Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Variational Calculus with Sums of Elementary Tensors of Fixed Rank

by

Mike Espig, Wolfgang Hackbusch, Thorsten Rohwedder, and Reinhold Schneider

Preprint no.: 52 2009

Variational Calculus with Sums of Elementary Tensors of Fixed Rank

Mike Espig [*] Wolfgang Hackbusch [*]
Thorsten Rohwedder [†] Reinhold Schneider [†]

August 28, 2009

Abstract

In this article we introduce a calculus of variations for sums of elementary tensors and apply it to functionals of practical interest. The survey provides all necessary ingredients for applying minimization methods in a general setting. The important cases of target functionals which are linear and quadratic with respect to the tensor product are discussed, and combinations of these functionals are presented in detail. As an example, we consider the solution of a linear system in structured tensor format. Moreover, we discuss the solution of an eigenvalue problem with sums of elementary tensors. This example can be viewed as a prototype of a constrained minimization problem. For the numerical treatment, we suggest a method which has the same order of complexity as the popular alternating least square algorithm and demonstrate the rate of convergence in numerical tests.

Keywords: Low tensor rank, sums of elementary tensors, variational calculus in tensor format.

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

1 Introduction

Approximation of solutions of high dimensional partial differential or integral equations by low rank tensors has yielded promising results, see e.g. [1, 2, 11, 12, 14, 15]. A tensor $u \in \mathbb{R}^{n^d}$ of order d requires in general a storage complexity of n^d. If u can be approximated by a low rank tensor

$$u \approx \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} u_{j,\mu},$$

the memory requirement reduces to drn, and the complexity of algebraic operations grows only linearily with respect to the order d. However, when using iterative methods for computing a low-rank tensor, one usually has to face the problem that the involved algebraic operations increase the tensor rank in each iteration step. To overcome this issue, efficient recompression methods have been developed in [1, 2, 5, 8, 9] to approximate a given sum of elementary tensors by low rank tensors. Moreover, the convergence of such approximate iterations is known, see [16] for an analysis. The subject of this article is in contrast to this approach. We will show that the representation of a tensor in low rank format allows in many cases of practical interest a direct optimization procedure on the manifold of tensors of a fixed rank r. Thus, we will solve the original problem directly in the low tensor rank format instead of solving a high dimensional problem indirectly by the use of approximative iterative schemes. This approach has some advantages: We can be sure that the solution is at least locally optimal with respect to the problem dependent functional, and we circumvent numerical problems that may arise during the compression step of approximate iterations even if the original task lacks this kind of approximation problems. An example where such problems may occur is the second numerical experiment discussed in this article, an eigenvalue problem for which we know a priori that there exists a low rank solution. If we would apply the indirect iterative methods discussed above, we have to approximate all iterands during the iterative process, while it is unclear that an iterand can be well approximated by low rank tensors.

The article is structured as follows: In the next section, we introduce notations and problems to be treated by our calculus. In the following part, we compute the derivatives of a general functional formulated on the set of rank r-tensors. Moreover, we will also discuss special parts of concrete target functionals. We specialize the treatment to functionals of particular practical interest in Section 4.

The next part deals with the numerical treatment of the optimization process by a suitable cg algorithm. Finally, we present numerical results for the two examples from Section 4.

2 Setting, notations and problem formulation

Let $d \in \mathbb{N}$. For $\nu \in \{1 \ldots, d\}$, let $\Omega_\nu \subseteq \mathbb{R}^n$, $H_\nu := L_2(\Omega_\nu)$ and $V_\nu \subseteq H_\nu$ reflexive Banach spaces, where each V_ν is dense in and continuously embedded into H_ν; for example, one might consider $V_\nu = H_0^s(\Omega_\nu)$. In this paper, we will be concerned with the tensor product space $\bigotimes_{\nu=1}^d V_\nu$, cf. [13, 19]. To keep notations simple, we will restrict our treatment to the case that $H_1 = \ldots = H_n =: H$ is a real Hilbert space, and will also suppose $V_1 = \ldots = V_n =: V$, although the above general case may be treated analogeously with the neccessary modifications. We denote by

$$\mathcal{V} := \mathcal{V}_d := \bigotimes_{\nu=1}^d V$$

the d-fold tensor product space over V. In the following, we will fix the parameter d and drop the suffix d in most cases to keep notations simple.
The norms on $H, \mathcal{H} = \bigotimes_{\nu=1}^d H$ and $\mathcal{V}$ will be denoted by $||.||_H, ||.||_{\mathcal{H}}$ and $||.||_{\mathcal{V}}$, respectively. For a convenient formulation of the problems we have in mind, we will use duality pairings of the following form,

$$\langle g, u \rangle := g(u), \;\; u \in V, \; g \in V'; \quad \langle G, U \rangle := G(U), \; U \in \mathcal{V}, \; G \in \mathcal{V}'.$$

Note that if we have $g \in H' \subseteq V'$ in the above situation, $\langle g, u \rangle$ may be identified with the inner product $\langle g, u \rangle_H$ on H; an analogeous statement holds for $G \in \mathcal{H}' \subseteq \mathcal{V}'$.

An elementary tensor $W \in \mathcal{V}$ is a tensor of the form $W = \bigotimes_{\nu=1}^d w_\nu \in \mathcal{V}$, $w_\nu \in V$. A tensor $U \in \mathcal{V}$ is called a *tensor of rank* r if it can be written as a sum of r elementary tensors,

$$U = \sum_{i=1}^r \bigotimes_{\nu=1}^d u_{i,\nu} \; .$$

The set of all tensors $U \in \mathcal{V}$ of rank r will be denoted by $\mathcal{K}^r$. Note that $\mathcal{K}^r$ is a cone, i.e. $U \in \mathcal{K}^r$ implies $\alpha U \in \mathcal{K}^r$, for all $\alpha \in \mathbb{R}$, but $\mathcal{K}^r$ is not a vector space,

and not even convex since for $W_1, W_2 \in \mathcal{K}^r$, there only holds $W_1 + W_2 \in \mathcal{K}^{2r}$, but in general we have $W_1 + W_2 \notin \mathcal{K}^r$.

Notation 2.1. *Let X a vector space, $Y \subset X$ and $f : Y \to \mathbb{R}$. We will use in the short notation $\mathfrak{M}(f, Y)$ for the set of minimizers of the induced minimization problem, i.e.*

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

Problem 2.2. *Given a functional $F : \mathcal{V} \to \mathbb{R}$ and an admissible set $\mathcal{M} \subset \mathcal{V}$, we are searching for a minimizer of the modified optimization problem where the original admissible set $\mathcal{M}$ is confined to tensors of rank at most r, i.e. we are searching for*

$$U \in \mathfrak{M}(F, \mathcal{M} \cap \mathcal{K}^r). \tag{2}$$

Let us mention a few basic examples which are important in several practical applications in high dimensions.

(i) The low rank approximation $F(W) = \|U - W\|_H^2$, $W \in \mathcal{K}^r$ for given $U \in \mathcal{V}$.

(ii) The solution of equations $AU = B$ or $g(U) = 0$ where $A, g : \mathcal{V} \to \mathcal{V}'$. Here we have $F(W) = \|AW - B\|_{\mathcal{V}'}$ resp. $\|g(W)\|_{\mathcal{V}'}$.

