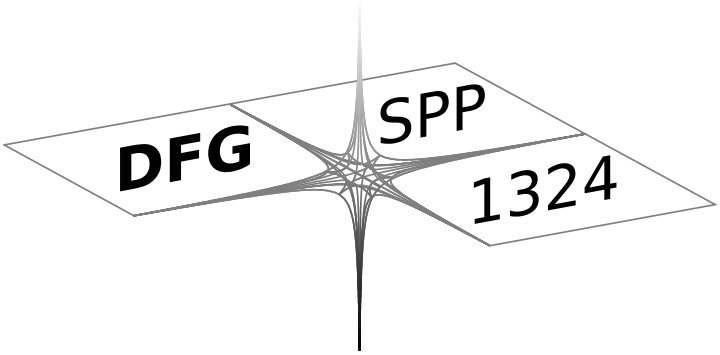
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

On the Low-rank Approximation by the Pivoted Cholesky Decomposition

H. Harbrecht, M. Peters, R. Schneider

Preprint 76



Edited by

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

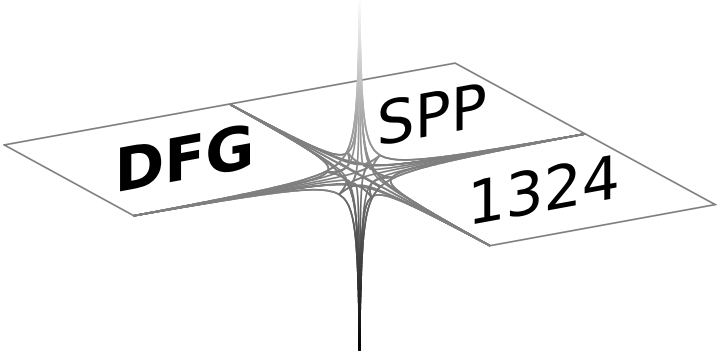
DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

On the Low-rank Approximation by the Pivoted Cholesky Decomposition

H. Harbrecht, M. Peters, R. Schneider

Preprint 76



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

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

ON THE LOW-RANK APPROXIMATION BY THE PIVOTED CHOLESKY DECOMPOSITION

HELMUT HARBRECHT, MICHAEL PETERS, AND REINHOLD SCHNEIDER

Dedicated to the victims of the earthquake 2010 in Chile

Abstract. The present paper is dedicated to the application of the pivoted Cholesky decomposition to compute low-rank approximations of dense, positive semi-definite matrices. The resulting approximation error is rigorously controlled in terms of the trace norm. Exponential convergence rates are proved under the assumption that the eigenvalues of the matrix under consideration exhibit a sufficiently fast exponential decay. By numerical experiments it is demonstrated that the pivoted Cholesky decomposition leads to very efficient algorithms to separate the variables of bi-variate functions.

1. Introduction

Many problems in science and engineering lead to hugh, densely populated matrices which are symmetric and positive semi-definite. Often they arise from the discretization of *symmetric and positive semi-definite functions*, especially in the context of partial differential equations.

Consider a domain $\Omega \subset \mathbb{R}^n$ and a symmetric bi-variate function $f \in L^2(\Omega \times \Omega)$. Then, we call the function f positive semi-definite if the corresponding (symmetric) Hilbert-Schmidt operator

$$\mathcal{K} : L^2(\Omega) \to L^2(\Omega), \quad (\mathcal{K}u)(\mathbf{x}) = \int_\Omega f(\mathbf{x},\mathbf{y})u(\mathbf{y})\,\mathrm{d}\mathbf{y}$$

is positive semi-definite

$$\int_\Omega \int_\Omega f(\mathbf{x},\mathbf{y})u(\mathbf{x})u(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} = (\mathcal{K}u,u)_{L^2(\Omega)} \geq 0 \quad \text{for all} \quad u \in L^2(\Omega).$$

Symmetric and positive semi-definite functions appear in many applications. For example, the second statistical moment of a stochastic function, the *two-point correlation*, falls into this class of functions. Furthermore, to compute the singular value decomposition of a function $g \in L^2(\Omega_1 \times \Omega_2)$, also known as the *Karhunen-Loève expansion*, one has to compute the eigenvalues of the Hilbert-Schmidt operator with

the symmetric and positive semi-definite kernel function

$$f(\mathbf{x},\mathbf{y}) = \int_{\Omega_1} g(\mathbf{x},\mathbf{z}) g(\mathbf{y},\mathbf{z})\,\mathrm{d}\mathbf{z}.$$

Further applications arise from quantum chemistry [4, 14, 17] or in inverse problems [13].

When spending n degrees of freedom to approximate functions from $L^2(\Omega)$, an approximation of $f \in L^2(\Omega\times\Omega)$ yields a symmetric and positive semi-definite matrix of size $n\times n$. Since n is typically large, this causes serious obstructions when dealing numerically with such problems. Therefore, we intend here to compute a *symmetric* low-rank approximation to f, namely

$$f_m(\mathbf{x},\mathbf{y}) := \sum_{i=1}^{m} \psi_i(\mathbf{x})\psi_i(\mathbf{y}) \quad \text{such that} \quad \|f - f_m\|_{L^2(\Omega\times\Omega)} \leq \varepsilon.$$

The best approximation in $L^2(\Omega\times\Omega)$ is of course the expansion into the largest eigenpairs (λ_i,φ_i) of the related Hilbert-Schmidt operator. However, the eigenpair computation is quite expensive as it needs the access to the full matrix.

An alternative approach for computing a symmetric low-rank approximation is provided by the *pivoted Cholesky decomposition*. This approach is quite similar to the *adaptive cross* or *mosaic-skeleton approximation* [1, 2, 6, 7, 24]. Nevertheless, there are important advantages in comparison with standard adaptive cross approximation.

(1) We are able to rigorously control the approximation error in terms of the trace norm.
(2) The pivoting strategy is exact, i.e., we need to search for the pivot element only on the main diagonal of the matrix which is easily accessible.
(3) In case of a sufficiently fast exponential decay of the eigenvalues of the matrix $\mathbf{A}$ we can proof that the pivoted Cholesky decomposition converges exponentially. For given $\varepsilon > 0$ it computes a rank-m approximation $\mathbf{A}_m$ such that $\mathrm{trace}(\mathbf{A} - \mathbf{A}_m) \leq \varepsilon$ and m being proportional to $|\log(\varepsilon/n)|$.
(4) The algorithm is extremely easy to implement, having only $\mathcal{O}(m^2 n)$ complexity.

An exponential decay of the eigenvalues is for example obtained in case of analytical functions and $\Omega \subset \mathbb{R}$ [23]. Nevertheless, our proof is purely algebraically. We do not need the function's smoothness in difference to the convergence analysis of adaptive cross approximation [1].

Numerical experiments (see Sect. 4) even indicate that the pivoted Cholesky decomposition always converges *optimal* in the sense of the rank m being proportional to

the number of terms required for the singular value decomposition. This is observed in particular in the case of functions of finite smoothness where it is known that the eigenvalues decay only algebraically.

We emphasize that the pivoted Cholesky decomposition is a well established algorithm in numerical linear algebra, see [8, 11, 12] and the reference therein. It is implemented for instance in the LINPACK library [5], particularly to factorize positive semi-definite matrices. However, except for publications in quantum chemistry where the method is quite popular, we are not aware of a paper in which it has been used in the context of low-lank approximations. It is a purely algebraically black-box method to compute low-lank approximations of matrices, where the approximation error is controlled in the trace norm. Since it is a *symmetric* low-rank decomposition, it can in particular be used for the fast computation of the largest eigenpairs (cf. Section 4), as required for the Karhunen-Loève expansion.

