Approximation of bi-variate functions: singular value decomposition versus sparse grids

Michael Griebel

Institut für Numerische Simulation, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany
griebel@ins.uni-bonn.de

AND

Helmut Harbrecht[*]

Mathematisches Institut, Universität Basel, Rheinsprung 21, 4051 Basel, Switzerland
[*]Corresponding author: helmut.harbrecht@unibas.ch

We compare the cost complexities of two approximation schemes for functions $f \in H^p(\Omega_1 \times \Omega_2)$ which live on the product domain $\Omega_1 \times \Omega_2$ of sufficiently smooth domains $\Omega_1 \subset \mathbb{R}^{n_1}$ and $\Omega_2 \subset \mathbb{R}^{n_2}$, namely the singular value/Karhunen–Lòeve decomposition and the sparse grid representation. Here, we assume that suitable finite element methods with associated *fixed* order r of accuracy are given on the domains Ω_1 and Ω_2. Then, the sparse grid approximation essentially needs only $\mathcal{O}(\varepsilon^{-q})$, with $q = \max\{n_1, n_2\}/r$, unknowns to reach a prescribed accuracy ε, provided that the smoothness of f satisfies $p \geqslant r((n_1 + n_2)/\max\{n_1, n_2\})$, which is an almost optimal rate. The singular value decomposition produces this rate only if f is analytical, since otherwise the decay of the singular values is not fast enough. If $p < r((n_1 + n_2)/\max\{n_1, n_2\})$, then the sparse grid approach gives essentially the rate $\mathcal{O}(\varepsilon^{-q})$ with $q = (n_1 + n_2)/p$, while, for the singular value decomposition, we can only prove the rate $\mathcal{O}(\varepsilon^{-q})$ with $q = (2\min\{r, p\}\min\{n_1, n_2\} + 2p\max\{n_1, n_2\})/(2p - \min\{n_1, n_2\})\min\{r, p\}$. We derive the resulting complexities, compare the two approaches and present numerical results which demonstrate that these rates are also achieved in numerical practice.

Keywords: singular value decomposition; sparse grids; complexity.

1. Introduction

The efficient approximate representation of multivariate functions is an important task in numerical analysis and scientific computing. In this paper, we concentrate on functions which are defined on the product $\Omega_1 \times \Omega_2$ of two domains. Already for this simple situation, there exists a large amount of applications. For example, radiosity models and radiative transfer (Widmer *et al.*, 2008) exhibit a product structure. Here, Ω_1 denotes the spatial three-dimensional domain of the geometric object under consideration and Ω_2 is the sphere $\mathbb{S}^2$. Moreover, in the case of space-time discretizations of parabolic problems, Ω_1 is the time interval, whereas Ω_2 is the spatial domain (Griebel & Oeltz, 2007; Stevenson & Schwab, 2009).

Then, there are various phase-space problems where both Ω_1 and Ω_2 are three-dimensional cubes or the full three-dimensional real space. Examples are the Boltzmann equation, kinetic equations or the Langevin equation; see, e.g., Balescu (1997). Furthermore, non-Newtonian flow can be modelled by a coupled system that consists of the Navier–Stokes equations for the flow in a three-dimensional

geometry described by Ω_1 and of the Fokker–Planck equation in a $3(k-1)$-dimensional configuration space on Ω_2. Here, k denotes the number of atoms in a chain-like molecule which constitutes the non-Newtonian behaviour of the flow; for details see Barrett *et al.* (2009), Le Bris & Lelièvre (2009) and Lozinski *et al.* (2010). Note that the domain of the configuration space is itself again a product of $k-1$ spheres.

Another example is two-scale homogenization. After unfolding (Cioranescu *et al.*, 2008), it gives rise to the product of the macroscopic physical domain and the periodic microscopic domain of the cell problem; see Matache (2002). For multiple scales, a general product appears here which still can be written as the product of two domains, one containing, for example, the macroscopic scale and the other consisting of the product of the domains of the different microscales (Hoang & Schwab, 2005).

Also, the two-point correlation functions of linear elliptic boundary value problems with stochastic source terms, i.e., $\mathrm{A}u(\omega) = f(\omega)$ in Ω, are known to satisfy a deterministic partial differential equation (PDE) on the product domain $\Omega \times \Omega$. Namely, the two-point correlation satisfies the equation $(A \otimes A) \, \mathrm{Cor}_u = \mathrm{Cor}_f$ in $\Omega \times \Omega$; see Schwab & Todor (2003a). Higher-order moments then involve larger tensor products (Schwab & Todor, 2003b). This approach extends to stochastic diffusion functions and more general PDEs with stochastic coefficient functions as well as to stochastic domains (Harbrecht *et al.*, 2008a; Harbrecht, 2010).

