Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Use of Tensor Formats in Elliptic Eigenvalue Problems

by

Wolfgang Hackbusch, Boris N. Khoromskij, Stefan A. Sauter, and Eugene E. Tyrtyshnikov

Preprint no.: 78 2008

Use of Tensor Formats in Elliptic Eigenvalue Problems

Wolfgang Hackbusch, Boris N. Khoromskij*
Max-Planck-Institut für Mathematik in den Naturwissenschaften,
Inselstr. 22-26, D-04103 Leipzig, Germany; {wh, bokh}@mis.mpg.de

Stefan Sauter
Universität Zürich, Switzerland; stas@math.uzh.ch

Eugene E. Tyrtyshnikov†
Institute of Numerical Mathematics, Russian Academy of Sciences,
Gubkina 8, 119333 Moscow, Russia; tee@inm.ras.ru

Abstract

We investigate approximations by finite sums of products of functions with separated variables to eigenfunctions of multivariate elliptic operators, and especially conditions providing an exponential decrease of the error with respect to the number of terms. The results of the consistent use of tensor formats can be regarded as a base for a new class of iterative eigensolvers with almost linear complexity in the univariate problem size.

The results of numerical experiments clearly indicate the linear-logarithmic scaling of low-rank tensor method in the univariate problem size. The algorithms work equally well for the computation of both, minimal and maximal eigenvalues of the discrete elliptic operators.

AMS Subject Classification: 65F30, 65F50, 65N35, 65F10
Key words: elliptic operators, spectra, eigenfunctions, separation of variables, separable approximations, matrix approximations, low-rank matrices, Kronecker products, multi-dimensional matrices, tensors, numerical methods, iterative algorithms

1 Introduction

Recent advances in tensor approximation methods applied to the functions and operators in $\mathbb{R}^d$ lead to the natural idea of solving multi-dimensional boundary and eigenvalue problems in tensor-product formats. This idea has been first time formulated in [2] in the very general setting. The recent results on efficient methods of tensor approximation and their application in electronic structure calculations can be found in [11, 3, 8, 15, 18] and others.

In this paper we investigate approximations by finite sums of products of functions with separated variables to eigenfunctions of elliptic operators with smooth coefficients, and especially conditions providing an exponential decrease of error in the number of terms.

We consider the model eigenvalue problem: Find a pair $(\lambda, u) \in \mathbb{R} \times H_0^1(\Omega) \setminus \{0\}$ such that

$$\begin{array}{rl} \Lambda u = \lambda u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{array} \tag{1.1}$$

with the elliptic differential operator Λ of the form

$$\Lambda := -\operatorname{div}(A \operatorname{grad} u) + \langle b, \nabla u \rangle + cu, \tag{1.2}$$

*The work of the second author was partially supported by FIM, ETH Zürich during his visit of the ETH/University of Zürich in May - July, 2008.

†This work was partially supported by DFG, Russian Fund of Basic Research (grant 08-01-00115) and a Priority Research Grant of the Department of Mathematical Sciences of the Russian Academy of Sciences.

where $\Omega \in \mathbb{R}^d$ is some bounded or unbounded tensor-product domain, and the operator coefficients A, b, c in (1.2) are supposed to be smooth (analytic) in Ω.

We will prove that the eigenfunctions of problem (1.1) allow separable approximation that converges exponentially in the number of terms.

Problem (1.1) is discretised by the Galerkin FEM with tensor-product basis functions, so that the arising stiffness and mass matrices of size $n^{\otimes d}$ (with d-fold product $n^{\otimes d} = n \times ... \times n$) are represented in the low Kronecker rank format with the storage requirements and computational complexity of order $O(dn)$.

In this paper, for the ease of presentation, we use simple iterative solvers such as the power method or the Lanczos iteration, though algorithms of better choice can be easily adapted to our concept. Due to the above mentioned approximation results for the continuous solutions, and relying on the rank-structured representation of all matrices involved, we propose to solve the corresponding high-dimensional algebraic eigenvalue problem of the size $n^{\otimes d}$ in the low tensor-rank format. To this end, we introduce the so-called "truncated iterations", where most of the intermediate vectors have to be approximated in some fixed rank-structured tensor product form. The corresponding rank truncation performed at each iteration makes use of recently developed methods (cf. [18, 22, 8]). For the class of rank structured matrices, our algorithm can be shown to have storage and complexity bounds of order $O(r^d n + rnd)$, or even $O(dRrn)$, where r, R are the small (often fixed) rank parameters with the theoretical bounds $r = O(\log n)$ and $R = O(\log n |\log \varepsilon|)$.

We notice that in the case $d = 2$ the rank truncation operator is realised by the "truncated SVD" method applied to the rank-R matrix. It is, in practice, a finite algorithm (with complexity at most $O(nR^2 + R^3)$) providing the best rank-r approximation to the current iterand (see numerics in Sections 2 and 6.2). Hence, numerical results for $d = 2$ can be viewed as the reference cases, demonstrating nearly optimal performance of the proposed techniques.

The results of the consistent use of tensor formats can be regarded as a base for a new class of iterative eigensolvers with almost linear complexity in the univariate problem size.

The rest of the paper is organised as follows. In Section 2 we give motivating numerical illustrations for the Laplace operator in $d = 2$, which clearly indicate a spectacular gain by the truncated Lanczos iteration compared with the standard full-format implementation. In Section 3, we briefly describe the tensor product formats for representing multivariate functions of the continuous and discrete arguments. Section 4 proves the existence of separable approximation for the eigenfunctions of an elliptic operator posed in $\mathbb{R}^d$. This result is of principal significance for understanding the rigorous mathematical basis for applying tensor methods in multidimensional setting. Section 5 describes the favourable properties of tensor formats in discrete elliptic eigenvalue problems, while Section 6 presents numerical illustrations for certain spectral problems in dimensions $d = 2, 3$.

2 Motivating Numerics

In this research we are motivated by the nice solution structure for the 2D Laplace operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

and the eigenproblem

$$-\Delta u(x, y) = \lambda u(x, y), \quad (x, y) \in [0, \pi]^2.$$

If u is zero on the boundary, then the eigenvalues and eigenfunctions are

$$\lambda_{kl} = k^2 + l^2, \quad u_{kl}(x, y) = \sin kx \ \sin ly, \quad k, l = 1, 2, \dots .$$

A matrix counterpart of the Laplace operator can be taken in the form

$$Mu = \lambda u, \quad M = A \otimes I + I \otimes A,$$

where $\otimes$ denotes the Kronecker (tensor) product, I is the identity and

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \dots & \dots & \dots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

Let A and I be of order n. Then the matrix eigenvalue problem $Mx = \lambda x$ can be solved explicitly. The eigenvalues are

$$\lambda_{kl} = 4\sin^2\frac{\pi k}{2(n+1)} + 4\sin^2\frac{\pi l}{2(n+1)}, \quad 1 \le k, l \le n,$$

and the corresponding eigenvectors are exactly represented as tensor products:

$$x_{kl} = u^k \otimes v^l,$$

where u^k and v^l are n-dimensional vectors with the entries

$$u^k_s = \sin\frac{\pi k s}{n+1}, \quad 1 \le s \le n; \qquad v^l_t = \sin\frac{\pi l t}{n+1}, \quad 1 \le t \le n.$$

In general, the cost of traditional iterative eigenvalue algorithms is higher than linear, possibly $O(n^2)$. The tensor structure of eigenvectors allows to modify the eigensolvers so that the cost of one iteration reduces to $O(n)$.

As an example of an iterative eigensolver we discuss the Lanczos method.

Algorithm 2.1 *Let M be a real symmetric matrix of order N. The following steps produce the Ritz values approximating the eigenvalues of M:*

- *Choose an initial vector p_1 with $||p_1|| = 1$ and set $p_0 = 0$, $b_0 = 0$.*
- *For $k = 1, 2, ...$ compute*

$$\begin{aligned} z_k &= Mp_k, \\ a_k &= (z_k, p_k), \\ q_k &= z_k - a_k p_k - b_{k-1}p_{k-1}, \\ b_k &= ||q_k||, \\ p_{k+1} &= q_k / b_k. \end{aligned}$$

- *Compute the Ritz values as the eigenvalues of the projected $k \times k$ matrix*

$$M_k = P_k^\top M P_k, \quad P_k = [p_1, ..., p_k],$$

which is the symmetric tridiagonal matrix consisting of the values a_k, b_k.

