langle G_\ell^T, \ldots, \langle G_1^T, H_1\rangle, \ldots, H_\ell\rangle.$$

If $G_1 = H_1, \ldots, G_k = H_k$ *are lefthand orthogonal, we have*

$$\langle G^T, H\rangle \;=\; \langle G_\ell^T, \ldots, \langle G_{k+1}^T, H_{k+1}\rangle \ldots, H_\ell\rangle \;=\; \langle (G_{k+1} \otimes \ldots \otimes G_\ell)^T, H_{k+1} \otimes \ldots \otimes H_\ell\rangle.$$

Likewise

$$\langle G, H^T\rangle = \langle G_1, \ldots, \langle G_\ell, H_\ell^T\rangle, \ldots, H_1^T\rangle.$$

If $G_k = H_k, \ldots, G_\ell = H_\ell$ are righthand orthogonal, we have

$$\langle G, H^T\rangle \;=\; \langle G_1, \ldots, \langle G_{k-1}, H_{k-1}^T\rangle, \ldots, H_1^T\rangle \;=\; \langle G_1 \otimes \ldots \otimes G_{k-1}, (H_1 \otimes \ldots \otimes H_{k-1})^T\rangle.$$

Note that this decouples the Kronecker product and reduces the amount of computations as well as the involved dimensions. For shorter notation we set the empty scalar product (i.e. $\ell = 0$) as I, the identity matrix of suitable size. We offer a last remark on a property which is central to the ADF core step.

REMARK 2.19 (Non-uniqueness of representations). *Considering $A^{(\cdot)} : H \mapsto A^H$, one has a mapping which is non-injective as for the representations $H^1 : H_1, \ldots, (H_\ell J), \ldots, H_d$ and $H^2 : H_1, \ldots, (JH_{\ell+1}), \ldots, H_d$ one has $A^{H^1} = A^{H^2}$, where ℓ is arbitrary and J a square matrix of appropriate dimension. This degree of freedom is not a disadvantage, indeed we exploit this property in Theorem 3.1 by use of the LMaFit principle (Theorem 2.1). Furthermore one can always assume, using the operations $orth^\ell$ and $orth^r$, that all matrix blocks H_i, $i < h$ are lefthand orthogonal as well as that all matrix blocks H_i, $i > h$ are righthand orthogonal. Then also $||A^G||_F = ||G_h||$.*

3. The ADF Algorithm. We approach the main Problem 1.3 by the ADF Algorithm 1. As a basic step of the algorithm one has to solve the lefthand and righthand problem, i.e., optimize with respect to $G_1, \ldots, G_h$ or $G_d, \ldots, G_h$ while keeping the other variables fixed. In the following we give a detailed description for solving these problems, as well as for the choice of the overrelaxation parameter α (Remark 3.8 and Algorithm 2).

3.1. The ADF Core Step. The core step we now give solves the lefthand respectively righthand problem in the sense that $G_1, \ldots, G_h$, respectively $G_d, \ldots, G_h$, are each optimized individually as it is required for the update of G in Algorithm 1.

THEOREM 3.1 (Core step of the ADF algorithm). *Given the lefthand problem one can, as one sequence, optimize consecutively $G_1, \ldots, G_h$ in order $s = 1 \to h$, respectively $G_h, \ldots, G_d$ in order $s = h \leftarrow d$ when given the righthand problem. Without loss of generality we assume that $G_1 \ldots G_{h-1}$ are given lefthand orthogonal, respectively $G_{h+1} \ldots G_d$ righthand orthogonal (cf. Remark 2.19). The solution of the lefthand or righthand problem is given by*

$$G_s(\boldsymbol{j}) = \sum_{i \in \mathcal{I}, i_s = \boldsymbol{j}} Z_i (G_1(i_1) \ldots G_{s-1}(i_{s-1}))^T (G_{s+1}(i_{s+1}) \ldots G_d(i_d))^T = Q_s^T Z_{(s)}(\boldsymbol{j}) P_s^T$$

where

$$Q_s := (G_1 \otimes \ldots \otimes G_{s-1})\rfloor, \qquad P_s := \lceil (G_{s+1} \otimes \ldots \otimes G_d),$$

followed by

$$G_s \leftarrow \begin{cases} orth^\ell(G_s), & \text{if } s < h \\ G_s, & \text{if } s = h \\ orth^r(G_s), & \text{if } s > h. \end{cases}$$

Proof. (lefthand) We carry out the proof for $s < h$ and $s = h$, assuming that the steps before were performed as described above. Thereby $G_1 \ldots G_{s-1}$ are lefthand orthogonal.