Finally, we find the product of two domains in quantum mechanics, for example, for the Schrödinger equation for helium; systems with more than two electrons then involve multiple product domains, of course.

In general, some problems are directly given on the product of two domains, while for other problems the domains themselves are products of lower-dimensional domains. Then, the domain of an *n*-dimensional problem with, for example, *n* being some power of two, can be split into the tensor product of two domains of dimension $n/2$ which can be recursively further split until a terminal situation (a one-dimensional domain or a truly higher-dimensional but nontensor-product domain) is reached. Related representation methods have recently been considered in Hackbusch & Kühn (2009) or Oseledets & Tyrtyshnikov (2009), Grasedyck (2010), Bebendorf (2011), Hackbusch (2012).

In this article, we consider the simple case of two domains Ω_1 and Ω_2 only. Here, our analysis covers the situation of the first bisection step in the above-mentioned recursion. To this end, for $i = 1, 2$, let Ω_i denote a domain in $\mathbb{R}^{n_i}$ (or alternatively also an n_i-dimensional manifold in $\mathbb{R}^{n_i+1}$). We consider the numerical approximation of bi-variate functions. In other words, we suppose that on the individual subdomains we have chosen fixed sequences of nested trial spaces

$$V_0^{(i)} \subset V_1^{(i)} \subset V_2^{(i)} \subset \cdots \subset L^2(\Omega_i),$$

each of which consists of ansatz functions of the same, fixed approximation order *r*. On the basis of these ansatz spaces, we intend to compare the efficient approximation of functions $f(\mathbf{x}, \mathbf{y}) \in L^2(\Omega_1 \times \Omega_2)$ by either the truncated singular value decomposition

$$f_M(\mathbf{x}, \mathbf{y}) := \sum_{\ell=1}^{M} \sqrt{\lambda_\ell} \varphi_\ell(\mathbf{x}) \psi_\ell(\mathbf{y}),$$

or by the generalized sparse grid approach

$$\hat{f}_J(\mathbf{x}, \mathbf{y}) := \sum_{j_1/\sigma + j_2\sigma \leqslant J} \sum_{k_1 \in \nabla_{j_1}^{(1)}} \sum_{k_2 \in \nabla_{j_2}^{(2)}} \beta_{(j_1, k_1), (j_2, k_2)} \xi_{j_1, k_1}^{(1)}(\mathbf{x}) \xi_{j_2, k_2}^{(2)}(\mathbf{y}),$$

where $\sigma > 0$ is an appropriately chosen parameter. In the first representation, $\{\varphi_\ell\}_{\ell=1}^M$ and $\{\psi_\ell\}_{\ell=1}^M$ are sets of orthonormal functions. They are unknown *a priori*, cannot in general be derived analytically and need thus to be approximated in the ansatz spaces $\{V_j^{(i)}\}$. In other words, the approximation involves, in most applications, both a truncation after M terms *and* an approximate computation of the singular values and the associated left- and right-singular vectors. In the second representation, $\{\xi_{j,k}^{(i)}\}_{k\in\nabla_j^{(i)}, j\in\mathbb{N}}$ are in general multilevel or wavelet bases associated with the trial spaces, where the index j refers to the level of resolution and the index k refers to the locality of the basis function (the precise definition will be given in Section 4).

For our comparison we consider the smoothness of the function f to be measured in isotropic Sobolev norms. The result of this paper is then as follows: given a function $f \in H^p(\Omega_1 \times \Omega_2)$, we have to use $\mathcal{O}(\varepsilon^{-q})$, with $q = (2\min\{r,p\}\min\{n_1,n_2\} + 2p\max\{n_1,n_2\})/(2p - \min\{n_1,n_2\})\min\{r,p\}$, degrees of freedom for the computation of the truncated and approximated singular value decomposition to reach a specific prescribed accuracy ε. For the general sparse grid method with associated parameter $\sigma = \sqrt{n_1/n_2}$ (a precise definition is given in Section 4), it is known that this cost-complexity rate is of the order $\mathcal{O}(\varepsilon^{-q})$ with $q = \max\{(n_1+n_2)/p, n_1/r, n_2/r\}$ (Griebel & Harbrecht, 2013). We compare the two approaches and present numerical results, which demonstrate that these rates are also achieved in practice.