(iii) If $A : \mathcal{V} \to \mathcal{V}'$ is bounded, symmetric and coercive with respect to $||.||_{\mathcal{V}}$ and $B \in \mathcal{V}'$ given, we may instead of the first functional in (ii) focus on $F(W) := \frac{1}{2}\langle AW, W\rangle - \langle B, W\rangle$.
Note that in the case that $B = AU$ for fixed $U \in \mathcal{V}$, this task means finding a low rank approximation of U with respect to the energy norm induced by A (equivalent to the V-norm) instead of the Hilbert space norm H used in example (i).

(iv) Computation of the lowest eigenvalue of a symmetric operator $A : \mathcal{V} \to \mathcal{V}'$ by minimizing the Rayleigh quotient: $F(W) := \langle AW, W\rangle / \langle W, W\rangle$ over $\mathcal{M} = \mathcal{V}\backslash\{0\}$, This problem is equivalent to the constraint minimization problem

$$U = \mathfrak{M}\big(F, \ \{W \in \mathcal{V} : \ ||W||_{\mathcal{H}} = 1\}\big).$$

In the first three examples we have $\mathcal{M} \cap \mathcal{K}^r = \mathcal{K}^r$, while in the last example we have an additional constraint, namely $\mathcal{M} = \{W \in \mathcal{V} : \langle W, W\rangle = 1\}$. Note that in this case, $\mathcal{M} \cap \mathcal{K}^r \neq \emptyset$ due to the cone property of $\mathcal{K}^r$.

Solving the optimization problem (2) means finding a system of representants

$$\big(u_{i,\nu}\big)_{i=1}^{r}{}_{\nu=1}^{d} := \big\{u_{i,\nu} \in V : i \in \mathbb{N}_{\le r},\, \nu \in \mathbb{N}_{\le d}\big\},$$

where $\mathbb{N}_{\le k} := \{n \in \mathbb{N} : n \le k\}$, such that the minimizer U is representable by $U = \sum_{i=1}^{r} \bigotimes_{\nu=1}^{d} u_{i,\nu}$. Let us cast the unknown functions (resp. vectors) $u_{i,\nu}$, $i = 1, \dots, r$, $\nu = 1, \dots, d$, into a vector (resp. matrix),

$$\mathbf{u} := \big(u_{i,\nu}\big)_{i=1}^{r}{}_{\nu=1}^{d} \in V^{r\times d}.$$

If V is finite dimensional, which after discretization is always the case in practice, the required number of degrees of freedom is $\dim V^{d\times r} = dr\dim V$, i.e. it grows only linearly with respect to the dimension d. This fact makes the representation of tensors by sums of elementary tensors an attractive option in particular in high dimensions.

In order to find $\mathbf{u} \in V^{r\times d}$ for representation of the minimizer $U \in \mathcal{V}$, we introduce the (multilinear) mapping

$$U : V^{r\times d} \to \mathcal{V} = \bigotimes_{\nu=1}^{d} V \ , \mathbf{u} \mapsto U(\mathbf{u}) := \sum_{i=1}^{r} \bigotimes_{\nu=1}^{d} u_{i,\nu} \ .$$

Then our original optimization problem (2) takes the form

$$\text{Find } \mathbf{u} \in \mathfrak{M}(J, V^{r\times d}),$$

where we set $J := F \circ U : V^{r\times d} \to \mathbb{R}$.

Since the representation $\mathbf{u} \in V^{r\times d}$ of $U(\mathbf{u}) \in \mathcal{K}^r$ is neither unique nor stable, the above optimization problem inherits additional difficulties and redundance, which should be removed in advance. In particular, the border rank problem [4] can be abolished by bounding the norm of the single elementary tensors. If V is densely embedded in a Hilbert space H (e.g. $H = L_2(\Omega)$ or $H = \ell_2(\mathbb{N})$), it is often numerically advantageous to impose the following constraint conditions on the H-norm of the vectors $u_{i,\nu}$:

$$\langle u_{i,\nu}, u_{i,\nu}\rangle = \|u_{i,\nu}\|_H^2 = 1 \quad \text{for} \quad 1 \le \nu \le d-1,\ \ 1 \le i \le r, \tag{3}$$

$$\|u_{i,d}\|_H^2 \le C \quad \text{for} \quad 1 \le i \le r. \tag{4}$$

Note that this implies that the norm of the corresponding elementary tensors U_i constituting $\mathbf{u}$ is bounded, $||U_i||_{\mathcal{H}} = ||\otimes_{\nu=1}^{d} u_{i,\nu}||_{\mathcal{H}} \le C_i$ for all $i \in \mathbb{N}_{\le r}$. The set

of all $\mathbf{u} \in V^{r\times d}$, where the vectors $u_{i,\nu}$ satisfy conditions (3) and (4) is denoted by $M \subset V^{r\times d}$.

Alternatively, the redundancy in an elementary tensor may be reduced by bounding and equilibrating the norms of the elementary tensors U_i, see [5]:

$$\|u_{i,\nu}\|_H \le C \quad \text{for all} \quad 1 \le \nu \le d,\ 1 \le i \le r, \tag{5}$$

$$\|u_{i,\nu}\|_H = \|u_{i,\mu}\|_H \quad \text{for} \quad 1 \le \nu, \mu \le d,\ 1 \le i \le r. \tag{6}$$

We will impose these constraint conditions either by introducing penalty terms in the functional or by treating them explicitly. Therefore, we have arrived at the following optimization problem, which is from now on the basic problem under consideration.

Problem 2.3.

$$\textit{Find } \mathbf{u} \in \mathfrak{M}(J, M). \tag{7}$$

Remark 2.4. *M is a closed and bounded subset of $V^{r\times d}$ and $U : V^{r\times d} \to \mathcal{V}$ is a continuous mapping.*

If V is finite dimensional, if $F : \mathcal{V} \to \mathbb{R}$ is continuous and if $\mathcal{M} = \mathcal{V}$ or at least $\mathcal{M} \cap U(M) \neq \emptyset$ is closed, then $J : M \to \mathbb{R}$ is continuous. Under these premises, there exists a solution of the above problem (7). If V is infinite dimensional, the situation is more challenging due to the lack of compactness; note though that for $V = H$, i.e. in the Hilbert space case, the existence of a best rank r-approximation (cf. Problem (i) above) has recently been proven [7, 18].

3 Computation of the Derivatives

We would like to find a local minimizer by local first and second order methods, i.e. by means of differential calculus. We start by computing the derivatives for

$$U(\mathbf{u}) = U\Big(\big(u_{i,\nu}\big)_{i=1\,\nu=1}^{r\quad d}\Big) = \sum_{i=1}^{r} U_i(\mathbf{u}) := \sum_{i=1}^{r} \bigotimes_{\nu=1}^{d} u_{i,\nu}\,. \tag{8}$$

The Fréchet derivative $U'(\mathbf{u})$ of U at $\mathbf{u} \in V^{r\times d}$ is a linear mapping from $V^{d\times r}$ to $\mathcal{V}$. Due to the multilinearity of U, it may be expressed by the partial derivatives of U in direction $u_{k,\alpha} \in V$ which we will denote by $(U'(\mathbf{u}))_{(k,\alpha)} := dU(\mathbf{u})/du_{k,\alpha} \in L(V,\mathcal{V})$. These map $v \in V$ to

$$\begin{aligned}(U'(\mathbf{u}))_{(k,\alpha)}(v) &= \lim_{h\to 0}\frac{1}{h}\left[\bigotimes_{\nu=1}^{\alpha-1} u_{k,\nu} \otimes (u_{k,\alpha}+hv) \otimes \bigotimes_{\nu=\alpha+1}^{d} u_{k,\nu} - \bigotimes_{\nu=1}^{d} u_{k,\alpha}\right]\\ &= \bigotimes_{\nu=1}^{\alpha-1} u_{k,\nu} \otimes v \otimes \bigotimes_{\nu=\alpha+1}^{d} u_{k,\nu}.\end{aligned}$$