The rest of the paper is organizes as follows. In Section 2 we survey on the algorithmic details of the Cholesky decomposition for positive semi-definite matrices. Section 3 motivates the use of the pivoted Cholesky decomposition to calculate low-rank approximations of matrices. As our main result we present the purely algebraic convergence proof (Thm. 3.2). Section 4 is devoted to numerical applications arising from the context of partial differential equations. We compute the variance of an elliptic second order boundary value problem with stochastic right hand side. Then, we consider the fast eigenpair computation of symmetric and positive semi-definite Hilbert-Schmidt operators. Finally, we discuss the application of the pivoted Cholesky decomposition for the fast computation of the two-electron integral matrix in quantum chemistry.

To avoid the repeated use of generic but unspecified constants, we denote throughout this paper by $C \lesssim D$ that C is bounded by a multiple of D independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \sim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. Cholesky decomposition for positive semi-definite matrices

Let $\mathbf{0} \neq \mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric and positive semi-definite matrix with eigenvalues $\lambda_i \geq 0$, $i = 1, 2, \ldots, n$. Due to the identity $\operatorname{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i > 0$ there exists at least one positive diagonal entry $a > 0$. We assume without loss of generality that it is located at the $(1,1)$-position. Otherwise we find an index $1 < \ell \leq n$ such that the ℓ-th diagonal entry is positive and swap the first and the ℓ-th column and row. This might be expressed in terms of a symmetric permutation matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ satisfying $\mathbf{P}^2 = \mathbf{I}$. The Schur complement relative to the $(1,1)$-entry is then again a

symmetric and positive semi-definite matrix. For sake of clearness in representation we shall prove this statement.

Lemma 2.1. *Let the matrix*

$$\mathbf{A} = \left[\begin{array}{c|c} a & \mathbf{b}^T \\ \hline \mathbf{b} & \mathbf{C} \end{array}\right] \in \mathbb{R}^{n\times n}$$

be symmetric and positive semi-definite with $a > 0$. Then, the Schur complement

$$\mathbf{S} := \mathbf{C} - \frac{1}{a}\mathbf{b}\mathbf{b}^T \in \mathbb{R}^{(n-1)\times(n-1)} \tag{2.1}$$

is well-defined and also symmetric and positive semi-definite.

Proof. Since $a > 0$ the Schur complement is well-defined and also symmetric due to

$$\mathbf{S}^T = \mathbf{C}^T - \frac{1}{a}\big(\mathbf{b}\mathbf{b}^T\big)^T = \mathbf{C} - \frac{1}{a}\mathbf{b}\mathbf{b}^T = \mathbf{S}.$$

Consider $\mathbf{y} \in \mathbb{R}^{n-1}$ and set $x := -\mathbf{b}^T\mathbf{y}/a \in \mathbb{R}$. Then, it follows

$$0 \le \begin{bmatrix} x \\ \mathbf{y} \end{bmatrix}^T \mathbf{A} \begin{bmatrix} x \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} x \\ \mathbf{y} \end{bmatrix} \begin{bmatrix} ax + \mathbf{b}^T\mathbf{y} \\ x\mathbf{b} + \mathbf{C}\mathbf{y} \end{bmatrix} = \begin{bmatrix} x \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} 0 \\ \mathbf{S}\mathbf{y} \end{bmatrix} = \mathbf{y}^T\mathbf{S}\mathbf{y},$$

i.e., $\mathbf{S}$ is also positive semi-definite. □

As a consequence of this lemma, we can decompose the matrix $\mathbf{A}$ according to

$$\mathbf{P}\mathbf{A}\mathbf{P} = \frac{1}{a}\left[\begin{array}{c} a \\ \hline \mathbf{b} \end{array}\right]\left[\begin{array}{c} a \\ \hline \mathbf{b} \end{array}\right]^T + \left[\begin{array}{c|c} 0 & \mathbf{0}^T \\ \hline \mathbf{0} & \mathbf{S} \end{array}\right]. \tag{2.2}$$

Now, if $\mathbf{S} \neq \mathbf{0}$, we can repeat this procedure for $\mathbf{S}$ and obtain a decomposition

$$\mathbf{P}_2\mathbf{P}_1\mathbf{A}\mathbf{P}_1\mathbf{P}_2 = \frac{1}{a_1}\mathbf{P}_2\left[\begin{array}{c} a_1 \\ \hline \mathbf{b}_1 \end{array}\right]\left[\begin{array}{c} a_1 \\ \hline \mathbf{b}_1 \end{array}\right]^T\mathbf{P}_2 + \frac{1}{a_2}\left[\begin{array}{c} 0 \\ \hline a_2 \\ \hline \mathbf{b}_2 \end{array}\right]\left[\begin{array}{c} 0 \\ \hline a_2 \\ \hline \mathbf{b}_2 \end{array}\right]^T + \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & \mathbf{S}_2 \end{array}\right],$$

or generally

$$\mathbf{P}_j\cdots\mathbf{P}_2\mathbf{P}_1\mathbf{A}\mathbf{P}_1\mathbf{P}_2\cdots\mathbf{P}_j = \sum_{i=1}^{j}\widehat{\boldsymbol{\ell}}_i\widehat{\boldsymbol{\ell}}_i^T + \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}_j \end{array}\right] \tag{2.3}$$

with

$$\widehat{\boldsymbol{\ell}}_i := \frac{1}{\sqrt{a_i}}\mathbf{P}_j\mathbf{P}_{j-1}\cdots\mathbf{P}_{i+1}\left.\left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline a_i \\ \hline \mathbf{b}_i \end{array}\right]\right\} i-1 \text{ times} \tag{2.4}$$

and

$$\mathbf{S}_i := \mathbf{S}_{i-1} - \frac{1}{a_i}\mathbf{b}_i\mathbf{b}_i^T \tag{2.5}$$

for all $i = 1, 2, \ldots, j$. To obtain the final representation of the matrix $\mathbf{A}$ we multiply by the permutation matrices from both sides and get

$$\mathbf{A} = \sum_{i=1}^{j} \boldsymbol{\ell}_i\boldsymbol{\ell}_i^T + \mathbf{E}_j \text{ with } \boldsymbol{\ell}_i := \frac{1}{\sqrt{a_i}}\mathbf{P}_1\mathbf{P}_2\cdots\mathbf{P}_i \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline a_i \\ \hline \mathbf{b}_i \end{bmatrix} \text{ for all } i = 1, 2, \ldots, j.$$

Here, the remainder matrix $\mathbf{E}_j$ is given by

$$\mathbf{E}_j := \mathbf{P}_1\mathbf{P}_2\cdots\mathbf{P}_j \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{S}_j \end{array}\right] \mathbf{P}_j\cdots\mathbf{P}_2\mathbf{P}_1.$$

By introducing the matrix $\mathbf{L}_j = [\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots, \boldsymbol{\ell}_j]$, we can rewrite the decomposition in accordance with $\mathbf{A} = \mathbf{L}_j\mathbf{L}_j^T + \mathbf{E}_j$.

If it holds $\operatorname{rank}\mathbf{A} = k < n$, then the above algorithm terminates after k steps with $\mathbf{E}_k = \mathbf{0}$ since the matrix $\mathbf{A}_k := \mathbf{L}_k\mathbf{L}_k^T$ has obviously rank k due to $a_1, a_2, \ldots, a_k > 0$. If the algorithm completes without termination ahead of time and the $\mathbf{P}_i$ are always chosen as the identity, it produces the standard Cholesky decomposition.