We abbreviate $P := P_s$ and $Q := Q_s$. As Q is orthogonal we have

$$\begin{aligned} G_s &= \operatorname*{argmin}_{G_s} \|Z - A^G\|_F = \operatorname*{argmin}_{G_s} \|Z_{(s)} - A^G_{(s)}\| = \operatorname*{argmin}_{G_s} \|Z_{(s)} - QG_sP\| \\ &= \operatorname*{argmin}_{G_s} \|Q^TZ_{(s)} - G_sP\| = \operatorname*{argmin}_{G_s} \|(Q^TZ_{(s)})\rfloor - G_s\rfloor P\|_F \quad (LS), \end{aligned}$$

the latter being one single least squares problem to be solved for the matrix $G_s\rfloor$.
Case s < h : At this point we can, for $s < h$, use the LMaFit principle. We thereby know that

$$\operatorname{range}((Q^TZ_{(s)})\rfloor P^T)) = \operatorname{range}(G_s\rfloor),$$

which for some matrix R yields

$$(Q^TZ_{(s)})\rfloor P^TR = G_s\rfloor \Leftrightarrow Q^TZ_{(s)}(i_s)P^TR = G_s(i_s) \quad \forall i_s.$$

In principle we have found G_s, but we do certainly not want to calculate R. Therefore, we interprete R as part of G_{s+1} instead of G_s. This is a valid operation that does not change A^G, as

$$\ldots G_{s-1}(i_{s-1}) \left(Q^TZ_{(s)}(i_s)P^T\boldsymbol{R}\right) G_{s+1}(i_{s+1}) \ldots$$

$$= \ldots G_{s-1}(i_{s-1}) \left(Q^TZ_{(s)}(i_s)P^T\right) \left(\boldsymbol{R}G_{s+1}(i_{s+1})\right) \ldots \quad \forall (i_1, \ldots, i_d) \in \mathcal{I}.$$

Thereby R will vanish as the newly formed RG_{s+1} is next to be replaced. Hence we can, instead of the direct minimizer, choose

$$G_s(i_s) \leftarrow Q^TZ_{(s)}(i_s)P^T \quad \forall i_s. \qquad (*)$$

followed by lefthand orthogonalizing G_s. This again produces some R, which is treated likewise (cf. [6] for the matrix case).
Case s = h : One can see a more specific reason for the splitting into lefthand and righthand now. There is naturally a limit for R to be transferred to the right, at the latest G_d. This implies that we can neither use the LMaFit principle nor orthogonalize for one index, this being $s = h$. Yet in this case P is orthogonal. Starting at (LS) again, we can thereby without any further reasoning put P to the right as we did with Q. This immediately results in the same formula $(*)$, except that we do not orthogonalize. This completes the proof. □

The core step described above solves the lefthand (or righthand) problem without overrelaxation. The overrelaxation parameter α can be incorporated directly into the core step by modifying Z as follows.

LEMMA 3.2. *Let $A^G = G_1 \otimes \cdots \otimes G_d \in \mathbb{R}^{\mathcal{I}}$ be given, $\alpha \in \mathbb{R}$, $Z \in \mathbb{R}^{\mathcal{I}}$ and*

$$G^+_\mu := \operatorname*{argmin}_{\tilde{G}_\mu} \|Z - G_1 \otimes \cdots \otimes G_{\mu-1} \otimes \tilde{G}_\mu \otimes G_{\mu+1} \otimes \cdots \otimes G_d\|.$$

Then $G^\alpha_\mu := \alpha G^+_\mu + (1-\alpha)G_\mu$ fulfils

$$G^\alpha_\mu = \operatorname*{argmin}_{\tilde{G}_\mu} \|Z^\alpha - G_1 \otimes \cdots \otimes G_{\mu-1} \otimes \tilde{G}_\mu \otimes G_{\mu+1} \otimes \cdots \otimes G_d\|$$

for $Z^\alpha := \alpha Z + (1-\alpha)A^G$.

Proof. We insert $\alpha G_\mu^+ + (1-\alpha)G_\mu$ into the righthand side of the last equation and obtain

$$\begin{aligned}
&\|Z^\alpha - G_1 \otimes \cdots \otimes G_{\mu-1} \otimes (\alpha G_\mu^+ + (1-\alpha)G_\mu) \otimes G_{\mu+1} \otimes \cdots \otimes G_d\| \\
=&\|\alpha Z + (1-\alpha)A^G - G_1 \otimes \cdots \otimes G_{\mu-1} \otimes (\alpha G_\mu^+ + (1-\alpha)G_\mu) \otimes G_{\mu+1} \otimes \cdots \otimes G_d\| \\
=&\|\alpha Z - \alpha G_1 \otimes \cdots \otimes G_{\mu-1} \otimes G_\mu^+ \otimes G_{\mu+1} \otimes \cdots \otimes G_d\| \\
=&\alpha\|Z - G_1 \otimes \cdots \otimes G_{\mu-1} \otimes G_\mu^+ \otimes G_{\mu+1} \otimes \cdots \otimes G_d\|
\end{aligned}$$

and any other choice $\tilde{G}_\mu = \alpha\tilde{G}_\mu^+ + (1-\alpha)G_\mu$ does not give a smaller value since G_μ^+ is a minimizer. ∎