We note that $(U'(\mathbf{u}))_{(k,\alpha)}(v) \in \mathcal{V}$ may alternatively be obtained by evaluating the Fréchet derivative $U'(\mathbf{u})$ at $(0,\dots,0,v,0,\dots,0) = v\otimes \mathbf{e}_{k,\alpha} =: v\mathbf{e}_{k,\alpha} \in V^{r\times d}$, where $\mathbf{e}_{k,\alpha}$ denotes the unit row vector $(\delta_{k,\alpha})_{i,\mu} \in \mathbb{R}^{1,r\times d}$. If $\mathbf{u}$ is fixed, we will denote the partial derivatives at $\mathbf{u}$ by

$$U_k^\alpha(v) \;:=\; U'(\mathbf{u})(v\mathbf{e}_{k,\alpha}) = (U'(\mathbf{u}))_{(k,\alpha)}(v).$$

in the following to keep notations simpler.

Corollary 3.1. *The directional first order derivative of the functional $J := F\circ U : V^{r\times d} \to \mathbb{R}$ from (7) at point $\mathbf{u} \in V^{d\times r}$ in direction $v\mathbf{e}_{k,\alpha}$ is given by*

$$J'_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}) = F'_{U(\mathbf{u})}(U_k^\alpha(v)) = \langle F'_{U(\mathbf{u})}, U_k^\alpha(v)\rangle \tag{9}$$

For second order schemes, and possibly for preconditioning, we also need second order derivatives.

For $v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta} \in V^{d\times r}$, we obtain in the case that $k=\ell$, $\alpha<\beta$, that

$$U^{(2)}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta}) = \bigotimes_{\nu=1}^{\alpha-1} u_{k,\nu} \otimes v \otimes \bigotimes_{\nu=\alpha+1}^{\beta-1} u_{k,\nu} \otimes w \otimes \bigotimes_{\nu=\beta+1}^{d} u_{k,\nu} \tag{10}$$

The case $k=\ell$, $\alpha>\beta$ follows from (10) by symmetry, while $U^{(2)}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta}) = 0$ if $k\neq\ell$ or $\alpha=\beta$.

In analogy to the first order derivatives, we define for fixed $\mathbf{u}$

$$U^{\alpha,\beta}_{k,\ell}(v,w) \;:=\; U^{(2)}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta}) \;=\; (U^{(2)}_{\mathbf{u}})_{(k,\alpha),(\ell,\beta)}(v,w). \tag{11}$$

Corollary 3.2. *For the second derivative of the functional J there holds*

$$J^{(2)}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta}) \;=\; F^{(2)}_{U(\mathbf{u})}(U^{\alpha}_{k}(v), U^{\beta}_{\ell}(w)) + \langle F'_{U(\mathbf{u})}, U^{\alpha,\beta}_{k,\ell}(v,w)\rangle\ . \quad (12)$$

Proof. Obviously, we have

$$\begin{aligned} J^{(2)}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta}) &= F^{(2)}_{U(\mathbf{u})}(U'_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}), U'_{\mathbf{u}}(w\mathbf{e}_{\ell,\beta})) + F'_{U(\mathbf{u})}(U^{2}_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}, w\mathbf{e}_{\ell,\beta})) \\ &= F^{(2)}_{U(\mathbf{u})}(U^{\alpha}_{k}(v), U^{\beta}_{\ell}(w)) + F'_{U(\mathbf{u})}(U^{\alpha,\beta}_{k,\ell}(v,w)) \\ &= F^{(2)}_{U(\mathbf{u})}(U^{\alpha}_{k}(v), U^{\beta}_{\ell}(w)) + \langle F'_{U(\mathbf{u})}, U^{\alpha,\beta}_{k,\ell}(v,w)\rangle\ . \end{aligned}$$

■

For practical applications, let us take a closer look at the treatment of functionals which are linear or at most quadratic. First, for $U \in \mathcal{V}, W \in \mathcal{V}'$, let us consider a linear functional of the form $U \mapsto \langle W, U\rangle$, which induces a functional $b : \mathbf{u} \mapsto b(\mathbf{u}) := \langle W, U(\mathbf{u})\rangle$ on $V^{r\times d}$.

Corollary 3.3. *The directional derivative of b at point $\mathbf{u}$ with respect to $v\mathbf{e}_{k,\alpha}$ is given by*

$$b'_{\mathbf{u}}(v\mathbf{e}_{k,\alpha}) = \langle W, U_{k,\alpha}(v)\rangle.$$

If W is an elementary tensor, i.e. $W = \bigotimes_{\nu=1}^{d} w_\nu$, there holds for the above directional derivatives that

$$\langle W, U_{k,\alpha}(v)\rangle = \langle W, U_k\rangle_\alpha := \Big(\prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}} \langle w_\nu, u_{k,\nu}\rangle\Big) w_\alpha \in V'.$$

If W is a finite rank tensor, i.e. $W = \sum_{j=1}^{R}\bigotimes_{\nu=1}^{d} w_{j,\nu} =: \sum_{j=1}^{R} W_j$, there holds

$$(b'_{\mathbf{u}})_{(k,\alpha)} = \sum_{j=1}^{R} \langle W_j, U_k(\mathbf{u})\rangle_\alpha$$

with $U_k(\mathbf{u})$ from (8).

In a more explicit form we have

$$(b'_{\mathbf{u}})_{k,\alpha} = \sum_{j=1}^{R} \Big(\prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}} \langle w_{j,\nu}, u_{k,\nu}\rangle\Big) w_{j,\alpha} = \sum_{j=1}^{R} b_{j,\alpha,k} w_{j,\alpha}, \quad (13)$$

where we let $b_{j,\alpha,k} := \prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}} \langle w_{j,\nu}, u_{k,\nu}\rangle$ for brevity.

Lemma 3.4. *If $V = \mathbb{R}^n$, the complexity for the computation of the directional derivative of $b : V^{r\times d} \to \mathbb{R}$ is $\mathcal{O}(dRrn)$.*

Proof. Let $\alpha \in \mathbb{N}_{\le d}$, $j \in \mathbb{N}_{\le R}$ and $k \in \mathbb{N}_{\le r}$. Similar to [6, Remark 21], the values $b_{j,\alpha,k}$ can be computed in $\mathcal{O}(dRrn)$. In addition we have to compute

$$(b'_{\mathbf{u}})_{k,\alpha} = \sum_{j=1}^{R} \prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}} \langle w_{j,\nu}, u_{k,\nu}\rangle w_{j,\alpha} = \sum_{j=1}^{R} b_{j,\alpha,k} w_{j,\alpha}.$$

This needs $2n(R-\frac{1}{2})$ operations for every α and k. Hence, the overall complexity is $\mathcal{O}(dRrn)$. ∎

Remark 3.5. *Let $W \in C(\Omega^d) \subseteq \otimes_{i=1}^d H = L^2(\Omega^d)$ and $z \in \Omega^d$. Then we define in analogy to (3)*

$$\langle W, U_i\rangle_\nu(z) := \langle W, U_{i,\nu}(\delta_z)\rangle,$$

where $\langle f, \delta_z\rangle := \delta_z(f) := f(z)$ denotes the Dirac distribution at the point $z \in \Omega^d$. Note that if W is not available as a low rank tensor, the computation of $\langle W, U_i\rangle_\nu(z)$ for this general case requires high-dimensional integration over $\Omega_\nu \subseteq \mathbb{R}^{d-1}$.