Now we consider the same computations with the following vector truncation. If $x \in \mathbb{R}^{n^2}$, $T_\varepsilon(x)$ is defined as a vector y of the form

$$y = \sum_{t=1}^{r} u_t \otimes v_t, \quad u_t, v_t \in \mathbb{R}^n,$$

with minimal r such that

$$||x - y|| \le \varepsilon$$

for a fixed ε. Note that if x is in the above tensor format and a matrix M is of the tensor form (cf. Section 3.5)

$$M = \sum_{s=1}^{R} A_s \otimes B_s, \quad R \ll n,$$

then the matrix-vector multiplication $z = My$ has a tensor format with Rr summands. A natural idea to reduce this number is to replace z with $T_\varepsilon(z)$. Then, the Lanczos algorithm transforms to the following tensor-format scheme.

Algorithm 2.2 *The Lanczos steps with tensor approximation of computed vectors reads as follows:*

- *Choose an initial vector p_1 so that $p_1 = T_\varepsilon(p_1)$ with $||p_1|| = 1$ and set $p_0 = 0$, $b_0 = 0$.*

- *For $k = 1, 2, \ldots$ compute*

$$\begin{aligned} z_k &= Mp_k, \\ a_k &= (z_k, p_k), \\ q_k &= T_\varepsilon(z_k - a_k p_k - b_{k-1} p_{k-1}), \\ b_k &= ||q_k||, \\ p_{k+1} &= q_k / b_k. \end{aligned}$$

- *Compute the Ritz values from the generalized eigenvalue problem*

$$M_k u = \lambda N_k u,$$

where $M_k = P_k^\top M P_k$, $N_k = P_k^\top P_k$.

We can compare the performance of these two algorithms. The truncation parameter was chosen so that every vector in the Krylov subspace basis is approximated in the tensor format so that the tensor rank (the number of tensor-product terms in the sum) does not exceed 7 or the relative approximation accuracy is less than 10^{-2}. In the table below we can see the timings obtained on a 2.7 GHz workstation:

n	1000	2000	4000	6000
Lanczos time (sec)	2.8	12.1	76.7	224.9
Tensor Lanczos time (sec)	0.4	0.7	1.5	2.2

The accuracy in the maximal eigenvalue was about 10^{-3} in both methods, even a bit better with truncations. The table shows that the time of the tensor-truncated Lanczos grows linearly in n while the original Lanczos time increases actually faster as n^2. For $n = 6000$ we observe an acceleration by a factor 100.

Thus, we are interested to find if we might have the same effect when using tensor formats in matrix eigensolvers for more general elliptic problems,

$$\Lambda u = \lambda u, \quad u : [0, \pi]^d \to \mathbb{R},$$

where Λ is the elliptic operator with smooth enough (or separable) coefficients. For this purpose, we next introduce the tensor representations.

3 Description of tensor formats

3.1 Tensor spaces and tensor representations

Several continuous and discrete spaces considered in this paper are tensor spaces of order d, where in our application d equals the spatial dimension of the eigenvalue problem. Let

$$\mathbb{W} = W_1 \otimes W_2 \otimes \ldots \otimes W_d \tag{3.1}$$

be the notation for the underlying tensor space. By definition, each $w \in \mathbb{W}$ can be written as a sum

$$w = \sum_k w_k^{(1)} \otimes w_k^{(2)} \otimes \ldots \otimes w_k^{(d)} \qquad (w_k^{(j)} \in W_j). \tag{3.2}$$

An interesting subset are those elements which require only R terms. They form the set

$$\mathcal{M}_R = \left\{ w \in \mathbb{W} : \; w = \sum_{k=1}^{R} w_k^{(1)} \otimes w_k^{(2)} \otimes \ldots \otimes w_k^{(d)}, \; w_k^{(j)} \in W_j \right\}.$$

We say that elements $w \in \mathcal{M}_R$ with $w \notin \mathcal{M}_{R-1}$ have the *tensor rank* R. Obviously, tensors $w \in \mathcal{M}_R$ can be represented by the description of Rd elements $w_k^{(j)} \in W_j$. Hence, the cost is linear in d.

As in the Galerkin method, the replacement of the spaces W_j by subspaces $V_j \subset W_j$ $(1 \le j \le d)$ is of interest. They lead to the tensor subspace

$$\mathbb{V} = V_1 \otimes V_2 \otimes \ldots \otimes V_d \subset \mathbb{W}. \tag{3.3}$$

Setting $r_j := \dim V_j$ and choosing a (without loss of generality orthonormal) basis $\left\{\phi_k^{(j)} : 1 \le k \le r_j\right\}$ of V_j, we can represent each $v \in \mathbb{V}$ by

$$v = \sum_{\mathbf{k}} b_{\mathbf{k}} \phi_{k_1}^{(1)} \otimes \phi_{k_2}^{(2)} \otimes \ldots \otimes \phi_{k_d}^{(d)}, \tag{3.4}$$

where the multi-index $\mathbf{k} = (k_1, \ldots, k_d)$ runs over all $1 \le k_j \le r_j$, $1 \le j \le d$.

Let $\mathbf{r} = (r_1, \ldots, r_d) \in N^d$ be a tuple of dimensions. We define

$$\mathcal{M}_{\mathbf{r}} = \left\{ v \in \mathbb{W} \text{ as in (3.4) with } b_{\mathbf{k}} \in \mathbb{R},\ \{\phi_k^{(j)} : 1 \le k \le r_j\} \subset W_j \text{ orthonormal system} \right\}.$$

A representation of $w \in \mathcal{M}_{\mathbf{r}}$ in the form (3.4) is also called a Tucker representation of Tucker rank $\mathbf{r}$. A representation of $w \in \mathcal{M}_{\mathbf{r}}$ in the Tucker format (3.4) requires $\prod_{j=1}^d n_j$ reals and the representation of $\sum_{j=1}^d n_j$ vectors $\phi_k^{(j)}$.

3.2 Tensor approximation, tensor truncation

For any element $w \in \mathbb{W}$ there are numbers R and tuples $\mathbf{r}$ such that $w \in \mathcal{M}_{R^*}$ and $w \in \mathcal{M}_{\mathbf{r}^*}$, but the ranks R^* and $\mathbf{r}^*$ may be rather large. A representation by one of the tensor formats (3.2) or (3.4) is only of interest if the respective ranks are small enough. Therefore, given $w \in \mathbb{W}$ we search for approximations $v \in \mathcal{M}_R$ or $v \in \mathcal{M}_{\mathbf{r}}$ with suitably small ranks R or $\mathbf{r}$. When we fix the set $\mathcal{S} := \mathcal{M}_R$ or $\mathcal{S} := \mathcal{M}_{\mathbf{r}}$, the smallest error is described by

$$\sigma(w, \mathcal{S}) := \inf_{v \in \mathcal{S}} \|w - v\|.$$

In the following we will give examples where $\sigma(w, \mathcal{S})$ decays exponentially with the rank R or $\min\{r_j : 1 \le j \le d\}$, respectively. In general, the infimum in the definition of $\sigma(w, \mathcal{S})$ cannot be replaced by a minimum, since for $\mathcal{S} = \mathcal{M}_R$ a minimiser is not necessarily existing. Although, the minimiser exists for $\mathcal{S} = \mathcal{M}_{\mathbf{r}}$, its computation can be performed only approximately. Therefore, in practice, one has to determine a $v \in \mathcal{S}$ such that $\|w - v\|$ comes close to $\sigma(w, \mathcal{S})$. The replacement of w by such a $v \in \mathcal{S}$ is called the *tensor truncation* to $\mathcal{S}$ and denoted by

$$w \mapsto v = \begin{cases} T_R w & \text{if } \mathcal{S} = \mathcal{M}_R, \\ T_{\mathbf{r}} w & \text{if } \mathcal{S} = \mathcal{M}_{\mathbf{r}}. \end{cases} \tag{3.5}$$

Heuristic methods for computing the rank structured approximations in different problem settings are discussed in [5, 19, 27, 18, 8, 7].

In the particular case of $d = 2$, the difficulties mentioned above do not appear. The minimiser of $\inf_{v \in \mathcal{S}} \|w - v\| = \min_{v \in \mathcal{S}} \|w - v\|$ is the result of the truncated singular value decomposition. Furthermore, the representations by $\mathcal{M}_R$ and $\mathcal{M}_{\mathbf{r}}$ with $\mathbf{r} = (R, R)$ coincide.

3.3 Application to function spaces

Let $I = I_1 \times I_2 \times \ldots \times I_d$ be the product of (possibly infinite) intervals $I_j \subset \mathbb{R}$. Then $L^2(I)$ is the tensor space $L^2(I_1) \otimes L^2(I_2) \otimes \ldots \otimes L^2(I_d)$. The tensor product $w = \bigotimes_{j=1}^d w^{(j)}$ of $w^{(j)} \in L^2(I_j)$ corresponds to the pointwise product $w(x) = \prod_{j=1}^d w^{(j)}(x_j)$.