NOTATION 3.3 (Denoting current and old representations). *During one sweep of updating $G_1, \ldots, G_h$ in the lefthand problem, we use the notation G^- for the representation before the sweep and G for the updated representation. Analogously for the righthand problem.*

THEOREM 3.4 (Practical ADF core step (lefthand)). *Under the assumptions from Theorem 3.1 the update $G_1, \ldots, G_h$ for the lefthand problem with overrelaxation parameter α is given by*

$$G_s(\boldsymbol{j}) = \underbrace{\langle G_{s-1}^T, \ldots \langle G_1^T, G_1^- \rangle \ldots G_{s-1}^- \rangle}_{(LS_s^1)} \, G_s^-(\boldsymbol{j}) \, \underbrace{\langle G_{s+1}^- \ldots \langle G_h^-, {G_h^-}^T \rangle \ldots {G_{s+1}^-}^T \rangle}_{(LS_s^2)} \tag{3.1}$$

$$+ \sum_{\substack{i=(i_1,\ldots,i_d)\in P \\ i_s=\boldsymbol{j}}} \alpha(M_i - A_i^{G^-}) \, \underbrace{(G_1(i_1)\ldots G_{s-1}(i_{s-1}))^T}_{(LM_s^1)_i} \, \underbrace{(G_{s+1}^-(i_{s+1})\ldots G_d^-(i_d))^T}_{(LM_s^2)_i} \tag{3.2}$$

$$G_s \leftarrow \begin{cases} orth^\ell(G_s), & \text{if } s < h \\ G_s, & \text{if } s = h. \end{cases} \tag{3.3}$$

(The short notations are used for Lemma 3.5.)

Proof. According to Theorem 3.1 and Lemma 3.2, we have

$$\begin{aligned}
G_s(\boldsymbol{j}) &= Q_s^T Z_{(s)}^\alpha(\boldsymbol{j}) P_s^T \\
&= \sum_{\substack{i=(i_1,\ldots,i_d)\in\mathcal{I} \\ i_s=\boldsymbol{j}}} Z_i^\alpha (G_1(i_1)\ldots G_{s-1}(i_{s-1}))^T (G_{s+1}(i_{s+1})\ldots G_d(i_d))^T \quad \forall \boldsymbol{j}.
\end{aligned} \tag{3.4}$$

$Z = A^{G^-}|_{\mathcal{I}\setminus P} + M|_P$ (cf. Algorithm 1) and $Z^\alpha = \alpha Z + (1-\alpha)A^{G^-}$ (cf. Lemma 3.2) yield

$$Z^\alpha = \underbrace{A^{G^-}}_{\hookrightarrow \text{First summand}} + \underbrace{\alpha(M|_P - A^{G^-}|_P)}_{\hookrightarrow \text{Second summand}}$$

which we insert it into (3.4).

First summand: Recall that $A_{(s)}^{G^-}$ can be expanded (Example 2.11). Using the definition for Q_s and P_s (Theorem 3.1) we have:

$$\begin{aligned}
&Q_s^T A_{(s)}^{G^-}(\boldsymbol{j}) P_s^T = \\
&(G_1 \otimes \ldots \otimes G_{s-1})\rfloor^T (G_1^- \otimes \ldots \otimes G_{s-1}^-)\rfloor \, G_s^-(\boldsymbol{j}) \, \lceil (G_{s+1}^- \otimes \ldots \otimes G_d^-)\rceil \lceil (G_{s+1} \otimes \ldots \otimes G_d)^T
\end{aligned} \tag{3.5}$$

With Remark 2.9 and the definition for the scalar product (Definition 2.12) we can simplify

$$(3.5) = \langle (G_1 \otimes \ldots \otimes G_{s-1})^T, (G_1^- \otimes \ldots \otimes G_{s-1}^-) \rangle \, G_s^-(i_s) \, \langle (G_{s+1}^- \otimes \ldots \otimes G_d^-), (G_{s+1} \otimes \ldots \otimes G_d)^T \rangle$$

Due to Corollary 2.18 and the fact that $G_{s+1}, \ldots, G_d$ still equal $G_{s+1}^-, \ldots, G_d^-$ and are righthand orthogonal as of index $h+1$ when optimizing lefthand we can simplify both scalar products

$$(3.5) = \langle G_{s-1}^T, \ldots \langle G_1^T, G_1^- \rangle \ldots G_{s-1}^- \rangle \, G_s^-(\boldsymbol{j}) \, \langle G_{s+1}^- \ldots \langle G_h^-, {G_h^-}^T \rangle \ldots {G_{s+1}^-}^T \rangle.$$