Note that we have only an upper bound for the truncated and approximated singular value decomposition. Therefore, our comparison of the singular value decomposition and the sparse grid approach (see Section 5) needs to be interpreted carefully. Indeed, the present paper is only a first attempt towards a 'fair' comparison of both approaches, which particularly takes into account that the eigenvectors of the approximated singular value decomposition need to be computed and stored in any practical numerical approach.

The remainder of this article is organized as follows. In Section 2, we give a short introduction to multilevel approximation. In Section 3, we describe the singular value decomposition of a bivariate function on $\Omega_1 \times \Omega_2$ and discuss its approximation properties in detail. Section 4 gives the basics of the so-called general sparse grid approximation of a bivariate function on $\Omega_1 \times \Omega_2$ and presents its error rates and cost complexities. In Section 5, we compare the two approximations. In Section 6, we present numerical results which show that the theoretical complexity rates are also achieved in practice. Section 7 concludes with some final remarks.

Throughout this paper, the notion 'essential' in connection with the complexity estimates means 'up to logarithmic terms'. Moreover, to avoid the repeated use of generic but unspecified constants, we denote 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. Approximation on the subdomains

Let $\Omega \subset \mathbb{R}^n$ be a sufficiently smooth, bounded domain. In general, one uses finite elements to approximate functions on $L^2(\Omega)$. In the present paper, we focus on the common h method, i.e., on finite elements of fixed approximation order. Then, particularly for applying multiscale techniques, one has a sequence of nested trial spaces

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2(\Omega) \tag{2.1}$$

such that

$$L^2(\Omega) = \overline{\bigcup_{j\in\mathbb{N}_0} V_j},$$

which is called *multiscale analysis*. Each space V_j is defined by a single-scale basis $\Phi_j = \{\phi_{j,k}\}$, i.e., $V_j = \text{span}\{\phi_{j,k} : k \in \Delta_j\}$, where Δ_j denotes a suitable index set with cardinality $\#(\Delta_j) \sim 2^{nj}$.

We say that the trial spaces have *(approximation) order* $r \in \mathbb{N}$ if

$$r = \sup \left\{ s \in \mathbb{R} : \inf_{v_j \in V_j} \|v - v_j\|_{L^2(\Omega)} \lesssim h_j^s \|v\|_s \; \forall v \in H^s(\Omega) \right\}, \tag{2.2}$$

where the quantity $h_j \sim 2^{-j}$ corresponds to the mesh width associated with the subspace V_j on Ω. Note that the integer $r > 0$ refers in general to the maximal order of polynomials that are locally contained in V_j.

Equation (2.2) implies that a given function $v \in H^p(\Omega)$, $0 \leqslant p \leqslant r$ can be approximated in V_j at a rate h_j^p, i.e.,

$$\inf_{v_j \in V_j} \|v - v_j\|_{L^2(\Omega)} \lesssim h_j^p \|v\|_{H^p(\Omega)}, \quad 0 \leqslant p \leqslant r. \tag{2.3}$$

Thus, when we approximate a function $v \in H^p(\Omega)$ with $0 \leqslant p \leqslant r$ by uniform mesh refinement, we obtain the rate h_j^p according to (2.3). Since the mesh size and the number of unknowns in V_j are related by $\dim(V_j) \sim 2^{jn} \sim h_j^{-n}$, we deduce that

$$N \sim \varepsilon^{-n/p} \tag{2.4}$$

unknowns have to be spent to achieve an approximation error ε. The best possible rate $N^{-n/r}$ is achieved if $p = r$, i.e., if $v \in H^r(\Omega)$.

For our subsequent analysis, to efficiently approximate bi-variate functions in $L^2(\Omega_1 \times \Omega_2)$, we shall fix the definitions, properties and cost complexities individually for each subdomain $\Omega_i \in \mathbb{R}^{n_i}$, $i = 1, 2$. That is, we fix two multiscale analyses

$$V_0^{(i)} \subset V_1^{(i)} \subset V_2^{(i)} \subset \cdots \subset L^2(\Omega_i), \quad i = 1, 2, \tag{2.5}$$

which are assumed to provide the *same* approximation order r. The construction of the truncated and approximated singular value decomposition in Section 3 and the sparse grid in Section 4 are based on these multiscale analyses. A comparison for the case of the *same* r, i.e., where the underlying finite element spaces in (2.5) for both the truncated and approximated singular value decomposition and the sparse grid, have the same order r, then takes place in Section 5. The question we address in this paper is as follows. Given a function $H^p(\Omega_1 \times \Omega_2)$, where $p > 0$ is arbitrary and where trial spaces with fixed order r are used in both approaches, which algorithm provides a cheaper approximation?