If w is an analytical function in all variables x_j, approximations by polynomials may lead to small errors. In the case of a uniform degree $r - 1$, the subspaces $V_j \subset W_j$ in (3.3) are $\mathbb{P}_{r-1}$ and any $v \in \mathbb{V} = \bigotimes_{j=1}^d V_j$ has the Tucker rank $\mathbf{r} = (r, \ldots, r)$. The error, which is an upper bound of $\sigma(w, \mathcal{M}_{\mathbf{r}})$, depends on the decay of the higher derivatives. The analysis in Section 5.1 will show exponential decay of $\sigma(w, \mathcal{M}_{\mathbf{r}})$.

Multivariate functions depending on the Euclidean norm as, e.g., the classical potentials $1/\|x\|$, $e^{-\lambda\|x\|}/\|x\|$, $e^{-\lambda\|x\|}$ can be rather well approximated in $\mathcal{M}_R$ leading to exponential decay of $\sigma(w, \mathcal{M}_R)$ with respect to R. For its computation and analysis see [12, 11, 23, 14, 17, 15, 4].

3.4 Application to grid functions

Galerkin discretisations with tensor product basis functions or finite difference schemes lead to grid points $x_{\mathbf{i}} = \left(x_{i_1}^{(1)}, \ldots, x_{i_d}^{(d)}\right)$, where $\mathbf{i} \in I = I_1 \times \ldots \times I_d$. Hence, the grid values $u(x_{\mathbf{i}}) = u_{\mathbf{i}}$ belong to $\mathbb{R}^I$ which is the tensor space

$$\mathbb{R}^I = \mathbb{R}^{I_1} \otimes \mathbb{R}^{I_2} \otimes \ldots \otimes \mathbb{R}^{I_d},$$

i.e., $W_j = \mathbb{R}^{I_j}$ from (3.1) have the dimension $n_j := \#I_j$. For simplicity we assume $n_j = n$ for all $1 \le j \le d$.

The representation of $w \in \mathcal{M}_R$ needs a storage of Rdn, while $w \in \mathcal{M}_{\mathbf{r}}$ with $\mathbf{r} = (r, \ldots, r)$ requires $rdn + r^d$ data.

3.5 Application to matrices

The index sets $I_1, \ldots, I_d$ and $J_1, \ldots, J_d$ give rise to the two tensor spaces $X := \mathbb{R}^{I_1} \otimes \ldots \otimes \mathbb{R}^{I_d}$ and $Y := \mathbb{R}^{J_1} \otimes \ldots \otimes \mathbb{R}^{J_d}$. Given matrices $A^{(j)} \in \mathbb{R}^{I_j \times J_j}$ $(1 \le j \le d)$, its *Kronecker product* $\mathcal{A} := A^{(1)} \otimes \ldots \otimes A^{(d)}$ is defined as the mapping

$$\mathcal{A} : X \to Y, \qquad x = x^{(1)} \otimes \ldots \otimes x^{(d)} \mapsto \mathcal{A}x = A^{(1)}x^{(1)} \otimes \ldots \otimes A^{(d)}x^{(d)} \in Y.$$

4 Regularity for elliptic eigenvalue problems

4.1 Polynomial approximation of analytic functions

To understand the separability property of eigenfunctions we analyse their regularity with respect to some classes of functions which allow the holomorphic extension to the complex plane.

The error estimates will be derived for the function set $\mathcal{A}_{M,\rho}(I)$ and for its multidimensional counterpart $\mathcal{A}_{M,\rho}(I^d)$, $d \ge 2$. The definition of the space $\mathcal{A}_{M,\rho}\left(I^d\right)$ requires several steps. For the interval $I := (-1, 1)$ and $\rho > 1$, Bernstein's regularity ellipse is given by (cf. [1])

$$\mathcal{E}_\rho := \{z \in \mathbb{C} : |z - 1| + |z + 1| \le \rho + \rho^{-1}\}.$$

Its semi-axes are $a = \frac{\rho+\rho^{-1}}{2}$ and $b = \frac{\rho-\rho^{-1}}{2}$, implying $a + b = \rho$.

Definition 4.1 *Let $I = (-1, 1)$ and $M > 0$, $\rho > 1$ be given constants. $\mathcal{A}_{M,\rho}(I)$ is the class of functions $f \in C^\infty(I)$ having a holomorphic extension to $\mathcal{E}_\rho(I)$ such that*

$$|f(z)| \le M \qquad \forall z \in \mathcal{E}_\rho(I).$$

Next, we introduce the multidimensional analogue of $\mathcal{A}_{M,\rho}(I)$ on the tensor domain $I^d := (-1, 1)^d$. Let $\mathcal{E}_\rho^{(j)} := I \times ... \times I \times \mathcal{E}_\rho \times I \times ... \times I$ with $\mathcal{E}_\rho$ to be inserted at the jth position.

Definition 4.2 *For given constants $M > 0$, $\rho > 1$, the set $\mathcal{A}_{M,\rho}(I^d)$ consists of all functions $f \in C^\infty(I^d)$ having holomorphic extensions to $\mathcal{E}_\rho^{(j)}$, for all $1 \le j \le d$, and satisfying*

$$\max_{1 \le j \le d} \{ \sup_{x \in \mathcal{E}_\rho^{(j)}} |f(x)| \} \le M.$$

The following remark recalls the well-known fact that controlling all higher derivatives of a function implies that it belongs to $\mathcal{A}_{M,\rho}(I)$ (see e.g. [20] for the proof).

Remark 4.3 *Assume that a function $u : I \to \mathbb{R}$ satisfies for some C_u, $\gamma_u \ge 0$*

$$\left\| \frac{\partial^n u}{\partial x^n} \right\|_{L^\infty(I)} \le C_u \gamma_u^n n! \quad \text{for all } n \in \mathbb{N}_0. \tag{4.1}$$

Then $u \in \mathcal{A}_{M,\rho}(I)$ holds with $\rho = 1 + \gamma_u^{-1} > 1$, $M = C \cdot C_u$.

For the continuous multivariate functions $f = f(x_1, ..., x_d) : \mathbb{R}^d \to \mathbb{R}$, we use the tensor product interpolant

$$\mathbf{I}_N f = I_N^1 ... I_N^d f \in P_N[I^d],$$

where $I_N^i f$ $(1 \le i \le d)$ denotes the interpolation polynomial of degree N with respect to the variables $x_i \in I := [-1, 1]$ interpolating f with respect to the variables $x_i \in I := [-1, 1]$ at the Chebyshev nodes.

Proposition 4.4 *Let $M > 0$ and $\rho > 1$ be given. For all $f \in \mathcal{A}_{M,\rho}(I^d)$ and $N > 1$ the estimate*

$$\|f - \mathbf{I}_N f\|_{C^0(I^d)} \le c\, M\, (\log N)^d\, \rho^{-N} \tag{4.2}$$

holds.

In the next section we derive the regularity results for solutions of elliptic eigenvalue problems which will imply Remark 4.3.

4.2 Regularity for elliptic eigenvalue problems with smooth data

Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. For $\ell \in \mathbb{N}$, we define the norms

$$\left\|\nabla^\ell u\right\|^2_{L^2(\Omega)} := \sum_{|\alpha|=\ell} \frac{\ell!}{\alpha!} \left\|D^\alpha u\right\|^2_{L^2(\Omega)} \quad \text{and} \quad \left\|\nabla^\ell u\right\|_{L^\infty(\Omega)} := \left\| \sqrt{\sum_{|\alpha|=\ell} \frac{\ell!}{\alpha!} \left|D^\alpha u\right|^2} \right\|_{L^\infty(\Omega)}.$$

For u replaced by vectors or matrices, the absolute value $|\cdot|$ is to be replaced by the Euclidean or spectral norm.