Second summand: As $(M|_P - A^{G^-}|_P)_i = 0$ for any $i \notin P$, we can reduce the summation from $\mathcal{I}$ to P without changing its outcome. We use again that $G_{s+1}, \ldots, G_d$ still equal $G_{s+1}^-, \ldots, G_d^-$ and receive the formula stated in the theorem. □

Analogously to the lefthand case we can treat the righthand case by

$$\begin{aligned} G_s(\boldsymbol{j}) = {} & \langle {G_{s-1}^-}^T, \ldots \langle {G_h^-}^T, G_h^- \rangle \ldots G_{s-1}^- \rangle \, G_s^-(\boldsymbol{j}) \, \langle G_{s+1}^- \ldots \langle G_d^-, G_d^T \rangle \ldots G_{s+1}^T \rangle \\ & + \sum_{\substack{i=(i_1,\ldots,i_d) \in P \\ i_s = \boldsymbol{j}}} \alpha(M_i - A_i^{G^-}) \, (G_1^-(i_1) \ldots G_{s-1}^-(i_{s-1}))^T (G_{s+1}(i_{s+1}) \ldots G_d(i_d))^T \end{aligned}$$

$$G_s \leftarrow \begin{cases} orth^r(G_s), & \text{if } s > h \\ G_s, & \text{if } s = h. \end{cases}$$

3.2. Computational Complexity of ADF. Each statement in this section which we give in its lefthand version can be treated analogously in the righthand case, leading to equivalent results.

LEMMA 3.5 (Successive computing). *In the lefthand core step (Theorem 3.4) occuring terms can be reduced to simpler successive computations, via:*

$$(LS_s^1) = \langle G_{s-1}^T, (LS_{s-1}^1), G_{s-1}^- \rangle, \tag{3.6}$$

$$(LS_s^2) = \langle {G^-}_{s+1}^T, (LS_{s+1}^2), G_{s+1}^- \rangle, \tag{3.7}$$

where $(LS_1^1) = 1$ *and* $(LS_h^2) = I$ *(the identity matrix). Likewise*

$$(LM_s^1)_i = G_{s-1}(i_{s-1})^T \; (LM_{s-1}^1)_i, \tag{3.8}$$

$$(LM_s^2)_i = (LM_s^2)_i \, G_{s+1}^-(i_{s+1})^T \tag{3.9}$$

where $(LM_1^1) = 1$ *and* (LM_h^2) *is inherited from the righthand step. Hence, while* (LS^1) *and* (LM^1) *are updated within the sequence,* (LS^2) *and* (LM^2) *are calculated before. Furthermore,* (LM_h^1) *and* (LM_h^2) *can be used to update* $A_P^{G^-}$.

COROLLARY 3.6 (Computational complexity). *Let* $r := max\{r_1, \ldots, r_{d-1}\}$, $n := max\{n_1, \ldots, n_d\}$ *and* $p := \#P$. *If we consider one full lefthand optimization sweep (*Z, G_1, …, G_h*), we can give an upper bound for the effort by analyzing the operations in Lemma 3.5 and Theorem 3.4:*

1. (3.6) *&* (3.7)*: 2n times* $(r \times r)$ *times* $(r \times r)$ *matrix multiplications*
2. (3.8) *&* (3.9)*: p times* $(1 \times r)$ *times* $(r \times r)$ *matrix multiplications*
3. (3.1)*: n times* $(r \times r)$ *times* $(r \times r)$ *matrix multiplications*
4. (3.2)*: p times* $(r \times 1)$ *times* $(1 \times r)$ *matrix multiplications*
5. (3.3)*: at most one QR decomposition of an* $nr \times r$ *matrix*

Each of these steps is performed about $d/2$ times. As the number of samples scales like $p \sim C_{ov} r^2 nd$, items 1, 3, 5 are negligible. This leaves us with a total effort for one lefthand sweep of

$$\mathfrak{E} = \mathcal{O}(pr^2 d) = \mathcal{O}(C_{ov} r^4 n d^2),$$

where the oversampling constant C_{ov} is assumed to be the uniform slice density, cf. Definition 1.1. Note that the constant hidden in the order is very small, but the number of necessary lefthand and righthand sweeps can be large.

COROLLARY 3.7 (Storage complexity). *Let $r := \max\{r_1, \ldots, r_{d-1}\}$ and $p := \#P$ and $n := \max\{n_1, \ldots, n_d\}$. As in the lefthand case (LS^2) and (LM^2) have to be calculated beforehand and stored for efficiency reasons, we require*

$$\mathfrak{S} = \mathcal{O}(prd) = \mathcal{O}(C_{ov} r^3 n d^2)$$

numbers to be stored.

3.3. Choice of the SOR Parameter α. By an optimized determination of the acceleration parameter α one can speed up the convergence of the ADF algorithm considerably. Therefore, after each sweep for the lefthand or righthand problem, we allow a relatively expensive search for a suitable α by testing increased (α^{up}) and reduced (α^{down}) values of α until the residual decays (or we break).