3. Low-rank approximation

As we have seen, if $\operatorname{rank}(\mathbf{A}) = m < n$, the above pivoted Cholesky decomposition will terminate after m steps with $\mathbf{E}_m = \mathbf{0}$. Thus, the following question arises: assume that a rank-m matrix $\mathbf{A}_m \in \mathbb{R}^{n\times n}$ exists such that

$$\|\mathbf{A} - \mathbf{A}_m\| \leq \varepsilon.$$

Can the Cholesky decomposition be used to find this approximation? In other words, does the computation of $m < n$ terms of (2.3)–(2.5) produce a rank-m approximation $\mathbf{A}_m = \mathbf{L}_m\mathbf{L}_m^T$ of $\mathbf{A}$ with $\|\mathbf{A} - \mathbf{A}_m\| \lesssim \varepsilon$?

To answer the above question we choose the trace norm as error measure. Then, the best possible reduction of the error in (2.2) is achieved if the trace norm of the Schur complement becomes small. This amounts to the problem

$$\operatorname{trace}\mathbf{S} = \operatorname{trace}\mathbf{A} - \frac{1}{a_{i,i}}\|\mathbf{a}_i\|_2^2 \to \min_{i=1}^{n}$$

where $\mathbf{a}_i$ denotes the i-th column vector of $\mathbf{A}$ and $a_{i,i}$ is the i-th diagonal entry.

Since the above minimization problem can only be solved when the complete matrix $\mathbf{A}$ is known, we shall use another strategy. Namely, we like to eliminate the value $a_{i,j}$ of largest modulus. By choosing $\mathbf{x} \in \mathbb{R}^n$ such that $x_i = \sqrt{a_{j,j}/a_{i,i}}$, $x_j = \pm\sqrt{a_{i,i}/a_{j,j}}$ and $x_k = 0$ otherwise, it follows $0 \leq \mathbf{x}^T\mathbf{A}\mathbf{x} = 2\big(\sqrt{a_{i,i}a_{j,j}} - |a_{i,j}|\big)$ and hence

$$|a_{i,j}| \leq \sqrt{a_{i,i}a_{j,j}} \leq \frac{a_{i,i} + a_{j,j}}{2} \leq \max_{i=1}^{n} a_{i,i}, \quad j = 1, 2, \ldots, n.$$

Consequently, the largest value in modulus lies on the diagonal of the matrix $\mathbf{A}$. Our strategy will thus be to choose the largest diagonal entry as pivot element. This is quite similar to the totally pivoted adaptive cross approximation [1, 2].

We emphasize that, due to (2.1), the series of pivot elements is strictly decreasing until the Schur complement vanishes or the algorithm is stopped. Putting all the above components together, we arrive at the following algorithm:

Algorithm 1: Pivoted Cholesky decomposition

Data: matrix $\mathbf{A} = [a_{i,j}] \in \mathbb{R}^{n\times n}$ and error tolerance $\varepsilon > 0$
Result: low-rank approximation $\mathbf{A}_m = \sum_{i=1}^{m} \boldsymbol{\ell}_i \boldsymbol{\ell}_i^T$ such that $\operatorname{trace}(\mathbf{A} - \mathbf{A}_m) \leq \varepsilon$
begin
  set $m := 1$;
  set $\mathbf{d} := \operatorname{diag}(\mathbf{A})$ and $error := \|\mathbf{d}\|_1$;
  initialize $\boldsymbol{\pi} := (1, 2, \ldots, n)$;
  while $error > \varepsilon$ **do**
    set $i := \arg\max\{d_{\pi_j} : j = m, m+1, \ldots, n\}$;
    swap π_m and π_i;
    set $\ell_{m,\pi_m} := \sqrt{d_{\pi_m}}$;
    for $m + 1 \leq i \leq n$ **do**
      compute $\ell_{m,\pi_i} := \left(a_{\pi_m,\pi_i} - \sum_{j=1}^{m-1} \ell_{j,\pi_m}\ell_{j,\pi_i}\right) \Big/ \ell_{m,\pi_m}$;
      update $d_{\pi_i} := d_{\pi_i} - \ell_{m,\pi_m}\ell_{m,\pi_i}$;
    compute $error := \sum_{i=m+1}^{n} d_{\pi_i}$;
    increase $m := m + 1$;
end

Notice that *only* all diagonal entries of the matrix $\mathbf{A}$ and the m rows associated with the pivot elements need to be evaluated to compute the rank-m approximation. All other matrix coefficients do not enter the computation. This makes the method highly attractive for the sparse approximation of smooth nonlocal operators (see Thm. 3.2). For operators with kernel functions that exhibit a singularity on the

diagonal $\mathbf{x} = \mathbf{y}$ it might be better to introduce a suitable partitioning of the matrix which leads to the original adaptive cross approximation as introduced in [1, 2].

Theorem 3.1. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ symmetric and positive semi-definite. Then, performing m steps of the pivoted Cholesky decomposition is of complexity $\mathcal{O}(m^2 n)$.*

Proof. The most expensive part in Algorithm 1 is the computation of the Cholesky vectors $\boldsymbol{\ell}_k$, $k = 1, 2, \ldots, m$. This requires

$$\sum_{k=1}^{m} \sum_{i=k+1}^{n} \sum_{j=1}^{k-1} 1 \leq \sum_{k=1}^{m} (k-1)n \leq \frac{m^2}{2} n$$

additions and multiplications each which proves the assertion. □

In case of sufficiently fast exponentially decaying eigenvalues we can proof that the pivoted Cholesky decomposition computes a rank-m approximation which exponentially approximates the matrix $\mathbf{A}$. For example, according to [23], the eigenvalues decay exponentially if the underlying function f is analytical on $\Omega \times \Omega \subset \mathbb{R}^2$. It even suffices to have piecewise analyticity in the sense of the smooth parts being the product domains $\Omega_i \times \Omega_j$ where $\overline{\Omega} = \bigcup_{i=1}^{k} \overline{\Omega}_i$.

Theorem 3.2. *Assume that the eigenvalues of $\mathbf{A} \in \mathbb{R}^{n\times n}$ satisfy*

$$4^m \lambda_m \lesssim \exp(-bm)$$

for some $b > 0$ uniformly in n. Then, the pivoted Cholesky approximation $\mathbf{A}_m$ with rank $m \sim |\log(\varepsilon/n)|$ satisfies $\operatorname{trace}(\mathbf{A}_m - \mathbf{A}) \lesssim \varepsilon$ uniformly as $\varepsilon > 0$ tends to zero.