Consider the eigenvalue problem for the differential operator

$$\Lambda u := -\nabla \cdot (A \nabla u) + \langle b, \nabla u \rangle + cu, \tag{4.3}$$

where $\langle a, b \rangle := \sum_{i=1}^d a_i b_i$ (without complex conjugation). We say that Λ is *uniformly elliptic* if $A \in \mathbf{L}^\infty_{\mathrm{sym}}\left(\Omega, \mathbb{R}^{d\times d}\right)$, $b \in \mathbf{L}^\infty\left(\Omega, \mathbb{R}^d\right)$, and $c \in L^\infty(\Omega)$ satisfy

$$0 < a_{\min} := \inf_{x\in\Omega} \inf_{v\in\mathbb{C}^d\setminus\{0\}} \frac{\langle A(x)\, v, \overline{v}\rangle}{\langle v, \overline{v}\rangle} \le \sup_{x\in\Omega} \sup_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle A(x)\, v, \overline{v}\rangle}{\langle v, \overline{v}\rangle} =: a_{\max} < \infty \tag{4.4a}$$

$$0 \le -\frac{1}{2} \operatorname{div} b + c. \tag{4.4b}$$

We assume that the coefficients of the operator Λ are *analytic*; i.e., there exist positive constants $C_{\mathcal{A}}$, C_b, C_c and $\gamma_{\mathcal{A}}$, γ_b, γ_c such that

$$\begin{aligned} \left\|\nabla^p A\right\|_{L^\infty(\Omega)} &\le C_{\mathcal{A}} \gamma_{\mathcal{A}}^p p! \quad \forall p \in \mathbb{N}_0, \\ \left\|\nabla^p b\right\|_{L^\infty(\Omega)} &\le C_b \gamma_b^p p! \quad \forall p \in \mathbb{N}_0, \\ \left\|\nabla^p c\right\|_{L^\infty(\Omega)} &\le C_c \gamma_c^p p! \quad \forall p \in \mathbb{N}_0. \end{aligned} \tag{4.5}$$

We consider two types of domain: Either $\Omega = \mathbb{R}^d$ or Ω is a bounded Lipschitz domain with analytic boundary, i.e., there is a finite family $\mathcal{U}$ of open subset in $\mathbb{R}^d$ along a family of bijective maps[1] $\left\{\chi_U : \overline{B_1} \to \overline{U}\right\}_{U\in\mathcal{U}}$ such that

$$\begin{aligned} &\forall U \in \mathcal{U}: && \chi_U \in C^{0,1}\left(\overline{B_1}, \overline{U}\right), \quad \chi_U^{-1} \in C^{0,1}\left(\overline{U}, \overline{B_1}\right), \\ &\forall U \in \mathcal{U}: && \chi_U\left(B_1^0\right) = U \cap \partial\Omega, \quad \chi_U\left(B_1^+\right) = U \cap \Omega, \quad \chi_U\left(B_1^-\right) = U \cap \mathbb{R}^d \setminus \overline{\Omega}, \\ &\exists C_\Gamma, \gamma_\Gamma \quad \forall U \in \mathcal{U}: && \left\|\nabla^p \chi_U\right\|_{L^\infty(B_1)} \le C_\Gamma \gamma_\Gamma^p p! \quad \forall p \in \mathbb{N}_0. \end{aligned} \tag{4.6}$$

Let us consider the eigenvalue problem: Find a pair $(\lambda, u) \in \mathbb{R} \times H_0^1(\Omega) \setminus \{0\}$ such that

$$\begin{aligned} \Lambda u &= \lambda u \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{4.7}$$

with Λ as in (4.3). Let $E(\lambda)$ denote the eigenspace for the eigenvalue λ.

[1] B_1 denotes the unit ball in $\mathbb{R}^d$ and $B_1^0 := \{x \in B_1 \mid x_d = 0\}$. For $\sigma \in \{+, -\}$, we set $B_1^\sigma := \{x \in B_1 \mid \sigma x_d > 0\}$.

Theorem 4.5 *Let Ω be an analytic, bounded Lipschitz domain which satisfies (4.6). Assume that the coefficients A, b, c satisfy (4.5). Then, any eigenfunction $u \in E(\lambda)$ of (4.7) (normalised to $\|u\|_{L^2(\Omega)} = 1$) is analytic. There exist constants $C, K > 0$ depending only on the constants in (4.5), (4.6), and on $a_{\min}$ and the spatial dimension d such that*

$$\left\|\nabla^{p+2} u\right\|_{L^2(\Omega)} \leq C K^{p+2} \max\left\{p, \sqrt{|\lambda|}\right\}^{p+2} \qquad \text{for all } p \geq 0. \tag{4.8}$$

Proof. The statement follows from [20, Theorem 5.3.10] as follows. First, let $|\lambda| \geq 1$ and consider (4.7) as the equation

$$-\varepsilon^2 \nabla \cdot (A \nabla u) + \left\langle \tilde{b}, \nabla u \right\rangle + (\tilde{c} - 1)\, u = f \qquad \text{in } \Omega \quad \text{with} \quad u|_{\partial\Omega} = 0$$

where $\varepsilon^2 = \lambda^{-1}$, $\tilde{b} = \lambda^{-1} b$, $\tilde{c} = c/\lambda$ and $f \equiv 0$. For the quantity $\mathcal{E}$ in [20, Theorem 5.3.10] we obtain the estimate

$$\mathcal{E}^{-1} := C_b + \frac{\sqrt{1 + C_c/|\lambda|}}{|\lambda|^{-1/2}} + 1 \leq 1 + C_b + \sqrt{|\lambda| + C_c} \leq C_1 \sqrt{|\lambda|},$$

where $C_1 := 1 + C_b + \sqrt{1 + C_c}$. The other quantities which appear in [20, Theorem 5.3.10] have to be substituted by

$$C_f \leftarrow 0, \quad C_c \leftarrow C_c + 1, \quad \mathcal{E} \leftarrow C_2 |\lambda|^{-1/2}, \quad \left(\frac{\mathcal{E}}{\varepsilon}\right)^2 \leftarrow C_2^2$$

with $C_2 := \left(\sqrt{1 + C_c} + C_b\right)^{-1}$. From (4.4) we conclude that

$$\begin{aligned}
\operatorname{Re} a(u,u) &= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \operatorname{Re}\left(\langle b, \nabla u \rangle \overline{u}\right) + c |u|^2 \\
&= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \frac{1}{2} \left\langle b, \nabla \left(|u|^2\right) \right\rangle + c |u|^2 \\
&= \int_\Omega \langle A \nabla u, \nabla \overline{u} \rangle + \left(-\frac{1}{2} \operatorname{div} b + c\right) |u|^2 \\
&\overset{(4.4)}{\geq} a_{\min} \|\nabla u\|_{L^2(\Omega)}^2
\end{aligned}$$

holds. Since u is an eigenfunction corresponding to λ and $\|u\|_{L^2(\Omega)} = 1$ we obtain

$$\|\nabla u\|_{L^2(\Omega)} \leq a_{\min}^{-1/2} \sqrt{\operatorname{Re} a(u,u)} = \sqrt{(\operatorname{Re} \lambda)/a_{\min}}.$$

Plugging these quantities into the estimate in [20, Theorem 5.3.10] we get

$$\left\|\nabla^{p+2} u\right\|_{L^2(\Omega)} \leq C K^{p+2} \max\left\{p, \sqrt{|\lambda|}\right\}^{p+2},$$

where C only depends on the constants $C_{\mathcal{A}}$, C_b, C_c, $\gamma_{\mathcal{A}}$, γ_b, γ_c, C_Γ, γ_Γ, $a_{\min}$. As explained in [20, Remark 5.3.11] the coercivity assumption which is imposed in [20, Theorem 5.3.10] is not required for this estimate. The proof of [20, Theorem 5.3.10] covers only the case $d = 2$. However, the only part therein, where $d = 2$ (instead of general d) is used explicitly, is the mapping lemma [20, Lemma 4.3.1]. Inspection of the proof shows that the case $d \geq 3$ can be handled analogously while, then, the constants in (4.8) in general depend also on the spatial dimension d.

The case $|\lambda| < 1$ is even simpler because we consider

$$-\varepsilon^2 \nabla \cdot (A \nabla u) + \langle b, \nabla u \rangle + (c - \lambda)\, u = f \qquad \text{in } \Omega \quad \text{with} \quad u|_{\partial\Omega} = 0$$

where $f = 0$ and $\varepsilon^2 = 1$. By repeating the steps in the first part of the proof with coefficients $\tilde{b} = b$, $\tilde{c} = c - \lambda$ with $|\lambda| < 1$ we obtain

$$\left\|\nabla^{p+2} u\right\|_{L^2(\Omega)} \leq C\, (pK)^{p+2}.$$

■

Remark 4.6 *Assume the suppositions of Theorem 4.5 except that $\Omega = \mathbb{R}^d$. Then the estimate*

$$\left\|\nabla^{p+2} u\right\|_{L^2(\mathbb{R}^d)} \le C K^{p+2} \max\left\{p, \sqrt{|\lambda|}\right\}^{p+2} \qquad \text{for all } p \ge 0$$

follows from [20, Prop. 5.5.1] by a simple repetition of the arguments of the previous proof.