REMARK 3.8 (Determination of the direction). *To handle the acceleration parameter α, we introduce a second parameter δ taking the role of an increment. Each lefthand or righthand sweep is run for two different accelerations $(\alpha^{up}, \alpha^{down})$, where we denote by α the previous one:*

$$\alpha^{up} := \alpha + \delta, \quad \alpha^{down} := max\{1, \alpha - \delta/5.\}$$

By this choice the overrelaxation parameter is at least $\alpha \geq 1$. Depending on the residuals of the results, one of the three directions, denoted by (up), (down) or (back), is chosen as specified in Algorithm 2. It determines the new α, δ as well as G. At the start we use $\alpha := 1$ (possibly adapted to the problem) and $\delta := \alpha/4$.

Algorithm 2 Choice of the SOR parameter α

Let G^- be the old representation received by having used α. Let G^{up}, G^{down} denote the representation obtained by either a lefthand or righthand sweep with parameter $\alpha^{up}, \alpha^{down}$ leading to residuals R^-, R^{up}, R^{down}.

Notation: $\searrow$= decrease, $\searrow\searrow$= heavy decrease, $\nearrow$= increase

if $R^{up} > R^-$ and $R^{down} > R^-$ **then**
  $\searrow\searrow$ α (but $\alpha \geq 1$); $\searrow$ δ; keep $G \leftarrow G^-$ and repeat step; $\Gamma \leftarrow$ (back);
else if $(R^{up} < R^{down})$ **then**
  If $\Gamma =$ (up) then $\nearrow$ δ, otherwise $\searrow$ δ;
  $\alpha \leftarrow \alpha^{up}$; $G \leftarrow G^{up}$; $\Gamma \leftarrow$ (up);
else if $(R^{down} < R^{up})$ **then**
  If $\Gamma =$ (down) then $\nearrow$ δ, otherwise $\searrow$ δ;
  $\alpha \leftarrow \alpha^{down}$; $G \leftarrow G^{down}$; $\Gamma \leftarrow$ (down);
end if

3.4. Generating Starting Values and Stopping Criteria. We continue by introducing an incremental approach for finding good initial values for the iterative optimization by the ADF algorithm.

REMARK 3.9 (Stepwise increase of r). *A successful strategy for finding good initial values for the optimization is to start with minimal rank $r_0 = \ldots = r_d = 1$, and each time the algorithm does not do sufficient progress anymore (cf. Remark 3.10), the ranks r_μ of G are increased until the target rank is reached. For the very first guess with TT ranks $r_1 = \ldots = r_{d-1} = 1$, we define*

$$(G_s(i))_{1,1} := \frac{1}{\sqrt{n}}, \quad \forall s, i.$$

G is uniform and fulfils all required orthogonality conditions, cf. Theorem 3.4. Finally, we need a reliable stopping criterion and decide when to increase the ranks.

REMARK 3.10 (Breaking criteria). *We break if one of the following criteria is met:*

- *the last 10 directions were (back), i.e., there is no reduction of the residual even if the SOR parameter α approaches 1.*
- *the last 5 quotients of old and new residuals are in average too small, and either the direction is (down) or the change in the two last quotients of residuals is too small.*

The term "too small" is related to a parameter ε. Most reasonable choices vary between 10^{-4} and 10^{-6}. The first criterion prevents that the algorithm gets stuck trying to lower α more and more. In the second one, instead of using some kind of absolute criteria regarding the residual, it is much better to look at relative improvements. As these can fluctuate a bit we take the average of several. In order to prevent too early breaking, the direction is supposed to be (down), as while α increases (in case of (up)) or is accidentially far too large (in case of (back)), there is still a chance that this improves. However, if this happens too slow, the algorithm will break anyway to prevent that the algorithm gets stuck. The final ADF algorithm is given in Algorithm 3

Algorithm 3 ADF algorithm

```
Initialize the representation G for r = 1 (Remark 3.9), α and δ;
while r ≤ r_final do
  for iter = 1, ..., iter_max do
    do the while loop lefthand, then righthand: {Remark 3.8}
    while Γ = (back) do
      for (up) and (down) direction optimize respective hand; {Theorem 3.4, Lemma
      3.5}
      choose direction; {Algorithm 2}
    end while
    if breaking criteria apply then
      stop and continue outer for loop; {Remark 3.10}
    end if
  end for
  adapt representation to r + 1; {Remark 3.9}
end while
```

4. Numerical Experiments.

4.1. Generating of Data and Details to Measurements. In order to obtain a sufficient slice density, cf. Definition 1.1, we generate the set P in a quasi-random way

as follows: For each direction $\mu = 1, \dots, d$ and each index $i_\mu \in \mathcal{I}_\mu$ we pick $C_{ov} r^2$ indices $i_1, \dots, i_{\mu-1}, i_{\mu+1}, \dots, i_d$ at random (uniformly). This gives in total $dnC_{ov}r^2$ samples (excluding some exceptions), where C_{ov} is the oversampling factor in the sense that the low rank tensor format inhibits at most dnr^2 degrees of freedom.