Proof. Without loss of generality we assume that $\mathbf{A}$ is permuted in such a way that the k-th pivot is found at the (k, k)-position for all $k = 1, 2, \ldots, n$. Then, $\mathbf{L}_m \in \mathbb{R}^{n\times m}$ is always lower triangular matrix. We partition the matrices $\mathbf{A}$ and $\mathbf{L}_m$ according to

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} \end{bmatrix}, \qquad \mathbf{L}_m = \begin{bmatrix} \mathbf{L}_{1,1} & \mathbf{0} \\ \mathbf{L}_{2,1} & \mathbf{0} \end{bmatrix}.$$

From

$$\mathbf{A}_m = \mathbf{L}_m \mathbf{L}_m^T = \begin{bmatrix} \mathbf{L}_{1,1}\mathbf{L}_{1,1}^T & \mathbf{L}_{1,1}\mathbf{L}_{2,1}^T \\ \mathbf{L}_{2,1}\mathbf{L}_{1,1}^T & \mathbf{L}_{1,2}\mathbf{L}_{2,1}^T \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1} & \mathbf{L}_{1,2}\mathbf{L}_{2,1}^T \end{bmatrix}$$

one readily infers that $\mathbf{L}_{1,1}\mathbf{L}_{1,1}^T$ is the (pivoted) Cholesky decomposition of $\mathbf{A}_{1,1}$. Consequently, denoting the m-th largest eigenvalue of a matrix $\mathbf{B}$ by $\lambda_m(\mathbf{B})$, we have

$$\frac{1}{\lambda_m(\mathbf{A}_{1,1})} = \left\|\mathbf{A}_{1,1}^{-1}\right\|_2 = \left\|\mathbf{L}_{1,1}^{-1}\right\|_2^2.$$

In accordance with [10] it holds for the pivoted Cholesky decomposition the (sharp) worst case estimate $\left\|\mathbf{L}_{1,1}^{-1}\right\|_2 \leq \sqrt{4^m + 6m - 1}/(3\ell_{m,m})$. Hence, we arrive at

$$\frac{1}{\lambda_m(\mathbf{A}_{1,1})} \leq \frac{4^m + 6m - 1}{9\ell_{m,m}^2} \leq \frac{4^m}{\ell_{m,m}^2}. \tag{3.6}$$

Since the trace error of $\mathbf{A} - \mathbf{A}_m$ is bounded by $(n-m)$-times the pivot element $\ell_{m,m}^2$, it follows immediately that

$$\operatorname{trace}(\mathbf{A} - \mathbf{A}_m) \leq 4^m(n-m)\lambda_m(\mathbf{A}_{1,1}) \leq 4^m n \lambda_m(\mathbf{A}_{1,1}). \tag{3.7}$$

Define $\mathbf{\Pi}_m : \mathbb{R}^n \to \mathbb{R}^n$ as the orthogonal projection onto the first m coordinates in $\mathbb{R}^n$. Then, the Courant-Fisher theorem implies

$$\begin{aligned}
\lambda_m(\mathbf{A}_{1,1}) &= \inf_{\substack{V \subset \mathbb{R}^m \\ \dim(V) = m-1}} \sup_{\substack{\mathbf{v} \in V^\perp \\ \|\mathbf{v}\|_2 = 1}} \|\mathbf{A}_{1,1}\mathbf{v}\|_2 \\
&= \inf_{\substack{V \subset \mathbb{R}^n \\ \dim(V) = m-1}} \sup_{\substack{\mathbf{v} \in V^\perp \\ \|\mathbf{v}\|_2 = 1}} \|\mathbf{\Pi}_m \mathbf{A} \mathbf{\Pi}_m \mathbf{v}\|_2 \\
&\leq \inf_{\substack{V \subset \mathbb{R}^n \\ \dim(V) = m-1}} \sup_{\substack{\mathbf{v} \in V^\perp \\ \|\mathbf{v}\|_2 = 1}} \|\mathbf{A}\mathbf{v}\|_2 \\
&= \lambda_m(\mathbf{A}).
\end{aligned}$$

Inserting this estimate into (3.7) gives finally

$$\operatorname{trace}(\mathbf{A} - \mathbf{A}_m) \leq 4^m n \lambda_m(\mathbf{A}) \lesssim n \exp(-bm).$$

This implies $|\log(\varepsilon/n)| \gtrsim m$ and thus the assertion. □

Remark 3.3. *Estimate* (3.6) *is sharp, i.e., the factor* 4^m *cannot be removed. A corresponding example is found in e.g.* [10]*. However, also in case of the approximation theory of the adaptive cross approximation, based on polynomial interpolation, such an exponentially growing factor appears, see* [1] *for the details.*

Remark 3.4. *The trace norm of a matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is equivalent to its* ℓ^2*-norm in case of exponentially decaying eigenvalues* $\lambda_m \sim \exp(-bm)$ *due to*

$$\|\mathbf{A}\|_2 = \max_{m=1}^{n} \lambda_m \sim 1$$

and

$$\operatorname{trace}(\mathbf{A}) = \sum_{m=1}^{n} \lambda_m \sim \sum_{m=1}^{n} \exp(-bm) \sim \int_0^n \exp(-bx)\,\mathrm{d}x \sim 1.$$

4. Applications and results

4.1. **Two-point correlation functions.** The two-point correlation function of second order elliptic problems with stochastic source terms are known to satisfy a deterministic partial differential equation of order 4 with the two-fold tensor product of the elliptic operator on the two-fold cartesian product $\Omega \times \Omega$ of the bounded physical domain $\Omega \subset \mathbb{R}^n$, i.e. in a computational domain of dimension $2n$.

Specifically, let A denote a linear, second order elliptic partial differential operator that maps the Hilbert space $H_0^1(\Omega)$ onto its dual $H^{-1}(\Omega)$. For a given stochastic load vector $f(\omega) \in H^{-1}(\Omega)$ with known expectation and two-point correlation, we consider the stochastic operator equation

$$Au = f.$$

Then the random solution's expectation $\mathbb{E}(u)$ satisfies the mean field equation

$$A\mathbb{E}(u) = \mathbb{E}(f) \tag{4.8}$$

while its two-point correlation is given by

$$(A \otimes A)\,\mathrm{Cor}(u) = \mathrm{Cor}(f), \tag{4.9}$$

see [18, 22] for details. Here, $\mathbb{E}(u)$ denotes the expectation or ensemble average for the random field $u \in H_0^1(\Omega)$ and $\mathrm{Cor}(u) = \mathbb{E}(u \otimes u)$ where now $\mathbb{E}(\cdot)$ denotes the expectation with respect to the product measure on the tensor product space $H_0^1(\Omega) \otimes H_0^1(\Omega)$ (see [18, 22]). Notice that $\mathrm{Cor}(f) \in H^{-1}(\Omega) \otimes H^{-1}(\Omega)$ is a symmetric and positive semi-definite function, that is

$$\int_\Omega \int_\Omega \big(\mathrm{Cor}(f)\big)(\mathbf{x},\mathbf{y})\psi(\mathbf{x})\psi(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \geq 0 \quad \text{for all} \quad \psi \in H^1(\Omega).$$

Nevertheless, one is mainly interested in the variance which is computed by

$$\big(\mathbb{V}(u)\big)(\mathbf{x}) = \big(\mathrm{Cor}(u)\big)(\mathbf{x},\mathbf{x}) - \big(\mathbb{E}(u)\big)^2(\mathbf{x}).$$

In general, second order elliptic partial differential equations like (4.8) are solved by finite elements. To compute the solution of the high-dimensional (4.9) sparse tensor product spaces have been applied in [9, 18, 22]. Whereas one can use also standard finite element methods if a low-rank approximation

$$\big(\mathrm{Cor}(f)\big)(\mathbf{x},\mathbf{y}) \approx \sum_{i=1}^{m} \psi_i(\mathbf{x})\psi_i(\mathbf{y})$$

is available. Then, the two-point correlation of u is simply computed by

$$\big(\mathrm{Cor}(u)\big)(\mathbf{x},\mathbf{y}) \approx \sum_{i=1}^{m} \varphi_i(\mathbf{x})\varphi_i(\mathbf{y}), \quad \text{where} \quad A\varphi_i = \psi_i,\ i = 1, 2, \ldots, m.$$

ε	value of σ				
	0.1	0.2	0.4	0.8	1.6
10^{-1}	85	46	27	14	9
10^{-2}	234	122	66	37	21
10^{-3}	442	236	123	68	38
10^{-4}	710	371	198	108	61
10^{-5}	1038	539	290	157	87
10^{-6}	1426	748	395	214	118

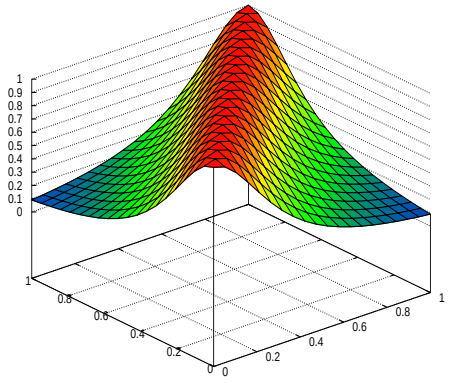


TABLE 4.1. The rank to derive the relative trace error ε.