5 Tensor formats in elliptic eigenvalue problems

5.1 Error estimate for low tensor-rank approximation

Now we are in a position to apply the regularity results in Theorem 4.5 to derive the low separation rank approximations to certain class of eigenfunctions in (4.7).

Corollary 5.1 *Assume that the assumptions of Remark 4.6 are fulfilled and let $\Omega_b = (-b, b)^d$ be some fixed hypercube in $\mathbb{R}^d$. Then the estimate (4.1) holds in Ω_b for each variable $x_1, ..., x_d$, with*

$$\rho = 1 + C_1^{-1}, \quad M = C_2 \sqrt{|\lambda|},$$

and with $C_1, C_2 > 0$ depending only on the constants C, K in Theorem 4.5 and $|\lambda|$ and a.

Proof. To prove the assertion we check that the bound (4.8) implies Remark 4.3 with the respective constants C_u, γ_u, where u, λ is the eigenpair in (4.7). Due to the Sobolev embedding theorem, the L^2-bound (4.8) implies the corresponding estimate in L^∞ norm

$$\left\|\nabla^{p+2} u\right\|_{L^\infty(B)} \le C(d) K^{p+2} \max\left\{p, \sqrt{|\lambda|}\right\}^{p+2},$$

such that $C(d)$ may depend on the dimension d (see also discussion in [21]). Now the estimate (4.1) follows with the constants specified in the assertion. ∎

Now we are able to derive the separable approximations for a class of elliptic eigenvalue problems.

Theorem 5.2 *Let the assumptions of Corollary 5.1 be satisfied. Then for any finite box Ω_b there exists an element $u_{\mathbf{r}} \in \mathcal{M}_{\mathbf{r}}$ with $\mathbf{r} = (r, ..., r)$, such that for the eigenfunction in (4.7) we have*

$$\|u - u_{\mathbf{r}}\|_{C^0(\Omega_b)} \le c M \left(\log r\right)^d \rho^{-r}, \tag{5.1}$$

where $\rho = 1 + C_1^{-1}$ and $M = C|\lambda|$ as defined above. The related representation of $u_{\mathbf{r}}$ in $\mathcal{M}_R$ has at most rank $R = r^{d-1}$.

Proof. Corollary 5.1 ensures that for the eigenfunction we have $u \in \mathcal{A}_{M,\rho}(\Omega_b)$ with the respective constants M, ρ. Hence, we apply the tensor product interpolant $\mathbf{I}_{r-1} = I^1_{r-1} ... I^d_{r-1} u \in P_{r-1}[\Omega_b]$ on Ω_b with respect to d variables as in Proposition 4.4 and obtain the bound (4.2). The representation in the format $\mathcal{M}_R$ is obtained by reordering of rank-1 summands. This completes our proof. ∎

For problems posed in $\mathbb{R}^d$, Theorem 5.2 implies the asymptotic estimate

$$r = O(|\log \varepsilon| + \log |\lambda|)$$

of the Tucker rank r on a finite hypercube, where $\varepsilon > 0$ is the desired approximation error.

To conclude this section we observe that for a special class of spectral problems, Theorem 5.2 can be applied in the finite hypercube Ω_b defined above. Specifically, we assume that the eigenfunction of problem (4.7) with $\Omega = \mathbb{R}^d$ exhibits the exponential decay,

$$|u(x)| \le C e^{-\beta \|x\|} \quad \text{as} \quad x \to \infty. \tag{5.2}$$

Given $\varepsilon > 0$, consider the "approximating" eigenvalue problem posed in the hypercube Ω_b of size $b = O(|\log \varepsilon|)$, with the corresponding eigenfunction u_b, and suppose that $\|u - u_b\| \le C\varepsilon$.

Remark 5.3 *Corollary 5.1 ensures that the eigenfunction u_b allows the same upper bound on the Tucker and canonical ranks as in Theorem 5.2. Hence, in this case the truncated iteration can be applied directly to the problem in finite domain Ω_b. Such a situation is typical in computational quantum chemistry.*

5.2 Application to the discrete problem

We apply the Galerkin approximation with respect to the tensor product piecewise linear basis functions. In the following, we make use of tensor representation of functions as well as the Kronecker tensor product representation of the elliptic operator Λ. Let us formulate the assumptions on the coefficients which ensure the respective tensor representation.

We consider the elliptic operators of the form

$$-\operatorname{div}(A\operatorname{grad} u) + \langle b, u\rangle + cu \equiv \Lambda_A u + \Lambda_C u, \tag{5.3}$$

where $A = \{a_{ij}(x)\}_{i,j=1}^d \in \mathbf{L}^\infty_{\rm sym}\left(\Omega, \mathbb{R}^{d\times d}\right)$, $b \in L^\infty\left(\Omega, \mathbb{R}^d\right)$ and $c \in L^\infty(\Omega)$ satisfy the conditions listed in §4.2. To simplify the discussion, we further set $b = 0$, though all the tensor constructions apply to this case as well. In addition, we make the following assumptions which lead to the low Kronecker rank representation of the discrete operator (the Galerkin stiffness matrix). We assume from the very beginning that the operator coefficients are separable, $a_{ij}, c \in \mathcal{M}_R$, with some moderate rank parameter $R \in \mathbb{N}$, i.e.,

$$a_{ij}(x) = \sum_{k=1}^R a^{(1)}_{ij,k}(x_1)\cdots a^{(d)}_{ij,k}(x_d), \quad i,j = 1,...,d,$$

and

$$c(x) = \sum_{k=1}^R c^{(1)}_k(x_1)\cdots c^{(d)}_k(x_d).$$

Hence, in the case of rank-1 test and trial functions $u(x) = \Pi_{\ell=1}^d u^{(\ell)}(x_\ell)$, $v(x) = \Pi_{\ell=1}^d v^{(\ell)}(x_\ell)$, the associated bilinear forms can be written using the product ansatz as follows

$$\langle \Lambda_A u, v\rangle_{L^2} = \sum_{k=1}^R \sum_{i,j=1}^d \prod_{\ell=1}^d \left\langle a^{(\ell)}_{ij,k}(x_\ell)\frac{\partial^{\delta_{j\ell}}}{\partial x_\ell}u^{(\ell)}(x_\ell), \frac{\partial^{\delta_{i\ell}}}{\partial x_\ell}\overline{v}^{(\ell)}(x_\ell)\right\rangle_{L^2(\Omega)}, \tag{5.4}$$

$$\langle \Lambda_C u, v\rangle_{L^2} = \sum_{k=1}^R \prod_{\ell=1}^d \left\langle c^{(\ell)}_k(x_\ell)u^{(\ell)}(x_\ell), \overline{v}^{(\ell)}(x_\ell)\right\rangle_{L^2(\Omega)}, \tag{5.5}$$

where δ_{ji} is the Kronecker delta.

In the general case, a basis of piecewise polynomial functions

$$\phi_{\mathbf{i}}(x) = \prod_{\ell=1}^d \phi_{i_\ell}(x_\ell), \quad \mathbf{i} \in \mathcal{I} = I^d := \{1,...,n\}^d, \tag{5.6}$$

can be used, where ϕ_{i_ℓ} are low order polynomials in the variable x_ℓ. For simplicity, we choose the Galerkin subspace $\mathbb{V} \subset (H^1_0(\Omega))^d$ of piecewise linear basis functions (cf. Section 3.3). The Galerkin approximation to the eigenvalue problem takes the form

$$\mathcal{L}U \equiv (\mathcal{A} + \mathcal{C})U = \lambda \mathcal{M}U, \quad U \in \mathbb{R}^{\mathcal{I}}, \tag{5.7}$$

with the Kronecker tensor product representation (cf. Section 3.5)

$$\mathcal{A} = \sum_{k=1}^R \sum_{i,j=1}^d \otimes_{\ell=1}^d A^{(\ell)}_{ij,k}, \quad \mathcal{C} = \sum_{k=1}^R \otimes_{\ell=1}^d C^{(\ell)}_k, \quad \mathcal{M} = \otimes_{\ell=1}^d M^{(\ell)}, \tag{5.8}$$

where $A^{(\ell)}_{ij,k}, C^{(\ell)}_k, M^{(\ell)} \in \mathbb{R}^{n\times n}$ are the tridiagonal matrices

$$A^{(\ell)}_{ij,k} = \left\{\left\langle a^{(\ell)}_{ij,k}(x_\ell)\frac{\partial^{\delta_{j\ell}}}{\partial x_\ell}\phi_p, \frac{\partial^{\delta_{i\ell}}}{\partial x_\ell}\phi_q\right\rangle_{L^2}\right\}_{p,q=1}^n,$$
$$C^{(\ell)}_k = \left\{\left\langle c^{(\ell)}_k(x_\ell)\phi_p, \phi_q\right\rangle_{L^2}\right\}_{p,q=1}^n, \quad M^{(\ell)} = \left\{\langle \phi_p, \phi_q\rangle_{L^2}\right\}_{p,q=1}^n.$$

Lemma 5.4 *The matrices $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{M}$ have the respective Kronecker ranks Rd^2, R, and 1.*

The storage requirements to represent these matrices scale linearly in the univariate problem size n,

$$Q(\mathcal{A}) = O(3d^3Rn), \quad Q(\mathcal{C}) = O(3dRn), \quad Q(\mathcal{M}) = O(3dn).$$

The same cost holds for the matrix-vector multiplication by a rank-1 vector.