As a control set C we use a set of the same cardinality as P that is generated in the same way. We give neither a limit to time nor to the number of iterations and use only the previously mentioned breaking criteria with varying $\varepsilon \in \{10^{-4}, 10^{-5}, 10^{-6}\}$. For the results of the tests we denote the ratio of known points $\rho = \#P/n^d$, the relativ residual $res_P = \|A - X\|_P / \|A\|_P$, the error on the control set $res_C = \|A - X\|_C / \|A\|_C$ and the *time* in seconds (s), minutes (m) or hours (h).

4.2. Approximation of a Full Rank Tensor with Decaying Singular Values. As a first example, we consider a tensor $A \in \mathbb{R}^{\mathcal{I}}$ given by the entries

$$A_{(i_1,\dots,i_d)} := \left(\sum_{\mu=1}^{d} i_\mu^2 \right)^{-1/2}.$$

We carry out three different tests, each one focusing on a different parameter, i.e. d (dimension), r (final rank) and n (size).

Each combination of parameters is tested 10 times for different random P and C, where $\langle res_C \rangle$ and $\langle res_P \rangle$ denote the geometric mean of the respective results and $\langle time \rangle$ the arithmetic mean of times. A plot of the convergence of $\langle res_P \rangle, \langle res_C \rangle$ for fixed $d = 8$ and $n = 20$ is given in Figure 4.1. We observe convergence for all choices of parameters. In Table 4.1 we list the detailed results for oversampling factor $C_{ov} = 10$ and stopping criterion $\varepsilon = 10^{-6}$.

r	$\langle res_C \rangle$	$\langle res_P \rangle$	$\langle time \rangle$
2	1.94e-02	8.15e-03	14 s
3	2.05e-03	3.84e-04	1.3 m
4	1.72e-03	7.50e-05	24 m
5	1.49e-03	4.50e-05	39 m
6	2.25e-03	1.84e-05	1.7 h
7	8.53e-04	7.36e-06	2.6 h

d	$\langle res_C \rangle$	$\langle res_P \rangle$	$\langle time \rangle$
5	1.30e-02	2.07e-03	6.3 s
8	2.98e-03	4.15e-04	32 s
13	3.07e-03	6.02e-04	2.9 m
21	6.29e-03	4.03e-03	4.3 m
34	7.66e-03	7.59e-03	27 s
55	4.63e-03	4.56e-03	2.1 m

TABLE 4.1

Convergence and timing with respect to the target rank r for fixed $d = 8, n = 20, C_{ov} = 10, \varepsilon = 10^{-6}$ (left). Convergence and timing with respect to the dimension d for fixed $r = 3, n = 8, C_{ov} = 10, \varepsilon = 10^{-6}$ (right).

We observe that the timing results can be quite inhomogenous due to the stopping criterion that does not limit the maximal number of iterative steps. However, the results show that an approximation in dimension $d = 55$ still works fine, although only extremely few points of the whole tensor (less than $10^{-50} \times n^d$) are known.

At last we consider the variation of the mode sizes $n \in \{8, 16, 32\}$ in Table 4.2. Here we also observe an independence of the mode size n, but we could observe in further numerical experiments that for large n and larger ranks r the results are less satisfying.

4.3. Reconstruction of a Low Rank Tensor without Noise. As second example, we consider quasi-random tensors with exact, common low TT ranks $A \in TT(r, \dots, r)$ (cf.