3. Singular value decomposition

3.1 *Definition and mapping properties*

We intend to numerically represent functions $f(\mathbf{x}, \mathbf{y}) \in L^2(\Omega_1 \times \Omega_2)$ on tensor-product domains $\Omega_1 \times \Omega_2$ in an efficient way. One way to solve this approximation problem is to use an ansatz by means of tensor products, which separates the variables $\mathbf{x}$ and $\mathbf{y}$. We first consider the approximation

$$f(\mathbf{x}, \mathbf{y}) \approx f_M(\mathbf{x}, \mathbf{y}) = \sum_{\ell=0}^{M} \alpha_\ell \varphi_\ell(\mathbf{x}) \psi_\ell(\mathbf{y}), \tag{3.1}$$

with certain coefficients $\alpha_\ell \in \mathbb{R}$ and normalized functions $\varphi_\ell \in L^2(\Omega_1)$ and $\psi_\ell \in L^2(\Omega_2)$. Such an approximation is called a *low-rank approximation*.

It is well known (see, e.g., Lòeve, 1978; Ghanem & Spanos, 1991 or Schwab & Todor, 2006) that, with respect to the number M of terms, the best possible representation of a function $f \in L^2(\Omega_1 \times \Omega_2)$ in the L^2 sense is given by the Karhunen–Lòeve/singular value decomposition. Then, $\alpha_\ell = \sqrt{\lambda_\ell}$ are the eigenvalues of the below-defined integral operator (3.2) with kernel (3.3). The decay of the eigenvalues is important for the fast convergence (in terms of M) of the series (3.1). As we will see in Section 3.2, the decay depends on the smoothness of the function f to be approximated. We analyse the truncation error in Section 3.3 and discuss the smoothness of the eigenfunctions in Section 3.4. In Section 3.5, we finally consider the numerical treatment of (3.1). Besides determining the coefficients $\{\alpha_\ell\}_{\ell \in \mathbb{N}}$, a numerical scheme needs to approximate the functions $\{\varphi_\ell\}_{\ell \in \mathbb{N}}$ and $\{\psi_\ell\}_{\ell \in \mathbb{N}}$ in appropriate trial spaces $V_{j_1}^{(1)}$ and $V_{j_2}^{(2)}$, respectively, up to an accuracy corresponding to that of (3.1). Recall that the trial spaces that we consider are elements of the multiscale analyses (2.5) that have the same approximation order r.

To derive the singular value decomposition, we shall consider the integral operator

$$\mathcal{S} : L^2(\Omega_1) \to L^2(\Omega_2), \quad (\mathcal{S}u)(\mathbf{y}) := \int_{\Omega_1} f(\mathbf{x}, \mathbf{y}) u(\mathbf{x}) \, d\mathbf{x}.$$

Its adjoint is

$$\mathcal{S}^\star : L^2(\Omega_2) \to L^2(\Omega_1), \quad (\mathcal{S}^\star u)(\mathbf{x}) := \int_{\Omega_2} f(\mathbf{x}, \mathbf{y}) u(\mathbf{y}) \, d\mathbf{y}.$$

Then, to obtain the low-rank representation (3.1), we need to compute the eigenvalues of the integral operator

$$\mathcal{K} = \mathcal{S}^\star \mathcal{S} : L^2(\Omega_1) \to L^2(\Omega_1), \quad (\mathcal{K}u)(\mathbf{x}) := \int_{\Omega_1} k(\mathbf{x}, \mathbf{x}') u(\mathbf{x}') \, d\mathbf{x}', \tag{3.2}$$

whose kernel function is given by

$$k(\mathbf{x}, \mathbf{x}') = \int_{\Omega_2} f(\mathbf{x}, \mathbf{y}) f(\mathbf{x}', \mathbf{y}) \, d\mathbf{y} \in L^2(\Omega_1 \times \Omega_1). \tag{3.3}$$

This is a Hilbert–Schmidt kernel. Thus, the associated integral operator $\mathcal{K}$ is compact. Moreover, since $\mathcal{K}$ is self-adjoint, there exists a decomposition into eigenpairs $(\lambda_\ell, \varphi_\ell)$ that fulfils

$$\mathcal{K}\varphi_\ell = \lambda_\ell \varphi_\ell, \quad \ell \in \mathbb{N},$$

with non-negative eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \to 0$ and eigenfunctions $\{\varphi_\ell\}_{\ell \in \mathbb{N}}$, which constitute an orthonormal basis in $L^2(\Omega_1)$.