Proof. The first assertion follows from the tridiagonal structure of the Kronecker factors in (5.8). Suppose that vector U has the rank-1 tensor representation

$$U = u^{(1)} \otimes ... \otimes u^{(d)}, \quad u^{(\ell)} \in \mathbb{R}^n.$$

Then the matrix-times-vector multiplication with our stiffness matrices is reduced to one-dimensional operations,

$$\mathcal{A}U = \sum_{k=1}^{R} \sum_{i,j=1}^{d} \otimes_{\ell=1}^{d} A_{ij,k}^{(\ell)} u^{(\ell)}, \quad \mathcal{C}U = \sum_{k=1}^{R} \otimes_{\ell=1}^{d} C_k^{(\ell)} u^{(\ell)}, \quad \mathcal{M}U = \otimes_{\ell=1}^{d} M^{(\ell)} u^{(\ell)},$$

which again leads to the linear cost in n. ∎

For example, in the case of the Laplace operator in $\mathbb{R}^d$ discretised by a finite difference scheme, we obtain the simple Kronecker rank-d representation

$$\mathcal{A} = A \otimes I_n \otimes ... \otimes I_n + I_n \otimes A \otimes I_n ... \otimes I_n + ... + I_n \otimes I_n ... \otimes A, \tag{5.9}$$

where $A = \text{tridiag}\{1, -2, 1\} \in \mathbb{R}^{n \times n}$ and I_n is the $n \times n$ identity matrix (cf. Section 3.5).

Now we are able to discuss iterative methods accomplished with low rank truncation for solving the discrete eigenvalue problem as in (5.7).

6 Numerics and concluding remarks

In this section we present numerical illustrations for 2D and 3D eigenvalue problems. Notice that the example for the "truncated" Lanczos algorithm applied to the 2D Laplace operator is already discussed in the introductory Section 2.

6.1 Iterative eigenvalue problem solvers with rank truncation

Without loss of generality we can set $\mathcal{M} = I$ in (5.7) since the matrix $\mathcal{M}$ has Kronecker rank one and, hence, it is diagonalisable at the expense $O(n \log n)$. Now consider the algebraic eigenvalue problem

$$\mathcal{L}U = \lambda U, \quad U \in \mathbb{R}^{\mathcal{I}}. \tag{6.1}$$

We are interested in the approximate solutions of (6.1) in the tensor class $\mathcal{M}_{\mathbf{r}} \subset \mathbb{R}^{\mathcal{I}}$.

Largest eigenvalues. A standard method for computing the eigenpair (λ, U) for the (single) largest eigenvalue is the simple power method. This includes the repeated matrix-vector multiplication which in our approach is accomplished with the rank truncation,

$$U^{(0)} \in \mathcal{M}_{\mathbf{r}}: \quad \widetilde{U}^{(m+1)} := \mathcal{L}U^{(m)}, \quad U^{(m+1)} = T_{\mathbf{r}}(\widetilde{U}^{(m+1)}),$$

where $T_{\mathbf{r}} : \mathbb{R}^{\mathcal{I}} \to \mathcal{M}_{\mathbf{r}}$ is the nonlinear truncation operator to $\mathcal{M}_{\mathbf{r}}$ defined in (3.5). $\|U^{(m+1)}\|/\|U^{(m)}\|$ will converge to the largest eigenvalue and $U^{(m)}/\|U^{(m)}\|$ to the associated eigenvector.

Lemma 6.1 *Each step of the "truncated" power iteration needs a storage of size $Q_P = O(3Rd^3rn + r^d)$. The same asymptotical complexity bound holds for the matrix-vector multiplication with rank-1 vectors.*

Smallest eigenvalues. To compute the minimal eigenvalue we apply the power method to the equivalent equation

$$\mathcal{L}^{-1}U = \frac{1}{\lambda}U, \quad U \in \mathbb{R}^{n^{\otimes d}}. \tag{6.2}$$

In the present paper we study the effect of numerical methods designed in the rank structured tensor formats described above. There are further, possibly better suited methods, which can be modified similarly by the rank truncation. In particular, one can apply preconditioned nested iterations (instead of evaluating the action of inverse matrix) as follows. Given $V^{(0)} \in \mathcal{M}_{\mathbf{r}}$:

$$\widetilde{V}^{(m+1)} := V^{(m)} - \omega\mathcal{B}\left(\mathcal{L}V^{(m)} - U\right), \quad V^{(m+1)} = T_{\mathbf{r}}(\widetilde{V}^{(m+1)}) \to \mathcal{L}^{-1}U, \tag{6.3}$$

where $\mathcal{B}$ can be chosen as the inverse of the shifted Laplacian. In the iteration (6.3) the truncation operator $T_{\mathbf{r}}$ can be applied to the intermediate vectors $\mathcal{L}V^{(m)}$ and $\mathcal{B}\left(\mathcal{L}V^{(m)} - U\right)$ as well.

6.2 Numerical illustrations

Below, we give numerical illustrations for the "truncated" power iteration applied to the inverse of an elliptic operator $\mathcal{L}^{-1}$, where we set $b = 0$. In the case of constant/separable coefficients , $\mathcal{L}^{-1}$ will be approximated by a low rank Kronecker product, e.g., obtained from the *sinc*-quadrature method [10, 11, 15].

6.2.1 Operators with constant coefficients

First we consider the finite difference analogue of the negative Laplacian $\mathcal{A}$ on the domain $\Omega = (0,1)^d$, as in (5.9) for $d = 2, 3$ with zero boundary conditions. The case $d = 2$ is of interest since in that case the truncation operator $T_{\mathbf{r}}$ can easily be realised by the *"truncated" SVD algorithm.*

We recall that the eigenvalues are

$$\lambda_{\mathbf{i}} = (n+1)^2 \sum_{\ell=1}^{d} 4\sin^2\frac{\pi i_\ell}{2(n+1)}, \quad \mathbf{i} \in \mathbb{N}^d \text{ with components } 1 \le i_\ell \le n,$$

and the corresponding eigenvectors are exact rank-1 tensors:

$$U_{\mathbf{i}} = u^{(i_1)} \otimes ... \otimes u^{(i_d)},$$

where $u^{(i_\ell)} \in \mathbb{R}^n$ with entries

$$u_s^{(i_\ell)} = \sin\frac{\pi i_\ell s}{n+1}, \quad 1 \le s \le n.$$

We implement the power iteration for the matrix $\mathcal{L} = \mathcal{A}^{-1}$, which is (approximately) represented in the rank-R Kronecker format in the form

$$\mathcal{L} \approx \mathcal{L}_R := \sum_{k=-M}^{M} c_k \bigotimes_{\ell=1}^{d} \exp(-t_k A^{(\ell)}) \approx \mathcal{A}^{-1}, \quad R = 2M+1, \quad t_k, c_k > 0, \tag{6.4}$$

with $A^{(\ell)} \in \mathbb{R}^{n\times n}$, providing exponential convergence in R (see [11]). In particular, we take

$$t_k = e^{k\mathfrak{h}}, \quad c_k = \mathfrak{h}t_k, \quad \mathfrak{h} = \pi/\sqrt{M},$$

which leads to the convergence rate

$$\left\|\mathcal{A}^{-1} - \mathcal{L}_R\right\| \le Ce^{-\pi\sqrt{M}}, \quad R = 2M+1.$$

For even better coefficients t_k, c_k compare [4]. We recall that the memory requirements for this algorithm are linear in n, i.e. $\mathcal{O}(dRn)$, while a linear complexity in the number of grid points would lead to $\mathcal{O}(n^d)$.