Apart from linear functionals, we now compute the derivatives of the functional $\mathbf{u} \mapsto G(U(\mathbf{u})) = \frac{1}{2}\langle AU(\mathbf{u}), U(\mathbf{u})\rangle$ with a symmetric operator $A : \mathcal{V} \to \mathcal{V}$, which is quadratic with respect to $U(\mathbf{u}) \in \mathcal{V}$.

Corollary 3.6. *The derivative $G'_{\mathbf{u}}$ can be written as*

$$(G'_{\mathbf{u}})_{(k,\alpha)} \quad = \quad \sum_{j=1}^{r} \langle AU_j(\mathbf{u}), U_k(\mathbf{u})\rangle_\alpha. \tag{14}$$

Remark 3.7. *If a linear operator $A : \mathcal{V} \to \mathcal{V}'$ can be decomposed into a finite sum of elementary tensors,*

$$A = \sum_{j=1}^{s} A_j = \sum_{j=1}^{s} \bigotimes_{\nu=1}^{d} A_{j,\nu}\ ,\ A_{j,\nu} : V \to V',$$

then $\langle AU_j(\mathbf{u}), U_k(\mathbf{u})\rangle_\alpha \in \mathcal{V}'$ is computable within polynomial cost, provided that the individual terms $\langle A_{j,\nu}u_{i,\nu}, u_{k,\nu}\rangle$ are computable. In this case the derivative

$G'_{\mathbf{u}}$ is expressed by

$$(G'_{\mathbf{u}})_{(k,\alpha)} = \sum_{i=1}^{r}\sum_{j=1}^{s}\langle A_j U_i(\mathbf{u}), U_k(\mathbf{u})\rangle_\alpha \tag{15}$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}\Big(\prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}}\langle A_{j,\nu}u_{i,\nu}, u_{k,\nu}\rangle\Big)A_{j,\alpha}u_{i,\alpha}. \tag{16}$$

In short-hand notation we can write for the derivatives of the quadratic functional G and the linear functional b, respectively,

$$G'_{\mathbf{u}} = \mathbf{A}_{\mathbf{u}}\mathbf{u},\ \ b'_{\mathbf{u}} = \mathbf{b}_{\mathbf{u}}.$$

It is worth mentioning that the matrix $\mathbf{A}_{\mathbf{u}}$ and the vector $\mathbf{b}_{\mathbf{u}}$ have a nice tensor structure, namely

$$\mathbf{A}_{\mathbf{u}} = \sum_{\alpha=1}^{d}\sum_{j=1}^{s}\mathbb{E}_\alpha\otimes G_{j,\alpha}(\mathbf{u})\otimes A_{j,\alpha},\quad \mathbf{b}_{\mathbf{u}} = \sum_{\alpha=1}^{d}\sum_{j=1}^{R} e_\alpha^{(d)}\otimes b_{j,\alpha}(\mathbf{u})\otimes w_{j,\alpha} \tag{17}$$

where $\mathbb{E}_\alpha\in\mathbb{R}^{d\times d}$, $G_{j,\alpha}\in\mathbb{R}^{r\times r}$, $e_\alpha^{(d)}\in\mathbb{R}^d$ and $b_{j,\alpha}\in\mathbb{R}^r$ with

$$(G_{j,\alpha}(\mathbf{u}))_{k,i=1}^{r} := \prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}}\langle A_{j,\nu}u_{i,\nu}, u_{k,\nu}\rangle,\quad (b_{j,\alpha,}(\mathbf{u}))_{k=1}^{r} := \prod_{\nu\in\mathbb{N}_{\le d}\setminus\{\alpha\}}\langle w_{j,\nu}, u_{k,\nu}\rangle$$

and $(\mathbb{E}_\alpha)_{\nu,\nu'=1}^{d} := \delta_{\alpha,\nu}\delta_{\alpha,\nu'}$, $(e_\alpha^{(d)})_{\nu=1}^{d} := \delta_{\alpha,\nu}$.

Lemma 3.8. *If $V=\mathbb{R}^n$, the complexity for the computation of the directional derivative of $G: V^{r\times d}\to\mathbb{R}$ is $\mathcal{O}(dsr^2n^2)$. If the components $A_{j,\nu}$ of A are sparse in the sense that matrix-vector products can be evaluated in $\mathcal{O}(n)$, the complexity reduces to $\mathcal{O}(dsr^2n)$.*

Proof. The proof is similar to Lemma 3.4. ■

4 Model Examples

We will present two basic examples. At first, we consider the minimization of the functional $F(U) := \frac{1}{2}\langle AU, U\rangle - \langle B, U\rangle$ prepared in the last section, where

$A : \mathcal{V} \to \mathcal{V}'$ is a symmetric linear operator, B and a linear functional $B \in \mathcal{V}'$, and both can be represented in the respective tensor formats

$$A = \sum_{j=1}^{R} \bigotimes_{\mu=1}^{d} A_{j,\mu} \ , \ A_{j,\mu} : V \to V' \ , \quad B = \sum_{j=1}^{s} \bigotimes_{\mu=1}^{d} b_{j,\mu} \ , \ b_{j,\mu} \in V' \ .$$

This kind of quadratic minimization problem can be easily extended to those quadratic minimization problems which are constrained by a set of linear side conditions. Formulating the corresponding minimization problem on the cone of rank r tensors $\mathcal{K}^r$ as in Section 2 yields the functional

$$J(\mathbf{u}) := \frac{1}{2}\langle AU(\mathbf{u}), U(\mathbf{u})\rangle - \langle B, U(\mathbf{u})\rangle = \frac{1}{2}G(\mathbf{u}) - b(\mathbf{u}) \ .$$

Usually we invoke the soft constraint conditions as described in Section 1, see (3) and (4). For sake of simplicity, let us neglect these weak constraints for a first view at the equations. The first order optimality condition is $J'_{\mathbf{u}}(w\mathbf{e}_{k,\alpha}) = 0$ for all $w \in V$, $k \in \mathbb{N}_{\leq r}, \alpha \in \mathbb{N}_{\leq d}$. With the results of the previous section, these conditions can be rewritten as the following nonlinear equations:

$$J'_{\mathbf{u}} = \mathbf{A}_{\mathbf{u}}\mathbf{u} - \mathbf{b}_{\mathbf{u}} = \mathbf{0} \in (V')^{r\times d}. \tag{18}$$

Herein, for fixed $\mathbf{u} \in V^{r\times d}$, we have $\mathbf{A}_{\mathbf{u}} : V^{r\times d} \to (V^{r\times d})'$ and $\mathbf{b}_{\mathbf{u}} \in (V^{r\times d})'$. With this notation at hand, we may propose an iteration of the form

$$\mathbf{A}_{\mathbf{u}^{(n)}}\mathbf{u}^{(n+1)} - \mathbf{b}_{\mathbf{u}^{(n)}} = \mathbf{0}.$$

Alternatively, we may use an iteration of steepest descent type for minimization the functional $J(\mathbf{u}) = \frac{1}{2}G(\mathbf{u}) - b(\mathbf{u})$ by using the gradient from (18). Note that the components of $\mathbf{A}_{\mathbf{u}^{(n)}} : V^{r\times d} \to (V^{r\times d})'$ map the space V into its dual space V'. If $V \neq H = L_2$, a gradient type algorithm often requires further preconditioning using a simply invertible operator $B : V \to V'$ with $\langle Bu, u\rangle \sim \|u\|_V^2$. We may for example use a preconditioned gradient type algorithm that uses $(\mathbf{A}_{\mathbf{u}^{(n)}})^{-1}$ as a preconditioner, i.e. as an approximate inverse of the Hessian. This explains how operator equations defined on Sobolev spaces, which are not of the simple product form, may be preconditioned in the present setting.