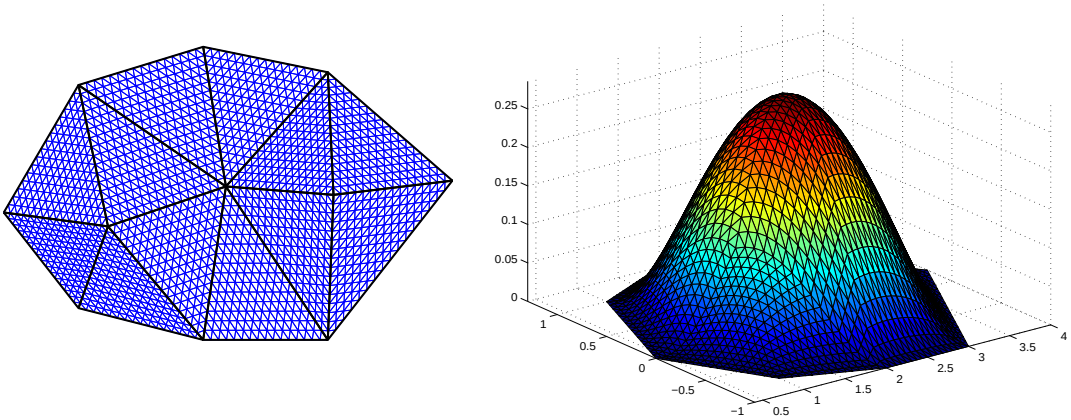


FIGURE 4.1. The computational domain Ω with triangulation on level 4 (left) and the approximate variance $\mathbb{V}(u)$ (right).

Especially, we have

$$\big(\mathbb{V}(u)\big)(\mathbf{x}) \approx \sum_{i=1}^{m} \varphi_i^2(\mathbf{x}) - \big(\mathbb{E}(u)\big)^2(\mathbf{x}).$$

Therefore, if n finite elements are applied, the complexity of solving (4.9) is reduced to $\mathcal{O}(n \cdot m)$, provided that one solve is of linear cost. If m is small this means a large reduction compared to a naive computation which is of quadratic complexity $\mathcal{O}(n^2)$.

To demonstrate the approach, we triangulate of the domain Ω by uniform refinement of a coarse grid triangulation as seen in the left plot of Fig. 4.1. This yields a finite element mesh with about 800000 elements and 400000 nodes. On the given

triangulation we like to interpolate the (analytical) two-point correlation

$$\big(\operatorname{Cor}(f)\big)(\mathbf{x},\mathbf{y}) = \frac{1}{\sigma + \|\mathbf{x}-\mathbf{y}\|_2^2}, \quad \sigma > 0, \tag{4.10}$$

by a piecewise linear function. To compute this interpolation, we perform the pivoted Cholesky decomposition until the relative trace norm is less than a given threshold parameter ε. In Tab. 4.1, we tabulated the number of Cholesky steps required to get the desired accuracy, depending on the size of σ. Notice that it holds $\operatorname{diam}(\Omega) = 4$. The plot which accompanies Tab. 4.1 shows an illustration of the covariance function (4.10) in one spatial dimension. The approximation of the variance $\mathbb{V}(u)$ in case of $\sigma = 0.2$, $A = -\Delta$, and $\mathbb{E}(f) \equiv 0$ is found in the right plot of Fig. 4.1.

Strang's first lemma together with the Aubin-Nitsche trick yields the error estimate

$$\|\operatorname{Cor}(u) - \widehat{\operatorname{Cor}(u)}\|_{L^2(\Omega\times\Omega)} \lesssim \|\operatorname{Cor}(f) - \widehat{\operatorname{Cor}(f)}\|_{L^2(\Omega\times\Omega)}.$$

Here, "$\widehat{\cdot}$" indicates numerical approximation. Therefore, the solution's accuracy is $\mathcal{O}(\varepsilon)$ provided that the L^2-error of low-rank approximation is ε.

4.2. **Generalized eigenvalue problems.** For computing a low-rank approximation of a symmetric and positive semi-definite function $f \in H^p(\Omega\times\Omega)$, $p \geq 0$, we may exploit the eigenvalue decomposition of the Hilbert-Schmidt operator

$$(\mathcal{K}u)(\mathbf{x}) = \int_\Omega f(\mathbf{x},\mathbf{y})u(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

This gives raise to the following decomposition

$$f(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{\infty} \lambda_k\varphi_k(\mathbf{x})\varphi_k(\mathbf{y}) \quad \text{where} \quad \mathcal{K}\varphi_k = \lambda_k\varphi_k,\ k = 1,2,\ldots$$

into orthonormal functions $\{\varphi_k\}$. It follows immediately from [23] that the sequence of eigenvalues $\{\lambda_k\}$ decays as $\lambda_k \lesssim k^{-p/n}$. In case of f being even analytical, one can proof the (sub-) exponential rate $\lambda_k \lesssim \sqrt[n]{\exp(-bk)}$ for some $b > 0$.

The Galerkin discretization of the eigenvalue problem yields a generalized eigenvalue problem

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}, \quad \mathbf{A} = [(\mathcal{K}\psi_i,\psi_j)]_{i,j}, \quad \mathbf{B} = [(\psi_i,\psi_j)]_{i,j}. \tag{4.11}$$

Here, the system matrix $\mathbf{A}$ is symmetric and positive semi-definite and the mass matrix $\mathbf{B}$ is symmetric and positive definite. We shall assume that the ansatz functions are stable and compactly supported such that the mass matrix is well-conditioned and has only $\mathcal{O}(n)$ coefficients.

Having the low-rank approximation

$$\mathbf{A} \approx \mathbf{A}_m := \mathbf{L}_m\mathbf{L}_m^T, \quad \mathbf{L}_m \in \mathbb{R}^{n\times m}$$

at hand, we can replace $\mathbf{A}$ by its low rank approximation to arrive at

$$\mathbf{L}_m\mathbf{L}_m^T\mathbf{x} = \widehat{\lambda}\mathbf{B}\mathbf{x}. \tag{4.12}$$

Multiplying from the left and from the right by $\mathbf{B}^{-1/2}$ yields

$$\mathbf{B}^{-1/2}\mathbf{L}_m\mathbf{L}_m^T\mathbf{B}^{-1/2}\widehat{\mathbf{x}} = \widehat{\lambda}\widehat{\mathbf{x}}, \quad \mathbf{x} = \mathbf{B}^{-1/2}\widehat{\mathbf{x}}.$$

Since $(\mathbf{B}^{-1/2}\mathbf{L}_m)(\mathbf{B}^{-1/2}\mathbf{L}_m)^T$ has the same eigenvalues as $(\mathbf{B}^{-1/2}\mathbf{L}_m)^T(\mathbf{B}^{-1/2}\mathbf{L}_m)$, we can replace the generalized eigenvalue problem (4.12) in $\mathbb{R}^{n\times n}$ by an equivalent standard eigenvalue problem in $\mathbb{R}^{m\times m}$:

$$\mathbf{L}_m^T\mathbf{B}^{-1}\mathbf{L}_m\widehat{\mathbf{x}} = \widehat{\lambda}\widehat{\mathbf{x}}, \quad \mathbf{x} = \mathbf{B}^{-1}\mathbf{L}_m\widehat{\mathbf{x}}. \tag{4.13}$$