The matrix-vector multiplication of $\mathcal{L}_R$ with a rank-1 vector in $\mathbb{R}^{n^{\otimes d}}$ takes $\mathcal{O}(dRn\log n)$ operations by using the diagonalisation

$$\exp(-t_k A^{(\ell)}) = F_s' \cdot D \cdot F_s, \quad D = \mathrm{diag}\{e^{-t_k\lambda_1}, ..., e^{-t_k\lambda_n}\},$$

where F_s is the *sin*-transform matrix of size n, and λ_i $(i = 1, ..., n)$ are the respective eigenvalues of the 1D Laplacian.

Example 1. We present the results for $\lambda = \lambda_{min}$ in the 2D case, computed in MATLAB by using *Intel(R) T230/1.66 GHz* processor. Since the eigenvector has rank one, we apply on each iteration step the truncation operator T_1 implemented via truncated SVD. In the next table we present numerical illustrations for a sequence of grids indicating the CPU time (sec.) for one matrix-vector multiplication with $\mathcal{L}_R$, and with fixed parameter $M = 10$ (the respective Kronecker rank of $\mathcal{L}_M$ is $R = 2M + 1 = 21$), and accomplished with the rank truncation $\mathcal{L}_R$ for $R = 21$ followed by a rank truncation. We present the number of power iterations on each grid, as well as the resulting relative errors $\delta_\lambda = |\lambda - \lambda_{h,R}|/|\lambda|$ and $\delta_u = \|U - U_{h,R}\|_1/\|U\|_1$.

p $(n = 2^p - 1)$	Time/it.	δ_λ	δ_u	it.
8	0.04	$7.3 \cdot 10^{-6}$	$5.0 \cdot 10^{-4}$	5
10	0.1	$2.5 \cdot 10^{-7}$	$1.1 \cdot 10^{-4}$	6
12	0.4	$1.0 \cdot 10^{-8}$	$2.0 \cdot 10^{-5}$	7
14	1.45	$3.8 \cdot 10^{-10}$	$4.2 \cdot 10^{-6}$	8
16	7.0	$2.2 \cdot 10^{-11}$	$1.17 \cdot 10^{-6}$	9

Table 6.1: Minimal eigenvalue for 2D Laplacian.

The results indicate the true asymptotical convergence of the truncated iteration in the mesh parameter $h = \pi/(n + 1)$, on a sequence of large $n \times n$ grids for $n = 2^p - 1$, $p = 8, 10, ..., 16$. Table 6.2.1 also indicates perfect linear scaling of the tensor method in n (compare with $\mathcal{O}(n^2)$ for traditional methods of linear complexity). The number of power iterations increases as $O(\log n)$ as expected.

Example 2. We consider the tensor computation of several smallest eigenvalues using the same tensor power iteration, but accomplished with successive Gram-Schmidt orthogonalisation at each iteration step. Here we solve the problem on a sequence of $n \times n$ grids for $n = 2^p - 1$, $p = 6, 8, ..., 16$. The number of power iterations observed is about $2 \le it \le 4$ for all grids with $p \ge 8$. Here the computation of the initial guess by means of a nested iteration involving coarser grids is very helpful. The next table presents the iteration history for three eigenpairs (λ_i, u_i) $(i = 1, 2, 3)$.

p	Time/it.	δ_{λ_1}	δ_{u_1}	δ_{λ_2}	δ_{u_2}	δ_{λ_3}	δ_{u_3}
6	0.02	$2.3 \cdot 10^{-6}$	$2.5 \cdot 10^{-4}$	$2.7 \cdot 10^{-5}$	$1.9 \cdot 10^{-3}$	$4.2 \cdot 10^{-6}$	$3.3 \cdot 10^{-3}$
8	0.03	$2.1 \cdot 10^{-7}$	$8.9 \cdot 10^{-5}$	$1.3 \cdot 10^{-6}$	$3.9 \cdot 10^{-4}$	$5.3 \cdot 10^{-7}$	$1.1 \cdot 10^{-3}$
10	0.1	$1.7 \cdot 10^{-8}$	$2.2 \cdot 10^{-5}$	$6.9 \cdot 10^{-8}$	$9.4 \cdot 10^{-5}$	$1.3 \cdot 10^{-8}$	$1.8 \cdot 10^{-4}$
12	0.39	$9.9 \cdot 10^{-10}$	$5.2 \cdot 10^{-6}$	$4.5 \cdot 10^{-9}$	$2.3 \cdot 10^{-5}$	$1.9 \cdot 10^{-9}$	$7.0 \cdot 10^{-5}$
14	1.6	$6.5 \cdot 10^{-11}$	$1.3 \cdot 10^{-6}$	$2.8 \cdot 10^{-10}$	$6.0 \cdot 10^{-6}$	$1.5 \cdot 10^{-10}$	$1.7 \cdot 10^{-5}$
16	6.9	$3.8 \cdot 10^{-12}$	$3.2 \cdot 10^{-7}$	$1.8 \cdot 10^{-11}$	$1.5 \cdot 10^{-6}$	$4.3 \cdot 10^{-11}$	$4.5 \cdot 10^{-6}$

Table 6.2: Several smallest eigenvalues for 2D Laplacian.

Again, these results indicate the linear scaling of the tensor method in n. The cost of one power iteration step increases like $O(n \log n)$ as expected. Furthermore, we expect the asymptotic behaviour

$$\delta_\lambda = O(\lambda h^2) \quad \text{and} \quad \delta_u = O(\sqrt{\lambda} h), \quad \text{as} \quad h \to 0,$$

which are in agreement with the above presented calculations. Notice that in our case the true scaling factor between the refined grids is 16 for the eigenvalues and 4 for the eigenfunctions.

Example 3. We apply the tensor method to compute the minimal eigenvalue of the 3D Laplacian on large $n \times n \times n$ grids with $n = 2^p - 1$. In this case, the action of the truncation operator T_1 is equivalent to the rank-1 Tucker approximation of the $n \times n \times n$ rank-R tensors arising at each iterative step.

We observe the asymptotic complexity $O(dn \log n)$ to achieve the theoretical error bounds for both the eigenvalues and eigenfunctions. Notice that the problem size on the finest grid (with $n = 2^{17}$, $d = 3$) exceeds $N = n^d = 2^{17d} \approx 10^{15}$, which is far beyond the facilities of modern super-computers. Hence tensor methods are mandatory for solving large scale multi-dimensional spectral problems.

p	Time/it.	δ_λ	δ_u	it.
6	0.03	$2.0 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$	4
8	0.05	$1.6 \cdot 10^{-5}$	$9.4 \cdot 10^{-4}$	4
10	0.12	$7.8 \cdot 10^{-7}$	$1.2 \cdot 10^{-4}$	5
12	0.51	$4.9 \cdot 10^{-8}$	$3.4 \cdot 10^{-5}$	5
14	2.2	$3.1 \cdot 10^{-9}$	$9.3 \cdot 10^{-6}$	5
16	10.6	$1.9 \cdot 10^{-10}$	$2.8 \cdot 10^{-6}$	5
17	22.3	$4.8 \cdot 10^{-11}$	$1.6 \cdot 10^{-6}$	5

Table 6.3: Minimal eigenvalue for 3D Laplacian on large spatial grids.

6.2.2 Operator with variable coefficients

In this section we consider eigenvalue problems for the elliptic operators of the form

$$\Lambda u = \Delta u + c(x)u, \quad x \in \mathbb{R}^d$$

with smooth coefficient $c(x)$.

In the subsequent examples we consider discretisations represented by matrices of the form

$$\mathcal{A} = \mathcal{M} + q\mathcal{C}, \quad q > 0, \tag{6.5}$$

where $\mathcal{M} > 0$ is the finite difference negative Laplacian in 2D as before, and $\mathcal{C}$ is the low Kronecker rank matrix discretising the zero-order term $c(x) = \sum_{k=1}^{R} c_k^{(1)}(x_1) \cdot c_k^{(2)}(x_2) \geq 0$ as in (5.3). We suppose that

$$0 \leq \langle \mathcal{C}x, x \rangle \leq \langle x, x \rangle \quad \forall\, x \in \mathbb{V}_n \setminus \{0\}.$$

Example 4. In this example we apply the Lanczos algorithm for computing the *largest eigenvalues* as described in §2. Consider matrices of the form

$$\mathcal{A} = \mathcal{M} - \sum_{t=1}^{R} D_t \otimes D_t,$$

where $\mathcal{M}$ is the negative discrete Laplacian and D_t are diagonal matrices with positive entries. We approximate the maximal eigenvalue by the standard Lanczos and truncated tensor Lanczos methods for the following two examples:

(A) the entries of D_t are grid values of the function $(1 + T_t(x))/10$, where T_t is the Chebyshev polynomial of degree t;

(B) the entries of D_t are random values uniformly distributed in $[0, 1]$.