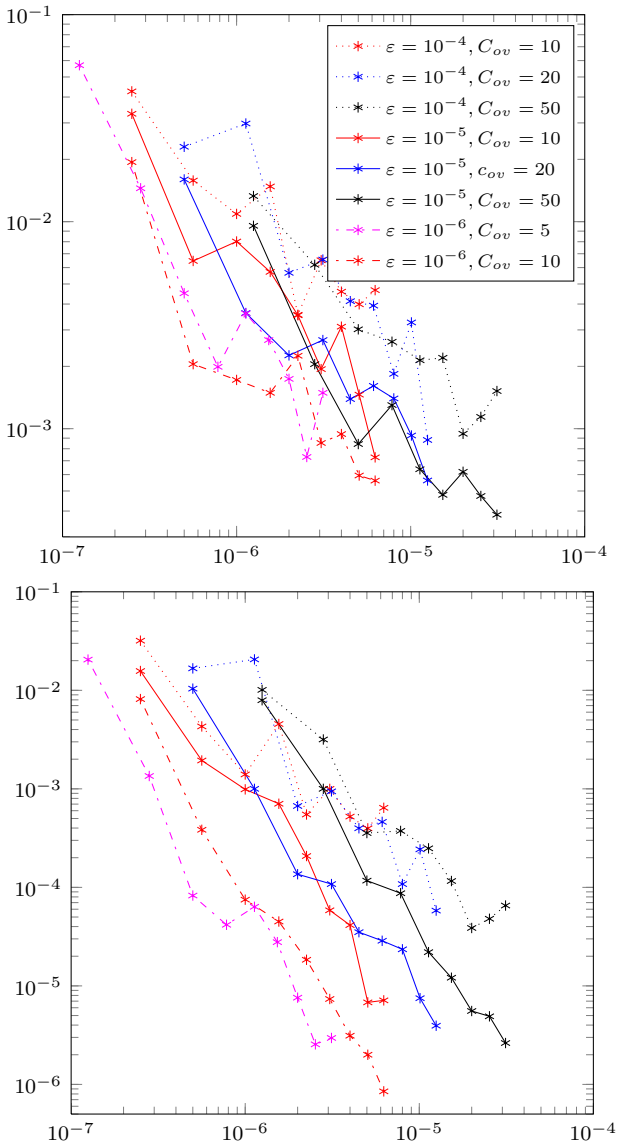


FIG. 4.1. *Plotted are the residuals* $\langle res_P \rangle$ *(right) as well as the control residuals* $\langle res_C \rangle$ *(left) as function of* ρ *for varying target ranks* $r = 2, \ldots, 10$ *. Each curve corresponds to one choice of the oversampling parameter* C_{ov} *and the stopping parameter* ε*.*

Definition 1.2). Each tensor is generated via a TT representation G as follows: Each entry of $G_1, \ldots, G_d$ is (uniformly) assigned with a random value in $[-0.5, 0.5]$. The tensor is then computed as $A = A^G$. Each combination of parameters is tested 20 times for different random P and C. We consider such a reconstruction successful if $res_C < 10^{-3}$. As an additional heuristic method we allow the algorithm after a failure to try the same task at most 5 times again. In these tests only portions of the previous P (which increase with the current rank) are used. However, no new information is anyhow used in these trials. Two diagrams, for $d = 4, 5$, with fixed $n = 20$ and $\varepsilon = 10^{-5}$ are given in Figure 4.2, displaying

n	$\langle res_C \rangle$	$\langle res_P \rangle$	$\langle time \rangle$
8	2.91e-03	1.56e-04	20 s
16	7.13e-03	3.87e-04	1.4 m
32	6.82e-03	4.01e-04	2.8 m

TABLE 4.2
Convergence and timing with respect to the mode size n for fixed $d = 8, r = 4, C_{ov} = 10, \varepsilon = 10^{-5}$.

the numbers of successes in 20 shades of gray, from white (0) to black (all 20).

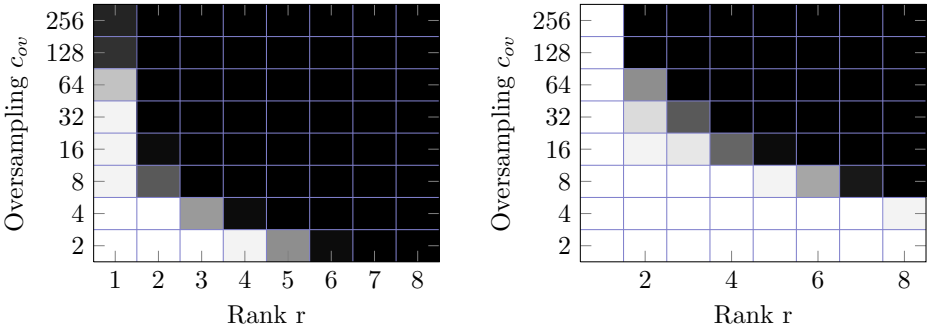


FIG. 4.2. *Displayed as shades of gray (white (0) to black (all 20)) are the numbers of successful reconstructions for varying target ranks $r = 1, \ldots, 8$ and oversamplings $C_{ov} = 2, 4, \ldots, 256$ with $d = 4$ (left) and $d = 5$ (right)*

We observe that the results for $d = 4$ are quite satisfying and that with increasing rank, less oversampling is needed. For $d = 5$, the performance is decreasing rather fast. In general the known portion of points ρ seems to effort a magnitude of about 10^{-3} to 10^{-2}.

4.4. Reconstruction of a Low Rank Tensor with Noise. As last test we exactly repeat the previous one but with perturbed tensors $\tilde{A} = A + 10^{-4}\nu\mathcal{E}$, where A is generated as before and $\nu := \|A\|_P/\sqrt{\#P}$. $\mathcal{E}$ is a tensor of same proportions; each entry being a (uniformly) quasi-randomly assigned value in $[-1, 1]$. A test is considered successful if $res_C < 10^{-3}$, where the control set residuum is still evaluated on A. However, no information about the non perturbed tensor is used in the algorithm. We desist from a diagram as there is no observable difference. It seems as if the algorithm ignores uniform noise as far as theoretically possible.