Next, let us turn to the side conditions (3) and (4) from Section 2. Condition (3) may be enforced by usage of the functional

$$h_1(\mathbf{u}) := \frac{1}{2}\sum_{j=1}^{r}\sum_{\mu=1}^{d-1}(\langle u_{j,\mu}, u_{j,\mu}\rangle - 1),$$

for which the derivative is easy to compute:

$$h_1{}'_{\mathbf{u}}(\mathbf{e}_{k,\alpha}w) = \langle u_{k,\alpha}, w\rangle.$$

The second side condition is treated like in the recent works [2, 5, 8, 9]. In this context, the function

$$h_2(\mathbf{u}) := \frac{1}{2}\sum_{j=1}^{r}\prod_{\mu=1}^{d}\|u_{j,\mu}\|^2$$

is important. The derivative of the penalty term is

$$h_2{}'_{\mathbf{u}}(\mathbf{e}_{k,\alpha}w) = \prod_{\mu=1,\mu\neq\alpha}^{d}\|u_{k,\mu}\|^2\langle u_{k,\alpha}, w\rangle.$$

As a second basic example, let us consider a quadratic minimization problem with quadratic constraints, namely Problem (iv) from Section 2, i.e. the computation of the eigenvector belonging to the lowest eigenvalue of a symmetric operator $A : \mathcal{V} \to \mathcal{V}'$. The corresponding functional is given by

$$F(U) := \frac{1}{2}\langle AU, U\rangle \quad \text{with the constraint} \quad \langle U, U\rangle = 1. \tag{19}$$

The corresponding optimization problem in tensor format thus reads

$$\text{Find} \quad \min\{J(\mathbf{u}) := \frac{1}{2}\langle AU(\mathbf{u}), U(\mathbf{u})\rangle : \langle U(\mathbf{u}), U(\mathbf{u})\rangle = 1\}.$$

The present calculus yields for the corresponding Lagrange functional

$$\begin{aligned} L(\mathbf{u},\lambda) &= \frac{1}{2}\langle AU(\mathbf{u}), U(\mathbf{u})\rangle - \lambda h(\mathbf{u}), \\ h(\mathbf{u}) &:= \frac{1}{2}\big(\langle U(\mathbf{u}), U(\mathbf{u})\rangle - 1\big), \end{aligned} \tag{20}$$

the first order optimality conditions

$$(\mathbf{A}_{\mathbf{u}} - \lambda\mathbf{M}_{\mathbf{u}})\mathbf{u} = \mathbf{0} \in (V')^{r\times d}, \tag{21}$$

that is, a nonlinear generalized eigenvalue problem, where for given $\mathbf{u} \in V^{r\times d}$, $\mathbf{A}_{\mathbf{u}}$ and $\mathbf{M}_{\mathbf{u}}$ map $V^{r\times d}$ into $(V^{r\times d})'$. Here, $\mathbf{A}_{\mathbf{u}}$ is the same as in (17), and

$$\mathbf{M}_{\mathbf{u}} = \sum_{\mu=1}^{d}\mathbb{E}_\mu \otimes H_\mu(\mathbf{u}) \otimes \mathbf{Id}_V \tag{22}$$

with

$$(\mathbb{E}_\mu)_{\nu,\nu'=1}^d := (\delta_{\mu,\nu}\delta_{\mu,\nu'})_{\nu,\nu'=1}^d \in \mathbb{R}^{d\times d},\ (H_\mu(\mathbf{u}))_{k,i=1}^r := \prod_{\nu\in\mathbb{N}_{\leq d}\setminus\{\mu\}} \langle u_{i,\nu}, u_{k,\nu}\rangle \in \mathbb{R}^{r\times r}.$$

Problem (21) is similar to the Hartree Fock and Kohn Sham equations or orbital minimization in multi-configuration methods used in quantum chemistry. Note that (18) also has a similar structure.

5 Conjugate Gradient Method for Minimization

So far we have developed all ingredients for applying steepest decent type algorithms. The most popular choice of minimization methods with sums of elementary tensors is a relaxation type method: For given ν, all $u_{i,\mu}$ with $\mu \neq \nu$ are kept fixed in this approach, and only the vectors $u_{i,\nu}$, $i = 1,\dots,r$ are optimized. This minimization step is then repeatedly alternated over all directions $\nu \in \mathbb{N}_{\leq d}$, resulting in the well known alternating least square (ALS) algorithm, see e.g. [1, 2]. Although it is known that the convergence behaviour of the ALS method is not optimal, the ALS method has some important advantages. It is fairly convenient to implement and the complexity of a single iteration step is small. In [5, 8, 6] a modified Newton method is used to solve minimization problems with sums of elementary tensors. Compared to the ALS algorithm the modified Newton method has a better rate of convergence but a single iteration step is more expensive. Moreover, we have to use special properties of the functional to make the modified Newton method efficient and spend more effort while implementing the algorithm. In order to overcome these problems, a conjugate gradient (CG) method is introduced in [9] which converges globally to a stationary point with a complexity similar to the ALS method. We will also use the CG method and describe it briefly in the following. For a detailed analysis we refer to [9].

The crucial part of the CG algorithm is the computation of the exact line search parameter $\alpha_k \in \mathbb{R}_{\geq 0}$. Given a direction $\mathbf{d}^k$, we have to find a solution of the one-dimensional nonlinear equation

$$p(\alpha_k) = \left\langle J'(\mathbf{u}^k + \alpha\mathbf{d}^k), \mathbf{d}^k \right\rangle = 0.$$

Normally we avoid the exact line search and use an Armijo type inexact line search. In our applications though, equations (13) and (16) show that the function p is a polynomial of degree at most $2d-1$. Hence we will apply a third order

Algorithm 1 Conjugate Gradient (CG) Method

Choose initial $\mathbf{u}^0 \in V^{r\times d}$ and parameter $\varepsilon \in \mathbb{R}_{>0}$. Define $k := 0$, $\mathbf{g}^0 := J'(u^0)$ and $\mathbf{d}^0 := -\mathbf{g}^0$.
while $\|\mathbf{g}^k\| > \varepsilon$ **do**
  Compute $\alpha_k := \min\left\{\alpha \in \mathbb{R}_{\geq 0} : p(\alpha) := \left\langle J'(\mathbf{u}^k + \alpha\mathbf{d}^k), \mathbf{d}^k\right\rangle = 0\right\}$.
  $\mathbf{u}^{k+1} := \mathbf{u}^k + \alpha_k\mathbf{d}^k$.
  $\mathbf{g}^{k+1} := J'(\mathbf{u}^{k+1})$.
  $\beta_k := \frac{\left\langle \mathbf{g}^{k+1}-\mathbf{g}^k, \mathbf{g}^{k+1}\right\rangle}{\|\mathbf{g}^k\|^2}$, $\gamma_k := \max\{0, \beta_k\}$.
  $\mathbf{d}^{k+1} := -\mathbf{g}^{k+1} + \gamma_k\mathbf{d}^k$.
  $k \mapsto k+1$.
end while

derivative-free procedure (3-PG) for finding zeros of a function, as described in [17]. The 3-PG method is globally R-order convergent for $f \in C^2[a,b]$, where $a, b \in \mathbb{R}$ with $f(a)f(b) < 0$. The order of convergence is defined by the real root of the polynomial $t \mapsto t^3 - t^2 - t - 1$ (≈ 1.8393). Moreover, the 3-PG method is equivalent to the Newton method for polynomials of degree three. An algorithmic description of the 3-PG method is presented below.