Compared to the original eigenvalue problem (4.11), this small eigenvalue problem is much cheaper to solve if $m \ll n$. Namely, the complexity is $\mathcal{O}(m^2n)$ instead of $\mathcal{O}(n^3)$ since $\mathbf{B}^{-1}\boldsymbol{\ell}_i$ $(i = 1, 2, \ldots, m)$ can iteratively be computed in linear complexity. It has only nonzero eigenvalues $\{\widehat{\lambda}_k\}_{k=1}^m$ which coincide with the nonzero eigenvalues of large problem (4.12). The Bauer-Fike theorem implies immediately an error bound relative to the eigenvalues $\{\lambda_k\}_{k=1}^n$ of original problem (4.11):

$$\begin{aligned}|\lambda_k - \widehat{\lambda}_k| &\le \|\mathbf{B}^{-1/2}(\mathbf{A} - \mathbf{A}_m)\mathbf{B}^{-1/2}\|_2 \\ &= \|\mathbf{B}^{-1}(\mathbf{A} - \mathbf{A}_m)\|_2 \\ &\lesssim \|\mathbf{A} - \mathbf{A}_m\|_2, \qquad k = 1, 2, \ldots, m.\end{aligned}$$

For numerical test calculations we choose the unit square $[0, 1]^2$ as computational domain and consider the following kernel functions. Illustrations of these kernel functions are found in Tabs. 4.2–4.5.

(1) The Gauss kernel

$$f(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{|x - y|^2}{\sigma^2}\right)$$

which is analytical.

(2) A jumping Gauss kernel

$$f(x, y) = \frac{1}{a(x, y)\sqrt{2\pi\sigma^2}} \exp\left(\frac{|x - y|^2}{\sigma^2}\right)$$

with $a(x, y) = 1$ if $x, y < 0.5$, $a(x, y) = 4$ if $x, y \ge 0.5$ and $a(x, y) = -2$ elsewhere.

(3) The Poisson kernel

$$f(x, y) = \frac{1}{\sqrt{\sigma}} \exp(-\sigma|x - y|)$$

whose eigenvalues decay algebraically like $\lambda_k \sim k^{-2}$.

ε	value of σ				
	1	0.5	0.1	0.05	0.01
10^{-1}	2	3	10	19	89
10^{-2}	3	5	15	28	137
10^{-3}	4	5	19	36	173
10^{-4}	5	6	21	39	187
10^{-5}	5	7	24	46	214
10^{-6}	5	8	27	50	238

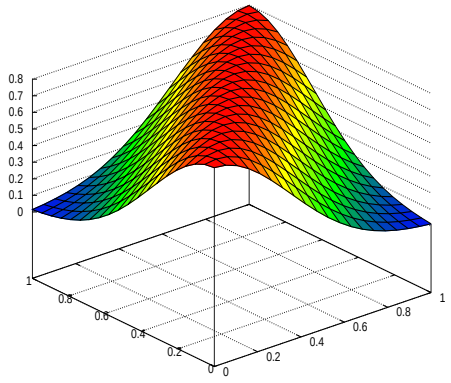


TABLE 4.2. Gauss kernel: The rank to derive the relative trace error ε.

ε	value of σ				
	1	0.5	0.1	0.05	0.01
10^{-1}	2	3	10	17	81
10^{-2}	3	4	15	28	131
10^{-3}	4	5	18	34	168
10^{-4}	4	6	21	39	186
10^{-5}	5	7	24	45	211
10^{-6}	5	8	26	50	234

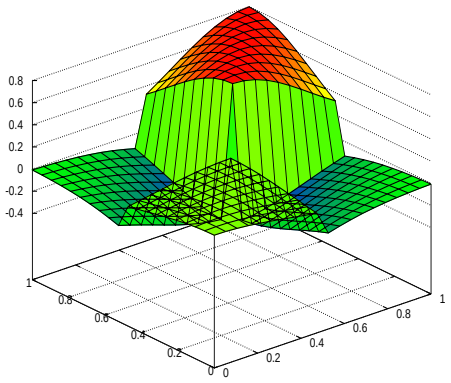


TABLE 4.3. Jumping Gauss kernel: The rank to derive the relative trace error ε.

(4) A random kernel which is given by

$$\mathbf{A} = \sum_{k=1}^{m} \lambda_k \mathbf{v}\mathbf{v}^T, \quad \lambda_k = \exp(-\sigma k), \quad \mathbf{v}_k^T \mathbf{v}_\ell = \delta_{k,\ell}.$$

We generate $m = 2000$ vectors $\{\mathbf{v}_k\}$ as n-dimensional arrays of in $[0, 1]$ uniformly distributed numbers which are then orthogonalized.

We subdivide the interval $[0, 1]$ into n equidistant intervals and apply continuous L^1-normalized piecewise linear ansatz and test functions. The integer n is chosen as 10^5 in case of the Poisson kernel and the random kernel, whereas it is chosen as 10^6 in case of the Gaussian type kernels. For different choices of the parameters σ and ε we compute the low-rank approximation by the pivoted Cholesky decomposition. The results are tabulated in the Tabs. 4.2–4.5.

ε	value of σ				
	1	10^{-1}	10^{-2}	10^{-3}	10^{-4}
10^{-1}	5	1	1	1	1
10^{-2}	36	5	1	1	1
10^{-3}	376	36	5	1	1
10^{-4}	3616	376	36	5	1

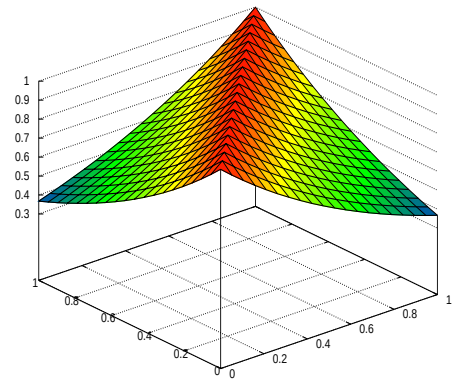


TABLE 4.4. Poisson kernel: The rank to derive the relative trace error ε.

ε	value of σ				
	1	0.5	0.1	0.05	0.01
10^{-1}	3	6	29	61	333
10^{-2}	6	11	56	115	610
10^{-3}	8	15	81	167	873
10^{-4}	10	21	106	216	1126
10^{-5}	13	25	130	266	1375
10^{-6}	15	30	154	315	1618

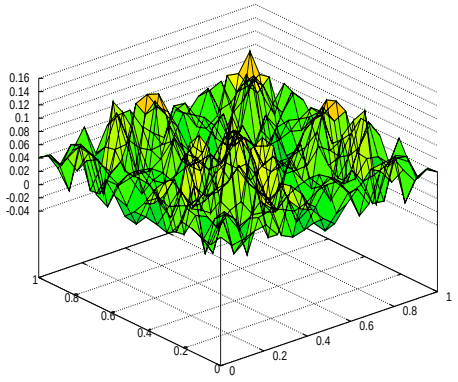


TABLE 4.5. Random kernel: The rank to derive the relative trace error ε.