5. Conclusions. In this article we presented an alternating directions fitting algorithm that aims at finding a low tensor rank approximation to a tensor whose entries are known only in a small subset of all indices. It is important to use a certain oversampling factor in order to obtain a reasonable reconstruction of the tensor, and in our numerical experiments it turns out that this factor depends on the dimension but can be decreased with increasing rank, where it approaches $C_{ov} = 2$ in dimension $d = 4$. The SOR-type solver of the first order optimality conditions is able to minimize the residual on the known set of indices (the samples) for moderate ranks $r < 10$ and dimension $d < 50$. A modification or extension is necessary in order to treat varying TT ranks $r_1, \ldots, r_{d-1}$ instead of a uniform rank. Also, large mode sizes $n > 100$ possibly require smoothness conditions and a refined sampling strategy. The influence of noise on the reconstruction is rather harmless, where the noise can be unstructured or of rank structure but of smaller magnitude than the desired target accuracy. It seems that the low rank format introduces an automatic

regularization in the same way as the singular value truncation pronounces low frequency components.

REFERENCES

[1] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, *Scalable tensor factorizations for incomplete data*, Chemometrics and Intelligent Laboratory Systems, 106 (2011), pp. 41–56.

[2] Jonas Ballani, Lars Grasedyck, and Melanie Kluge, *Black box approximation of tensors in hierarchical Tucker format* , Linear Algebra Appl., 438 (2013), pp. 639–657.

[3] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, IOP Science, 27 (2011), p. 025010.

[4] Lars Grasedyck, *Hierarchical Singular Value Decomposition of Tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.

[5] Wolfgang Hackbusch and Stefan Kühn, *A new scheme for the tensor representation*, The journal of Fourier analysis and applications, 15 (2009), pp. 706–722.

[6] Sebastian Holtz, Thorsten Rohwedder, and Reinhold Schneider, *The alternating linear scheme for tensor optimization in the tensor-train format*, SIAM J. Sci. Comput., 34 (2012), pp. A683–A713.

[7] M. Kluge, *Sampling rule for tensor reconstruction in hierarchical tucker format*. in preparation, 2013.

[8] Tamara G. Kolda and Brett W. Bader, *Tensor decompositions and applications*, SIAM Review, 51 (2009), pp. 455–500.

[9] D. Kressner, M. Steinlechner, and B. Vandereycken, *Low-rank tensor completion by Riemannian optimization*, tech. report, École Polytechnique Fédérale de Lausanne. Technical Report 19.2013.

[10] Akshay Krishnamurthy and Aarti Singh, *Low-rank matrix and tensor completion via adaptive sampling*. arXiv:1304.4672, April 2013.

[11] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, 2009 IEEE 12th International Conference on Computer Vision (ICCV), (2009), pp. 2114–2121.

[12] Y. Liu and F. Shang, *An efficient matrix factorization method for tensor completions*, IEEE Signal Process. Lett., 20 (2013), pp. 307–310.

[13] Ivan Oseledets, *Tensor-train decomposition* , SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.

[14] Ivan V. Oseledets and Eugene E. Tyrtyshnikov, *Breaking the Curse of Dimensionality, Or How to Use SVD in Many Dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759.

[15] ———, *TT-cross approximation for multidimensional arrays*, Linear Algebra Appl., 432 (2010), pp. 70–88.

[16] H. Rauhut, R. Schneider, and Z. Stojanac, *Low rank tensor tensor recovery via iterative hard thresholding*. SampTA 2013, 10th International Conference on Sampling Theory and Application, Jacobs University Bremen, 2013.

[17] M. Signoretto, Q. Tran Dinh, L. De Lathauwer, and J.A.K. Suykens, *Learning with tensors: a framework based on convex optimization and spectral regularization*, tech. report, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2011. Accepted for publication in Machine Learning., Lirias number: 397075.

[18] C. Da Silva and F. J. Herrmann, *Hierarchical Tucker tensor optimization - applications to tensor completion*. SampTA 2013, 10th International Conference on Sampling Theory and Application, Jacobs University Bremen, 2013.

[19] G. Tomasi and R. Bro, *Parafac and missing values*, Chemometrics and Intelligent Laboratory Systems, 75 (2005), pp. 163–180.

[20] Guifre Vidal, *Efficient classical simulation of slightly entangled quantum computation*, Phys. Rev. Lett., 91 (2003), p. 147902.

[21] Z. Wen, W. Yin, and Y. Zhang, *Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm*, tech. report, Rice University CAAM, 2010. Technical Report 10-07.

[22] Steven R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett., 69 (1992), pp. 2863–2866.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[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. Effcient 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.