We now define, for all $\ell \in \mathbb{N}$ with $\lambda_\ell > 0$, the function $\psi_\ell \in L^2(\Omega_2)$ by

$$\psi_\ell(\mathbf{y}) = \frac{1}{\sqrt{\lambda_\ell}} (\mathcal{S}\varphi_\ell)(\mathbf{y}) = \frac{1}{\sqrt{\lambda_\ell}} \int_{\Omega_1} f(\mathbf{x}, \mathbf{y}) \varphi_\ell(\mathbf{x}) \, d\mathbf{x}. \tag{3.4}$$

This constitutes a second sequence of orthonormal functions since

$$(\psi_k, \psi_\ell)_{L^2(\Omega_2)} = \frac{1}{\sqrt{\lambda_k \lambda_\ell}} (\mathcal{S}\varphi_k, \mathcal{S}\varphi_\ell)_{L^2(\Omega_2)} = \frac{1}{\sqrt{\lambda_k \lambda_\ell}} (\mathcal{K}\varphi_k, \varphi_\ell)_{L^2(\Omega_1)}$$

$$= \frac{\lambda_k}{\sqrt{\lambda_k \lambda_\ell}} (\varphi_k, \varphi_\ell)_{L^2(\Omega_1)} = \delta_{k,\ell}.$$

If $\lambda_\ell = 0$ for some $\ell \in \mathbb{N}$, we can extend this collection of functions properly to obtain an orthonormal basis $\{\psi_\ell\}_{\ell \in \mathbb{N}}$ of $L^2(\Omega_2)$. Since

$$\sqrt{\lambda_\ell}\varphi_\ell(\mathbf{x}) = \frac{1}{\sqrt{\lambda_\ell}}(\mathcal{S}^\star \mathcal{S}\varphi_\ell)(\mathbf{x}) = (\mathcal{S}^\star \psi_\ell)(\mathbf{x}) = \int_{\Omega_2} f(\mathbf{x}, \mathbf{z})\psi_\ell(\mathbf{z})\,\mathrm{d}\mathbf{z} \qquad (3.5)$$

for all $\mathbf{x} \in \Omega_1$ and $\ell \in \mathbb{N}$, we finally obtain the representation

$$f(\mathbf{x}, \mathbf{y}) = \sum_{\ell=0}^{\infty} \sqrt{\lambda_\ell}\varphi_\ell(\mathbf{x})\psi_\ell(\mathbf{y}). \qquad (3.6)$$

With (3.4) and (3.5), this equation is easily verified by testing with the orthonormal basis $\{\varphi_k \otimes \psi_\ell\}_{k,\ell \in \mathbb{N}}$ of $L^2(\Omega_1 \times \Omega_2)$.

REMARK 3.1 The adjoint kernel $\tilde{k}(\cdot, \cdot)$ is obtained just by interchanging Ω_1 and Ω_2, i.e.,

$$\tilde{k}(\mathbf{y}, \mathbf{y}') = \int_{\Omega_1} f(\mathbf{x}, \mathbf{y})f(\mathbf{x}, \mathbf{y}')\,\mathrm{d}\mathbf{x} \in L^2(\Omega_2 \times \Omega_2).$$

Then, one has the integral operator

$$\tilde{\mathcal{K}} = \mathcal{S}\mathcal{S}^\star : L^2(\Omega_2) \to L^2(\Omega_2), \quad (\tilde{\mathcal{K}}u)(\mathbf{y}) := \int_{\Omega_2} \tilde{k}(\mathbf{y}, \mathbf{y}')u(\mathbf{y}')\,\mathrm{d}\mathbf{y}'.$$