A typical decay of $|p(\alpha)|$ with respect to the number of 3-PG iterations is shown in Figure 1. It is remarkable that only function evaluations of the function p are necessary for the favorable rate of convergence.

Remark 5.1. *According to Lemma 3.4 and Lemma 3.8, the complexity of the computation of the gradient J' is $\mathcal{O}(drn(srn+R))$ ($\mathcal{O}(drn(sr+R)$ in the sparse case). Since the most expensive part in the CG method is the calculation of the gradient, the complexity of the CG method is*

$$k_{\max}\mathcal{O}(drn(srn+R)) \tag{23}$$

($k_{\max}\mathcal{O}(drn(sr+R))$ in the sparse case), where $k_{\max}$ denotes the maximal number of iterations in Algorithm 1.

Algorithm 2 3-PG Algorithm

Choose initial $a, b \in \mathbb{R}$ with $p(a)p(b) < 0$ and parameter $\varepsilon \in \mathbb{R}_{>0}$, $C, D \in \left(\frac{1}{2}, 1\right)$. Define $R := a$, $p_a := p(a)$, $p_b = p(b)$ and compute

$$p[b,a] := \frac{p_b - p_a}{b-a}, \quad \alpha = b - \frac{p_b}{p[b,a]}, \quad p_\alpha = p(\alpha).$$

while $|p_\alpha| > \varepsilon$ **do**
  if $p_\alpha p_b < 0$ **then**
    $R := b$.
  end if
  Compute $p[\alpha, b] := \frac{p_\alpha - p_b}{\alpha - b}$, $p[\alpha, a] := \frac{p_\alpha - p_a}{\alpha - a}$ and

$$Q(a,b,\alpha) := \frac{(\alpha - a)p[\alpha,b] + (b-\alpha)p[\alpha,a]}{b-a}.$$

  if $Q(a,b,\alpha) = 0$ **then**
    $y := \frac{R+\alpha}{2}$.
  else
    $y := \alpha - \frac{p_\alpha}{Q(a,b,\alpha)}$.
  end if
  if $(y-\alpha)(y-R) > 0$ or $[|y-R| > C|\alpha - R|$ and $|p_\alpha| > D|p_b|]$ **then**
    $y := \frac{R+\alpha}{2}$.
  end if
  $a = b$, $p_a = p_b$, $b = \alpha$, $p_b = p_\alpha$, $\alpha = y$, $p_\alpha = p(\alpha)$.
end while

6 Numerical Experiments

6.1 Unconstrained Minimization Problem

The first numerical test is the Poisson equation in d dimensions with Dirichlet boundary condition. We consider

$$\begin{aligned} -\Delta u &= f \quad \text{in} \quad \Omega := [0,1]^d \\ u &= 0 \quad \text{on} \quad \partial\Omega, \end{aligned}$$

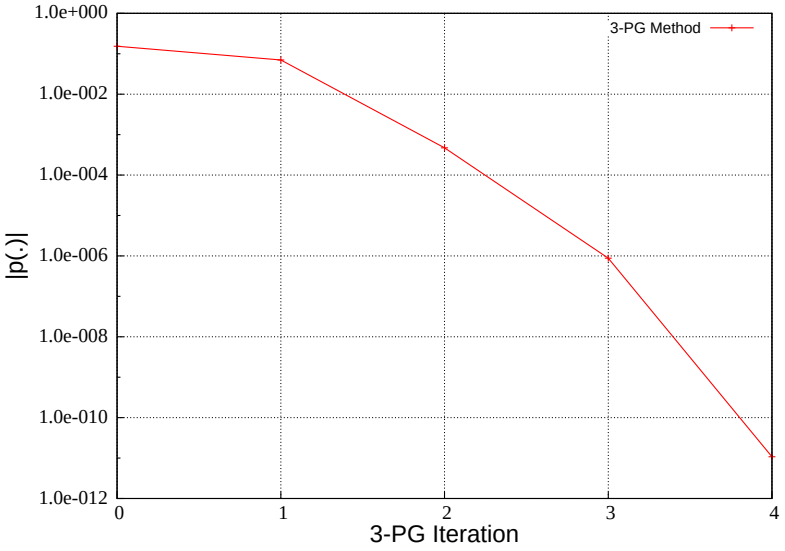

Figure 1: Decay of $|p(\alpha)|$.

with a separable right-hand side $f(x_1, \ldots, x_d) := \sum_{i=1}^{R} \prod_{\mu=1}^{d} f_{i,\mu}(x_\mu)$. A standard finite difference discretization on uniform grids leads to a linear system $AU(\mathbf{u}) = b$ with

$$A \;=\; \mathbf{T} \otimes \mathbf{Id} \otimes \ldots \otimes \mathbf{Id} + \cdots + \mathbf{Id} \otimes \ldots \otimes \mathbf{Id} \otimes \mathbf{T}, \quad b = \sum_{i=1}^{R} \bigotimes_{\mu=1}^{d} b_{i,\mu},$$

where the matrix $\mathbf{T}$ is a discretized version of the second derivative, e.g.

$$T = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

For simplicity, the vectors $b_{i,\mu}$ are initialized with uniformly distributed pseudo-random numbers, where we set $\|b\|_2 = 1$. All plots display the convergence of the relative residual $\|AU(\mathbf{u}) - b\|_2$ with respect to the separation rank of $U(\mathbf{u})$. The results of CG method applied to the function $J(\mathbf{u}) = \frac{1}{2}\langle AU(\mathbf{u}), U(\mathbf{u})\rangle - \langle B, U(\mathbf{u})\rangle = \frac{1}{2}G(\mathbf{u}) - b(\mathbf{u})$ are shown in Figures 2 and 3 for $n = 100$ and $n = 1000$ respectively. The computation is done for various dimensions $d \in$

$\{25, 50, 100\}$. We observe that in all numerical experiments the value of the relative approximation error is less than 10^{-6} for separation ranks about 15. In order to compare our results, we set $\text{rank}(b) = R := 1$ and also compute a separable approximation of the inverse of A with the use of exponential sums, see [3]. In [3], Braess and Hackbusch analyse the best approximation of the inverse function $\frac{1}{\cdot} : [1, c] \to \mathbb{R}$ by exponential sums with respect to the maximum norm. Moreover, an upper bound of the approximation error is given there:

Lemma 6.1. *Let $k \in \mathbb{N}$, $s_k(\tau) := \sum_{l=1}^{k} \omega_l \exp(-\alpha_l \tau)$ with $\alpha_l, \omega_l \in \mathbb{R}_{>0}$. With the optimal choice of the parameter α_l and ω_l we have*

$$\sup_{\tau \in [1,c]} \left| \frac{1}{\tau} - s_k(\tau) \right| \leq 16 \exp\left(\frac{-k\pi^2}{\log(8c)} \right).$$

The parameters α_l and ω_l are precomputed for different k and c. The values are available at the web page [10]. From this approximation, it follows that for the optimal choice of α_l and ω_l,

$$\|A^{-1} - s_k(A)\|_2 \leq \frac{16}{\lambda_{\min}(A)} \exp\left(\frac{-k\pi^2}{\log(8\kappa(A))} \right),$$

where $s_k(A) = \sum_{l=1}^{k} \omega_l \otimes_{\mu=1}^{d} \exp(-\alpha_l T)$. As mentioned above, we observe that the value of the relative approximation residual is less than 10^{-6} for ranks around 15. We set $k := 15$ and use coefficients α_l, ω_l from the web page to compute the relative residual of the approximation with exponential sums, i.e. we compute

$$\rho := \|Au_e - d\|_2, \quad u_e := s_{15}(A)b = \sum_{l=1}^{15} \omega_l \Big(\otimes_{\mu=1}^{d} \exp(-\alpha_l T) b_\mu \Big).$$

For $n = 100$ and $n = 1000$ the value of the relative residual ρ is 1.52×10^{-3} and 4.56×10^{-2} respectively.

6.2 Constrained Minimization Problem

Our second example is the eigenvalue problem (19), which may serve as an example for a constrained minimization problem. For our numerical illustration, we only use a penalty method to enforce the side condition $\langle U(\mathbf{u}), U(\mathbf{u}) \rangle = 1$.

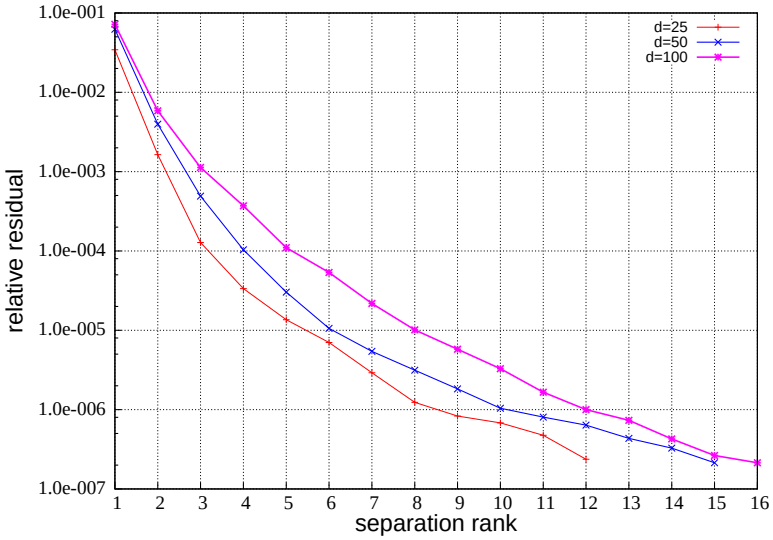

Figure 2: Relative residual error of the optimal low tensor rank approximation for the Poisson equation with Dirichlet boundary conditions and $n := 100$.

Herein, the constrained optimization problem (19) is replaced by a series of unconstrained problems the solutions of which converge to the solution of the original constrained problem. These unconstrained problems are formulated by adding a penalty term to the target function, and then solved by using the CG method as in the previous example, see Algorithm 1. This choice for the treatment of the side conditions may be taken as the CG algorithm is a first order method, while a second order approach like a modified Newton method would lead to ill-conditioned system matrices. In the following we will describe the penalty algorithm, using $L(\mathbf{u}, \lambda)$ and $h(\mathbf{u})$ as defined in equation (20).

One interesting application for an eigenvalue problem is the computation of the maximum norm. A straightforward approach gives a complexity linear in the number of entries in the tensor, i.e. the complexity is $\mathcal{O}(n^d)$. This fact makes the computation of the maximum norm especially in high dimensions nontrivial. In [5], it was shown that for a given sum of elementary tensors $u := \sum_{j=1}^{r} \otimes_{\mu=1}^{d} u_{j\mu} \in \bigotimes_{\mu=1}^{d} \mathbb{R}^n$, the computation of the maximum norm, i.e.

$$\|u\|_\infty := \max_{\underline{i}:=(i_1,\ldots,i_d)\in\mathbb{N}^d_{\leq n}} |u_{\underline{i}}| = \max_{\underline{i}:=(i_1,\ldots,i_d)\in\mathbb{N}^d_{\leq n}} \left| \sum_{j=1}^{r} \prod_{\mu=1}^{d} (u_{j\mu})_{i_\mu} \right|,$$

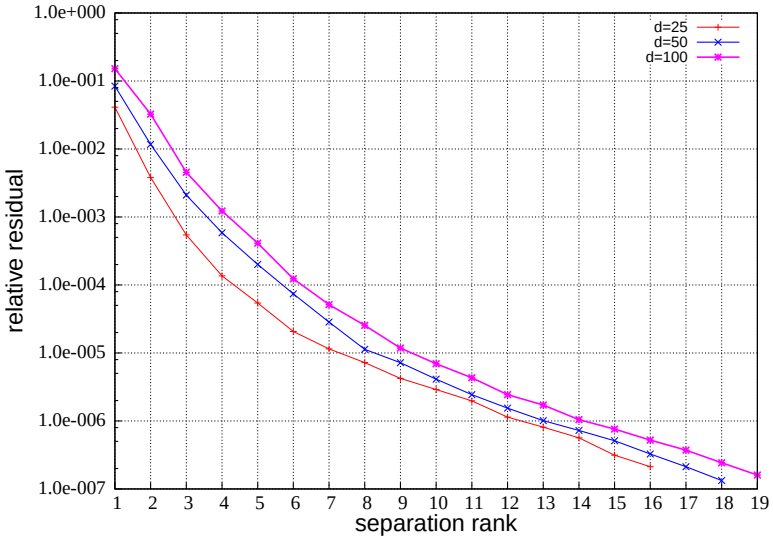

Figure 3: Relative residual error of the optimal low tensor rank approximation for the Poisson equation with Dirichlet boundary conditions and $n := 1000$.

is equivalent to the solution of a suitable eigenvalue problem. Let $\underline{i}^* := (i_1^*, \ldots, i_d^*)$ the multiindex where the maximum norm will appear, i.e. $|u_{\underline{i}^*}| = \|u\|_\infty$, and define

$$D(u) := \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} \operatorname{diag}\left((u_{j,\mu})_i\right)_{i=1}^{n}, \qquad E_{\underline{i}^*} := \bigotimes_{\mu=1}^{d} e_{i_\mu}^*,$$

where e_{i_μ} is the canonical unit vector from $\mathbb{R}^n$, with 1 in the i_μ-th entry and 0 elsewhere. We have

$$\begin{aligned} D(u)E_{\underline{i}^*} &= \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} \operatorname{diag}\left((u_{j,\mu})_i\right)_{i=1}^{n} e_{i_\mu^*} = \sum_{j=1}^{r} \bigotimes_{\mu=1}^{d} (u_{j,\mu})_{i_\mu^*} e_{i_\mu^*} \\ &= \sum_{j=1}^{r} \prod_{\mu=1}^{d} (u_{j,\mu})_{i_\mu^*} \bigotimes_{\mu=1}^{d} e_{i_\mu^*} = u_{\underline{i}^*} E_{\underline{i}^*}. \end{aligned}$$