We compare the results obtained after 50 iterations for both methods. The matrix size is $N = 300^2$, the truncation rank and the accuracy are set to 10 and 10^{-2}, respectively.

R	1	3	5	7	9
Standard Lanczos	7.989	7.957	7.925	7.900	7.893
Tensor Lanczos	7.977	7.940	7.917	7.893	7.906

Table 6.4: Maximal eigenvalues, case (A).

We observe that the computed eigenvalues in both methods are close to the truncation accuracy even for the random case. This suggests that tensor tools may be applied to a much broader class of matrices than required by the theory described above.

R	1	3	5	7	9
Standard Lanczos	7.862	7.615	7.302	6.800	6.460
Tensor Lanczos	7.852	7.608	7.292	6.789	6.452

Table 6.5: Maximal eigenvalues, case (B).

Example 5. The *minimal eigenvalue* is calculated for the problem (6.5) with the matrix $\mathcal{C}$ corresponding to the rank-1 potential $c(x) = \sin(\lambda x_1)\sin(\lambda x_2)$ in $(0,1)^2$. We perform the truncated power iteration with the matrix $\mathcal{L} = \mathcal{A}^{-1}$, such that the respective matrix-vector multiplication $y = \mathcal{L}U$ is implemented by the "truncated" iterative procedure, $y^p \to y$, as $p \to \infty$, where

$$y^{p+1} = y^p - \omega \mathcal{L}_R(\mathcal{A}y^p - U), \quad p = 0, 1, ...$$

with certain rank-r initial guess y^0 and with a proper relaxation parameter $\omega \in (0, 2)$. Here $\mathcal{L}_R$ is the rank-R approximation to the inverse of the shifted Laplacian described in (6.4).

Table 6.6 presents the results on the sequence of grids $n = 2^p$, $p = 10, 11, 12, 13$, for the truncation rank $r = 3$. We give the total CPU time (sec.), the number of power iterations and the scaling factor between the neighbouring grids. We expect an $O(n \log n)$ scaling provided that all nested iterations require the same number of loops and the same tensor rank for the preconditioner (of course, there are some fluctuations).

n	Time	δ_{Res}	Power iter.	Scaling
1024	7.3	$3.0 \cdot 10^{-4}$	10	–
2048	23.6	$1.5 \cdot 10^{-4}$	14	2.3
4096	63.9	$7.6 \cdot 10^{-5}$	14	2.7
8192	209.	$3.8 \cdot 10^{-5}$	17	2.7

Table 6.6: Minimal eigenvalue for $-\Delta + c(x)$ in 2D.

This table indicates the linear-logarithmic scaling in n as well as the robust convergence of the power iteration with the tensor modification.

6.3 Concluding remarks

The theoretical and numerical analysis of multi-dimensional eigenvalue problems presented in the paper clearly indicate that tensor structured methods for the approximation and solution of "smooth" spectral problems in $\mathbb{R}^d$ yield a promising basis for efficient solution methods in the modern high dimensional applications. Moreover, it seem that these methods are not restricted to smooth problems.

Acknowledgement. The authors are thankful to Dipl. Ing. Cristóbal Bertoglio for assistence with the numerical computations of Tables 6.1, 6.2 and Example 5.

References

[1] S.N. Bernstein: *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*. Paris: Gauthier-Villars 1926

[2] G. Beylkin and M. M. Mohlenkamp: *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, **99** (2002), 10246–10251.

[3] G. Beylkin and M. M. Mohlenkamp: *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., **26** (2005), 2133–2159.

[4] D. Braess and W. Hackbusch: *Approximation of* $1/x$ *by exponential sums in* $[1, \infty)$. IMA J. Numer. Anal., **25** (2005) 685–697.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-*1 *and rank-*$(R_1, ..., R_N)$ *approximation of higher-order tensors.* SIAM J. Matrix Anal. Appl., **21** (2000) 1324-1342.

[6] L. De Lathauwer, B. de Moor, and J. Vandewalle: *A multilinear singular value decomposition.* SIAM J. Matrix Anal. Appl. **21** (2000), 1253–1278.

[7] M. Espig: *Approximation mit Elementartensorsummen.* Doctoral thesis, Universität Leipzig, 2008.

[8] H.-J. Flad, B. Khoromskij, D. Savostianov, and E. Tyrtyshnikov: *Verification of the cross 3d algorithm on quantum chemistry data.* Rus. J. Numer. Anal. and Math. Modelling, **4** (2008), 1–16.

[9] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Data-sparse approximation to operator-valued functions of elliptic operators.* Math. Comp. **73**, no. 247 (2003), 1297–1324.

[10] I.P. Gavrilyuk, W. Hackbusch, and B.N. Khoromskij: *Tensor-product approximation to elliptic and parabolic solution operators in higher dimensions.* Computing **74** (2005), 131-157.

[11] 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** (2006), 177-202.

[12] W. Hackbusch, B.N. Khoromskij, and E.E. Tyrtyshnikov: *Hierarchical Kronecker tensor-product approximations.* J. Numer. Math. **13** (2005), 119–156.

[13] W. Hackbusch, B.N. Khoromskij, and E.E. Tyrtyshnikov: *Approximate iterations for structured matrices.* Numer. Math. **109** (2008), 365-383.

[14] B. N. Khoromskij: *Structured Rank-*$(r_1, ..., r_d)$ *Decomposition of Function-related Tensors in* $\mathbb{R}^d$. Comp. Meth. in Applied Math. **6** (2006), 194-220.

[15] B. N. Khoromskij: *On tensor approximation of Green iterations for Kohn-Sham equations.* Computing and Visualization in Scince, DOI:10.1007/s00791-008-0097-x, 2008.

[16] B. N. Khoromskij: *Multi-linear approximation of higher-dimensional convolution in linear cost.* Preprint 36, Max-Planck Institute for Mathematics in Sciences, Leipzig, 2008 (submitted).

[17] B. N. Khoromskij and V. Khoromskaia: *Low rank Tucker tensor approximation to the classical potentials.* Central European J. of Math., **5**(3) 2007, 1-28.

[18] B. N. Khoromskij and V. Khoromskaia: *Multigrid accelerated tensor approximation of function related multi-dimensional arrays.* Preprint 40, Max-Planck Institute for Mathematics in Sciences, Leipzig, 2008 (submitted).

[19] T. G. Kolda and B. W. Bader: *Tensor decompositions and applications.* Technical Report No SAND2007-6702, Sandia National Laboratories, Albuquerque, NM and Livermore, CA, November 2007.

[20] J.M. Melenk: *hp-finite element methods for singular perturbations.* Springer Lecture Notes in Mathematics, **1796**, (2002).

[21] M. Melenk and S. Sauter: *Convergence analysis for finite element discretizations of the Helmholtz equation. Part I: The full space problem.* 09-2008, Universität Zürich 2008.

[22] I.V. Oseledets, D.V. Savostyanov, and E.E. Tyrtyshnikov: *Tucker dimensionality reduction of three-dimensional arrays in linear time.* SIMAX, vol. 30, no. 3, pp. 939-956 (2008).

[23] E.E. Tyrtyshnikov: *Tensor approximations of matrices generated by asymptotically smooth functions.* Sbornik: Mathematics **194**, No. 5-6 (2003), 941–954 (translated from *Mat. Sb.* **194**, No. 6 (2003), 146–160).

[24] E.E. Tyrtyshnikov: *Kronecker-product approximations for some function-related matrices.* Linear Algebra Appl. **379** (2004), 423–437.

[25] S. Sauter and C. Schwab: *Randelementmethoden*. B.G. Teubner, 2004.

[26] C. Schwab: *p- and hp-Finite Element Methods*. Oxford University Press, 1998.

[27] A. Smilde, R. Bro, and P.Geladi: *Multiway analysis*. Wiley, 2004.

[28] E. Tadmor: *The exponential accuracy of Fourier and Chebychev differencing methods*. SIAM J. Numer. Anal. **23** (1986), 1-23.

[29] V. Thomée: *Galerkin finite element methods for parabolic problems*. Springer, Berlin, 1997.

[30] R.-A. Todor, and Ch. Schwab: *Convergence rate for sparse approximation of elliptic problems with stochastic coefficients*. IMA J. of Numer. Anal. **27** (2007), 232-261.

[31] L.R. Tucker: *Some mathematical notes on three-mode factor analysis*. Psychometrika **31** (1966) 279-311.