Again there exists a decomposition into eigenpairs,

$$\tilde{\mathcal{K}}\tilde{\varphi}_\ell = \tilde{\lambda}_\ell \tilde{\varphi}_\ell, \quad \ell \in \mathbb{N},$$

with non-negative eigenvalues $\tilde{\lambda}_1 \geqslant \tilde{\lambda}_2 \geqslant \cdots \geqslant \tilde{\lambda}_m \to 0$ and eigenfunctions $\tilde{\varphi}_\ell \in L^2(\Omega_2)$. We also obtain a second sequence of orthonormal functions $\tilde{\psi}_\ell \in L^2(\Omega_1)$, analogous to (3.4). The functions $\{\tilde{\varphi}_\ell\}_{\ell \in \mathbb{N}}$ and $\{\tilde{\psi}_\ell\}_{\ell \in \mathbb{N}}$ will be the same as before but now their roles are exchanged. Moreover, the eigenvalues λ_ℓ and $\tilde{\lambda}_\ell$ of $\mathcal{K}$ and $\tilde{\mathcal{K}}$, respectively, coincide.

We will prove the following auxiliary result concerning the mapping properties of the integral operators $\mathcal{S}$ and $\mathcal{S}^\star$. To this end, for $s > 0$, we mean by $H^{-s}(\Omega) := (H^s(\Omega))'$ the dual of the Sobolev space $H^s(\Omega)$ (which is usually denoted by $\tilde{H}^{-s}(\Omega)$). Moreover, for $s_1, s_2 \in \mathbb{R}$, let

$$H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2) := H^{s_1}(\Omega_1) \otimes H^{s_2}(\Omega_2)$$

denote the Sobolev space of dominating mixed derivatives.

LEMMA 3.2 Assume that $f \in H^p(\Omega_1 \times \Omega_2)$. Then, the operators

$$\mathcal{S} : H^{-s}(\Omega_1) \to H^{p-s}(\Omega_2), \quad \mathcal{S}^\star : H^{-s}(\Omega_2) \to H^{p-s}(\Omega_1)$$

are continuous for all $s \in [0, p]$.

Proof. From $H^p(\Omega_1 \times \Omega_2) \subset H_{\text{mix}}^{0,p}(\Omega_1 \times \Omega_2)$ it follows that $f \in H_{\text{mix}}^{0,p}(\Omega_1 \times \Omega_2)$. Therefore, the operator $\mathcal{S} : L^2(\Omega_1) \to H^p(\Omega_2)$ is continuous since

$$
\begin{aligned}
\|\mathcal{S}u\|_{H^p(\Omega_2)} &= \sup_{\|v\|_{H^{-p}(\Omega_2)}=1} (\mathcal{S}u, v)_{L_2(\Omega_2)} \\
&= \sup_{\|v\|_{H^{-p}(\Omega_2)}=1} (f, u \otimes v)_{L^2(\Omega_1 \times \Omega_2)} \\
&\leqslant \sup_{\|v\|_{H^{-p}(\Omega_2)}=1} \|f\|_{H_{\text{mix}}^{0,p}(\Omega_1 \times \Omega_2)} \|u \otimes v\|_{H_{\text{mix}}^{0,-p}(\Omega_1 \times \Omega_2)} \\
&\sim \|f\|_{H_{\text{mix}}^{0,p}(\Omega_1 \times \Omega_2)} \|u\|_{L^2(\Omega_1)}.
\end{aligned}
$$

Note that we have used here that $H_{\text{mix}}^{0,-p}(\Omega_1 \times \Omega_2) = L^2(\Omega_1) \otimes H^{-p}(\Omega_2)$.

By complete analogy one shows that $\mathcal{S}^\star : L^2(\Omega_2) \to H^p(\Omega_1)$ is continuous, which proves the desired assertion for $s = 0$. By duality one also infers the assertion for $s = p$. The assertion for $s \in (0, p)$ is finally obtained by interpolation. □

3.2 *Decay of the eigenvalues*

With the above lemma, we are now able to determine the decay rate of the eigenvalues of the integral operator $\mathcal{K} = \mathcal{S}^\star \mathcal{S}$ with kernel (3.3).