It turns out that the algorithm always converges even though the kernel functions do not match the assumptions of Thm. 3.2 except for the Gaussian type kernels in case of $\sigma = 0.1$ and $\sigma = 0.5$. Even in case of the algebraically decaying Poisson kernel the approach is feasible, although we cannot reach a high accuracy. Moreover, the convergence of the Gauss and the jumping Gauss kernel are nearly identical. Contrary to our theory the convergence analysis of the adaptive cross approximation does not include discontinuous kernel functions since it is based on (global) polynomial interpolation, cf. [1, 2].

In Figs. 4.2 and 4.3 we plotted the trace error in the m-th step of the algorithm versus the rank of the matrix $\mathbf{A}_m$ (indicated by red stars). The eigenvalues of the matrix $\mathbf{A}$ correspond to the green squares, whereas the eigenvalues of the underlying Hilbert-Schmidt operator (i.e., the solution of generalized eigenvalue problem) correspond

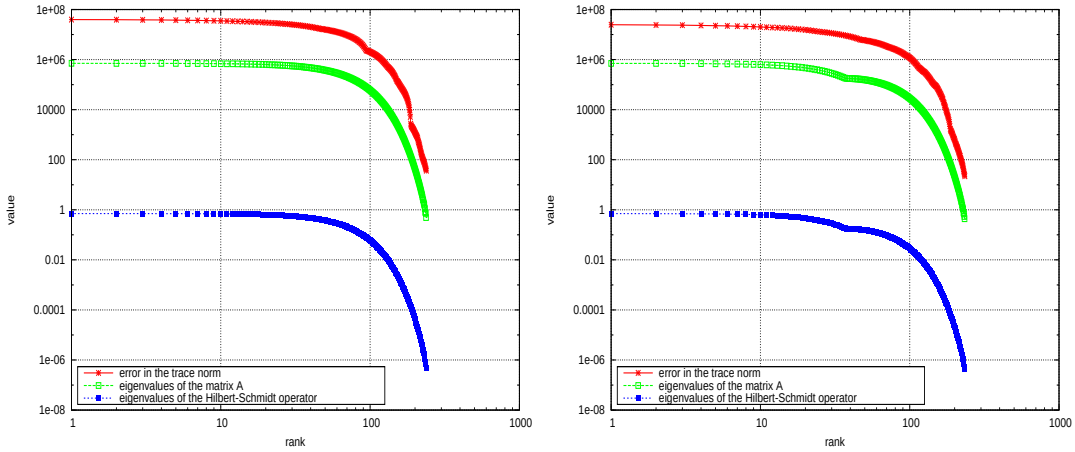


FIGURE 4.2. Gaussian (left) and jumping Gaussian (right) kernel: Eigenvalues and trace error versus the rank.

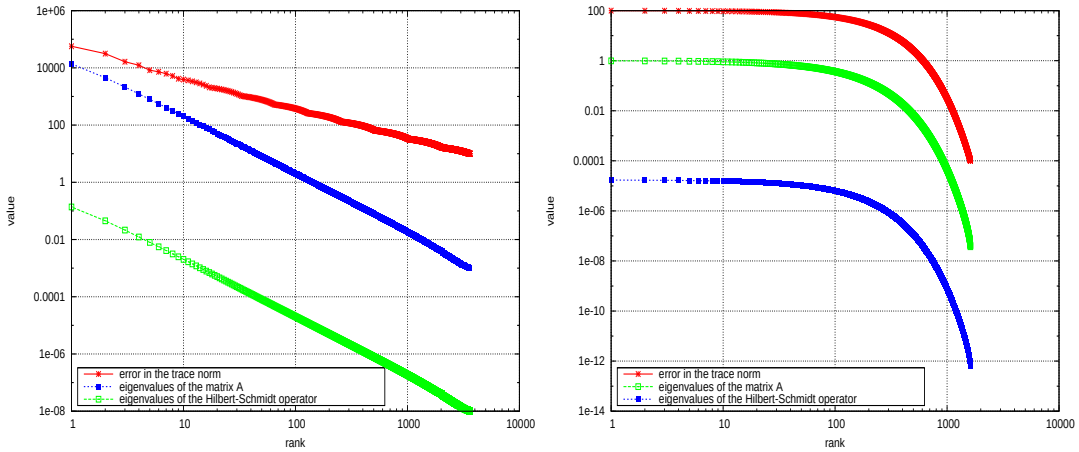


FIGURE 4.3. Poisson (left) and random (right) kernel: Eigenvalues and trace error versus the rank.

to the blue boxes. In case of the exponentially decreasing eigenvalues, i.e., in case of the Gauss, the jumping Gauss and the random kernel, all curves exhibit the same exponential decay which means that the trace norm converges optimally. This issues from the fact that the trace norm is equivalent to the ℓ^2-norm if the eigenvalues decay exponentially:

$$\sum_{k=1}^{\ell} \exp(-bk) \sim \int_0^{\ell} \exp(-bx)\,\mathrm{d}x \sim \exp(-b\ell).$$

In case of the Poisson kernel the eigenvalues decay only quadratically and thus the trace norm can converge at most only linearly. In fact, as figured out of Fig. 4.3, exactly this behavior is observed.

To compute the low-rank approximation of the Gaussian type kernels we needed only 4 minutes in case of $\varepsilon = 10^{-6}$ and $\sigma = 0.01$. Recall that $n = 10^6$ and $m \approx 240$. The transformation to the small eigenvalue problem (4.13) and its solution requires then only additional 3 minutes. This is much faster than the solution of the large eigenvalue problem (4.12) by the well-known ARPACK library [15] which provides an implicitly restarted Arnoldi/Lanczos method. Here, the solution requires 50 minutes which means that our approach leads to a speed-up of more than 10.

Our numerical experiments indicate that in general the exponential growth factor 4^m in Thm. 3.2 does not appear in praxis. Likewise, concerning the influence of the factor n which appears in the error estimate of Thm. 3.2, we did not observe an increase of the rank when n is increased while ε is fixed.

4.3. **Quantum chemistry.** In quantum chemistry, in particular electronic structure calculation the following quantities

$$V_{a,i}^{b,j} := \sum_{s,s'=\pm\frac{1}{2}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\overline{\varphi_a(\mathbf{r},s)\varphi_b(\mathbf{r}',s')}\varphi_i(\mathbf{r},s)\varphi_j(\mathbf{r}',s')}{\|\mathbf{r}-\mathbf{r}'\|_2}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'. \tag{4.14}$$

are called *two-electron integrals*. They play a crucial role in many circumstances, particularly in wave function methods like coupled cluster method etc. Here the functions $\varphi_i : \mathbb{R}^3 \times \{\pm\frac{1}{2}\} \to \mathbb{C}$, $(\mathbf{r},s) \mapsto \varphi_i(\mathbf{r},s) = \phi_i(\mathbf{r})\chi_i(s) \in H^1(\mathbb{R}^3) \times \{\pm\frac{1}{2}\}$, $i = 1,2,\ldots,K$, are given atomic basis functions, where χ_i is one of the two basis spin functions α, β for either spin up or spin down. It may be assumed, after basis transformation, that the functions $\varphi_i(\mathbf{r},s)$ are pairwise orthogonal. For instance, this is the case for the eigenfunctions of the Fock operator, called *canonical molecular orbitals*.