We have thus arrived at an eigenvalue problem $D(u)E_{\underline{i}^*} = u_{\underline{i}^*} E_{\underline{i}^*}$ for the given matrix $D(u)$, where we are looking for the eigenvalue $u_{\underline{i}^*}$ and corresponding eigenvector $E_{\underline{i}^*}$. It is remarkable that the tensor rank of the eigenvector is exactly one. Moreover, the tensor structure is significant, since $E_{\underline{i}^*}$ is a Kronecker product of canonical unit vectors. For the numerical test, we create a tensor $\tilde{u}$

Algorithm 3 Penalty Method

1: Choose initial $\mathbf{u}^0 \in V^{r\times d}$, $\lambda_o \in \mathbb{R}_{>0}$ and parameter $\varepsilon \in \mathbb{R}_{>0}$, $k := 0$.
2: **repeat**
3: Compute the solution $\mathbf{u}^{k+1}$ of the unconstrained minimization problem

$$\min_{\mathbf{u}\in V^{r\times d}} L(\mathbf{u}, \lambda_k)$$

by using the CG method and u^k as an initial guess, see algorithm 1.
4: Choose $\lambda_{k+1} > \lambda_k$, $k \mapsto k+1$.
5: **until** $|h(\mathbf{u}^{k+1})| < \varepsilon$

with separation rank 5 where all entries are initialized with uniformly distributed pseudo-random numbers in the interval $[-2, 0]$. In addition, we create a tensor $v := -(4 + \tilde{u}_{\underline{i}^*})E_{\underline{i}^*}$ where the multi-index $\underline{i}^*$ is also randomly generated. With the definition $u := \tilde{u} + v$ we make sure that $|u^*_{\underline{i}}| = \|u\|_\infty = 4$ and -4 is the smallest eigenvalue of $D(u)$ with the corresponding eigenvector $E_{\underline{i}^*}$. Hence the computation of the maximum norm is a good application for our second model example. In Table 1 the results of our numerical experiments are presented for $d \in \{25, 50, 100\}$ and $n := 250$. Since the penalty method uses the CG algorithm several times we count the total number of iterations in the CG method. In all calculations, we observe a moderate number of iterations and a good approximation of the solution of the eigenvalue problem. A typical decay of the gradient of the functional with respect to the number of CG iteration is presented in Figure 4.

d	$\frac{\left\|4+\left\langle D(u)E_{\underline{i}^*}, E_{\underline{i}^*}\right\rangle\right\|}{4}$	$\|J'(E_{\underline{i}^*})\|$	Overall CG iterations	Time [Sec.]
25	2.42×10^{-7}	2.49×10^{-11}	79	0.94
50	1.53×10^{-7}	6.22×10^{-11}	54	1.46
100	2.88×10^{-7}	2.91×10^{-11}	61	3.26

Table 1: Computation of the maximum norm of u for different d and n=250.

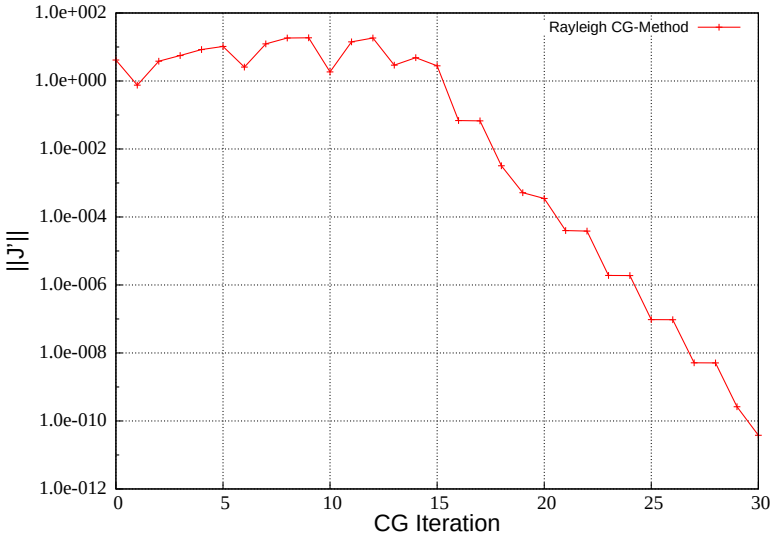

Figure 4: The decay of the gradient $\|J'(u^k)\|_2$ for $d := 50$, $\lambda_0 = 10$ (first call of the CG algorithm in the penalty method) and $n := 250$.

References

[1] G. Beylkin and M. J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proceedings of the National Academy of Sciences*, 99(16):10246–10251, 2002.

[2] G. Beylkin and M. J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM Journal on Scientific Computing*, 26(6):2133–2159, 2005.

[3] D. Braess and W. Hackbusch. Approximation of $1/x$ by Exponential Sums in $[1, \infty)$. *IMA Numer. Anal.*, 25:685–697, 2005.

[4] V. de Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. Technical Report SCCM-06-06, Stanford University, 2006.

[5] M. Espig. *Effiziente Bestapproximation mittels Summen von Elementartensoren in hohen Dimensionen*. PhD thesis, Universität Leipzig, 2008.

[6] M. Espig, L. Grasedyck, and W. Hackbusch. Black box low tensor rank approximation using fibre-crosses. *Constructive approximation*, 2009.

[7] M. Espig and W. Hackbusch. Characteristics and existence of best approximations with sums of elementary tensors. *in preparation*, 2009.

[8] M. Espig and W. Hackbusch. A new modified newtons method for the efficient low tensor rank approximation with sums of elementary tensors in high dimensions. *in preparation*, 2009.

[9] M. Espig and W. Hackbusch. A preconditioned conjugate gradient method for the efficient low tensor rank approximation with sums of elementary tensors in high dimensions. *in preparation*, 2009.

[10] Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig. http://www.mis.mpg.de/scicomp/exp_sum.

[11] I. P. Gavrilyuk, W. Hackbusch, and B. N. Khoromskij. Hierarchical tensor-product approximation to the inverse and related operators for high-dimensional elliptic problems. *Computing*, 74(2):131–157, 2005.

[12] L. Grasedyck. Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure. *Computing*, 72:247–265, 2004.

[13] W. H. Greub. *Multilinear Algebra*. Springer, 1967.

[14] W. Hackbusch and B. N. Khoromskij. Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. part I. Separable approximation of multi-variate functions. *Computing*, 76(3/4):177–202, 2006.

[15] W. Hackbusch and B. N. Khoromskij. Low-rank Kronecker-product approximation to multi-dimensional nonlocal operators. part II. HKT representation of certain operators. *Computing*, 76(3/4):203–225, 2006.

[16] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov. Approximate iterations for structured matrices. *Numerische Mathematik*, 109(3):365–383, 2008.

[17] P. Kosmol. A new class of derivative-free procedures for finding zeros of a function. *Computing*, 1993.

[18] A. Uschmajew. Convex maximization problems on non-compact stiefel manifolds with application to orthogonal tensor approximations. *submitted to Num. Math.*, 2009.

[19] T. Yokonuma. *Tensor Spaces and Exterior Algebra*. American Mathematical Society, 1991.