THEOREM 3.3 Consider $f \in H^p(\Omega_1 \times \Omega_2)$ with associated kernel k from (3.3) and associated integral operator $\mathcal{K}$ from (3.2). Then, the eigenvalues $\{\lambda_\ell\}_{\ell \in \mathbb{N}}$ of $\mathcal{K}$ decay like

$$\lambda_\ell \lesssim \ell^{-2p/\min\{n_1,n_2\}} \quad \text{as } \ell \to \infty. \tag{3.7}$$

Proof. We shall investigate the dependence of the decay of the eigenvalues of the integral operator $\mathcal{K} = \mathcal{S}^\star \mathcal{S}$ on the smoothness p of f. To this end, we introduce *new*[1] finite element spaces $U_N \subset L^2(\Omega_2)$, which consist of N discontinuous, piecewise polynomial functions of total degree $\lceil p \rceil$ on a quasi-uniform triangulation of Ω_2 with mesh width $h_N \sim N^{-1/n_2}$. Then, the Bramble–Hilbert lemma implies the following approximation result for the L^2-orthogonal projection $P_N : L^2(\Omega_2) \to U_N$:

$$\|(I - P_N)w\|_{L^2(\Omega_2)} \leqslant c_p N^{-p/n_2} \|w\|_{H^p(\Omega_2)},$$

provided that $w \in H^p(\Omega_2)$. Then, since $\mathcal{S}^\star P_N \mathcal{S} : L^2(\Omega_1) \to L^2(\Omega_1)$ is an operator of finite rank N, the min–max principle of Courant–Fischer implies

$$
\begin{aligned}
\lambda_{N+1} &= \min_{\substack{V \subset L^2(\Omega_1) \\ \dim V^\perp \leqslant N}} \max_{\substack{u \in V \\ \|u\|_{L^2(\Omega_1)}=1}} (\mathcal{K}u, u)_{L^2(\Omega_1)} \\
&\leqslant \max_{\substack{u \perp \mathrm{img}(\mathcal{S}^\star P_N \mathcal{S}) \\ \|u\|_{L^2(\Omega_1)}=1}} (\mathcal{S}^\star \mathcal{S}u, u)_{L^2(\Omega_1)} \\
&= \max_{\substack{u \perp \mathrm{img}(\mathcal{S}^\star P_N \mathcal{S}) \\ \|u\|_{L^2(\Omega_1)}=1}} (\mathcal{S}^\star (I - P_N)\mathcal{S}u, u)_{L^2(\Omega_1)} \\
&= \max_{\substack{u \perp \mathrm{img}(\mathcal{S}^\star P_N \mathcal{S}) \\ \|u\|_{L^2(\Omega_1)}=1}} ((I - P_N)\mathcal{S}u, (I - P_N)\mathcal{S}u)_{L^2(\Omega_2)} \\
&\leqslant \sup_{\|u\|_{L^2(\Omega_1)}=1} \|(I - P_N)\mathcal{S}u\|_{L^2(\Omega_2)}^2 \\
&\lesssim N^{-2p/n_2} \sup_{\|u\|_{L^2(\Omega_1)}=1} \|\mathcal{S}u\|_{H^p(\Omega_2)}^2.
\end{aligned}
$$

Since $\mathcal{S} : L^2(\Omega_1) \to H^p(\Omega_2)$ is continuous according to Lemma 3.2, we arrive at $\lambda_{N+1} \lesssim N^{-2p/n_2}$. Applying the same arguments to the operator $\tilde{\mathcal{K}}$, one gets the decay rate $\lambda_{N+1} \lesssim N^{-2p/n_1}$. Thus, $\lambda_{N+1} \lesssim N^{-2p/\min\{n_1,n_2\}}$. Finally substituting for $N + 1$ by ℓ yields the desired result. $\qquad\square$

REMARK 3.4 The min–max principle of Courant–Fischer is also used in Schwab & Todor (2006) to estimate the decay of the eigenvalues. There, however, only the regularity of the kernel function $k \in H^{p,p}_{\mathrm{mix}}(\Omega_1 \times \Omega_1)$ in the first variable enters the estimate, which gives the rate

$$\lambda_\ell \lesssim \ell^{-p/\min\{n_1,n_2\}} \quad \text{as } \ell \to \infty.$$

By using the representation $\mathcal{K} = \mathcal{S}^\star \mathcal{S}$, we are able to exploit the kernel's regularity in both variables which doubles the decay rate.

3.3 *Truncation error*

Altogether, if $f \in H^p(\Omega_1 \times \Omega_2)$, then Theorem 3.3 implies that the coefficients $\{\sqrt{\lambda_\ell}\}$ in the expansion (3.6) of f decay like $\sqrt{\lambda_\ell} \lesssim \ell^{-p/\min\{n_1,n_2\}}$. This leads to the following theorem.