In the formulation of second quantization, the *Hamilton operator* for the *electronic Schrödinger equation*

$$\mathcal{H}\Psi = E\Psi,$$

where

$$\mathcal{H}\Psi := \sum_{i=1}^{N}\left[-\frac{1}{2}\Delta_i - \sum_{j=1}^{M}\frac{Z_j}{\|\mathbf{r}_i-\mathbf{R}_j\|_2} + \frac{1}{2}\sum_{j\neq i}^{N}\frac{1}{\|\mathbf{r}_i-\mathbf{r}_j\|_2}\right]\Psi$$

is given by

$$\mathcal{H}\Psi = \sum_{p,q} h_q^p a_q^\dagger a_p + \frac{1}{2}\sum_{p,q,r,s} V_{q,s}^{p,r} a_r^\dagger a_s^\dagger a_p a_q$$

with the single-electron integrals

$$h_i^j := \sum_{s=\pm\frac{1}{2}} \int_{\mathbb{R}^3} \left\{ \overline{\nabla\varphi_i(\mathbf{r},s)}\nabla\varphi_j(\mathbf{r},s) - \sum_{k=1}^{M} Z_k \frac{\overline{\varphi_i(\mathbf{r},s)}\varphi_j(\mathbf{r},s)}{\|\mathbf{r}-\mathbf{R}_k\|_2} \right\} \mathrm{d}\mathbf{r}$$

and the two-electron integrals defined in (4.14). The symbols $a^\dagger$ and a denote creation and annihilation operators of second quantization.

The tensor $\mathbf{V}$ with K^4 entries can be casted in a matrix

$$\mathbf{V} = \left[V_{a,i}^{b,j}\right] \in \mathbb{R}^{K^2 \times K^2}$$

with row indices (a, i), $a, i = 1, 2, \ldots, K$, and columns indexed by (b, j). It is reasonable, e.g. for local basis functions or even for polynomials, that the K^2 functions $\varphi_a(\mathbf{r}, s)\varphi_i(\mathbf{r}, s)$ can be approximated simultaneously within a linear space of dimension $\mathcal{O}(K)$. Consequently, the matrix $\mathbf{V}$ is essentially of rank $\mathcal{O}(K)$ instead of K^2. This fact has firstly been observed in [3] and later used in *density fitting* or *resolution of identity* where an auxiliary basis set $\{\psi_p\}$ is constructed. We do not go into the details here and refer the reader to [25, 26, 27].

Since it is easy to see that $\mathbf{V}$ is symmetric and positive semi-definite, an $\mathcal{O}(K)$-rank approximation can also be computed by the pivoted Cholesky factorization, with the mentioned advantages and properties. The whole computation requires at most $\mathcal{O}(K^4)$ operation and reduces the memory requirement to $\mathcal{O}(K^3)$. In particular, the cost of the matrix-vector multiplication is reduced to $\mathcal{O}(K^3)$ while the approximation is satisfactory [14, 28]. In contrast to density fitting the accuracy can be controlled by the trace norm. Moreover, the sparsity of $\mathbf{V}$ could be maintained quite often during the Cholesky decomposition which reduces the complexity even to $\mathcal{O}(K)$ for both, computation and storage. This effect can also be exploited in linear scaling DFT calculation. With these methods at hand, for large systems and using local basis functions, the whole cost could be reduced to $\mathcal{O}(K)$ for linear scaling MP2 (Møller-Plesset 2nd order perturbation theory) [16, 19] or coupled cluster calculations [20, 21].

References

[1] M. Bebendorf. Approximation of boundary element matrices. *Numer. Math.*, 86:565–589, 2000.

[2] M. Bebendorf and S. Rjasanow. Adaptive low-rank approximation of collocation matrices. *Computing*, 70:1–24, 2003.

[3] N.H.F. Beebe and J. Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *Int. J. Quantum Chem.*, 7:683–705, 1977.

[4] L. Boman, H. Koch, and A.S. de Merás. Method specific Cholesky decomposition: Coulomb and exchange energies. *J. Chem. Phys.*, 129:134107, 2008.

[5] J.J. Dongarra, C.B. Moler, J.R. Bunch, and G.W. Stewart. *LINPACK users' guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1979.

[6] S.A. Goreinov, E.E. Tyrtyshnikov, and A.Y. Yeremin. Matrix-free iterative solution strategies for large dense linear systems. *Numer. Linear Algebra Appl.*, 4:273–294, 1997.

[7] S.A. Goreinov, E.E. Tyrtyshnikov, and N.L. Zamarashkin. A theory of pseudoskeleton approximations. *Linear Algebra Appl.*, 261:1–21, 1997.

[8] S. Hammarling, N. Higham, and C. Lucas. LAPACK-style codes for pivoted Cholesky and QR updating. In B. Kågström et al., editors, *Lecture Notes in Computer Science*, volume 4699, pages 137–146, Springer, Berlin/Heidelberg, 2007.

[9] H. Harbrecht, R. Schneider, and C. Schwab. Multilevel frames for sparse tensor product spaces. *Numer. Math.*, 110:199–220, 2008.

[10] N. Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 29:575–596, 1987.

[11] N. Higham. Analysis of the Cholesky decomposition of a semi-definite matrix. In M.G. Cox and S.J. Hammarling, editors, *Reliable Numerical Computation*, pages 161–185, Oxford University Press, Oxford, 1990.

[12] N. Higham. Cholesky factorization. In E.J. Wegman, Y.H. Said, and D.W. Scott, editors, *Wiley Interdisciplinary Reviews: Computational Statistics*, volume 1, pages 251–254, Wiley, New York, 2009.

[13] B. Kaltenbacher. Regularization by truncated Cholesky factorization: a comparison of four different approaches. *Journal of Complexity*, 23:225–244, 2007.

[14] H. Koch, A.S. de Merás, and T.B. Pedersen. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.*, 118:9481–9484, 2003.

[15] R.B. Lehoucq, D.C. Sorensen, and C. Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.

[16] J. Zienau, L. Clin, B. Doser , and C. Ochsenfeld. Cholesky-decomposed densities in Laplace-based second-order Møller-Plesset perturbation theory. *J. Chem. Phys.*, 130:204112, 2009.

[17] D.W. O'neal and J. Simons. Application of Cholesky-like matrix decomposition methods to the evaluation of atomic orbital integrals and integral derivatives. *Int. J. Quantum Chem.*, 36:673–688, 2004.

[18] T. von Petersdorff and C. Schwab. Sparse wavelet methods for operator equations with stochastic data. *Appl. Math.*, 51:145–180, 2006.

[19] M. Schütz, G. Hetzer, and H.-J. Werner. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.*, 111:5691–5705, 1999.

[20] M. Schütz and H.-J. Werner. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.*, 114:661–681, 2001.

[21] M. Schütz. A new, fast, semi-direct implementation of linear scaling local coupled cluster theory. *Phys. Chem. Chem. Phys.*, 4:3941–3947, 2002.

[22] C. Schwab and R. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.*, 95:707–734, 2003.

[23] C. Schwab and R. Todor. Karhunen-Loéve approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217:100–122, 2006.
[24] E.E. Tyrtyshnikov. Mosaic-skeleton approximation. *Calcolo*, 33:47-57, 1996.
[25] O. Vahtras, J.E. Almlöf, and M.W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, 213:514–518, 1993.
[26] F. Weigend, A. Köhn, and C. Hattig. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.*, 116:3175–3183, 2002.
[27] H.-J. Werner, F.R. Manby, and P.J. Knowles. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.*, 118:8149–8160, 2003.
[28] S. Wilson. Universal basis sets and Cholesky decomposition of the two-electron integral matrix. *Comput. Phys. Commun.*, 58:71–81, 1990.

Helmut Harbrecht, Institut für Angewandte Mathematik und Numerische Simulation, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany.

E-mail address: harbrecht@ians.uni-stuttgart.de

Michael Peters, Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany.

E-mail address: peters@ins.uni-bonn.de

Reinhold Schneider, Institut für Mathematik, Technische Universität Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany.

E-mail address: schneidr@math.tu-berlin.de

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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