---

[1] The present proof relies on an approximation argument. The new finite element spaces $\{U_N\}$ are introduced to obtain the optimal convergence rate with N degrees of freedom.

THEOREM 3.5 Let $f \in H^p(\Omega_1 \times \Omega_2)$ and $p > \min\{n_1, n_2\}/2$. Then it holds that

$$\left\| f - \sum_{\ell=0}^{M} \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)} \lesssim M^{1/2 - (p/\min\{n_1, n_2\})}. \tag{3.8}$$

Proof. Assume, without loss of generality, that $n_1 \leqslant n_2$. Then, owing to $\lambda_\ell \sim \ell^{-2p/n_1}$ from Theorem 3.3, we obtain by the orthonormality of the function sets $\{\phi_\ell\}$ and $\{\psi_\ell\}$ that

$$\left\| f - \sum_{\ell=0}^{M} \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)}^2 = \left\| \sum_{\ell=M+1}^{\infty} \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)}^2$$

$$= \sum_{\ell=M+1}^{\infty} \lambda_\ell \lesssim \sum_{\ell=M+1}^{\infty} \ell^{-2p/n_1}.$$

Since $2p/n_1 > 1$, we can estimate the sum by an integral as

$$\sum_{\ell=M+1}^{\infty} \ell^{-2p/n_1} \leqslant \int_{M}^{\infty} x^{-2p/n_1} \, dx = \frac{M^{1-2p/n_1}}{2p/n_1 - 1},$$

which leads to the desired result (3.8). $\square$

Consequently, to ensure the error bound

$$\left\| f - \sum_{\ell=0}^{M} \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)} \lesssim \varepsilon, \tag{3.9}$$

we need to choose the expansion degree M as

$$M \sim \varepsilon^{2 \min\{n_1, n_2\}/(\min\{n_1, n_2\} - 2p)}. \tag{3.10}$$

REMARK 3.6 The following comments are in order.

1. Regularity in terms of mixed derivatives does not further improve the results. The property $f \in H_{\mathrm{mix}}^{p,p}(\Omega_1 \times \Omega_2)$ again yields the estimate $\lambda_\ell \lesssim \ell^{-2p/\max\{n_1, n_2\}}$ for the eigenvalues of $\mathcal{K}$.

2. The use of the Sobolev regularity might give too low a decay rate (3.7). For example, for the exponential kernel $k(x, y) = \exp(-|x - y|)$ on the unit square, we have $k \in H^{3/2-\delta}((0, 1) \times (0, 1))$ for all $\delta > 0$, but we observe $\sqrt{\lambda_\ell} \sim \ell^{-2}$ instead of $\sqrt{\lambda_\ell} \sim \ell^{-3/2}$; see Fig. 1. Here, we expect that the use of Besov regularity in the approximation argument in the proof of Theorem 3.3 would give the correct decay since it allows for an adaptive refinement at the diagonal $x = y$.
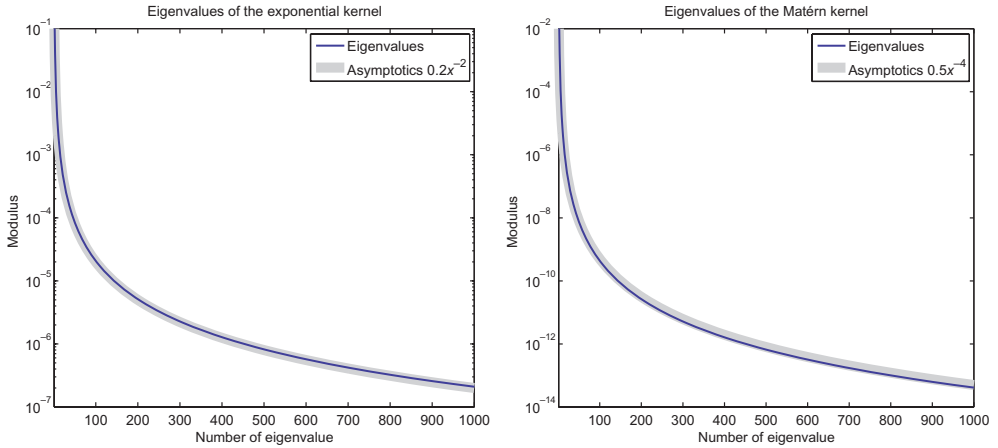
FIG. 1. The eigenvalues of the exponential kernel (left) and the Matérn kernel (right).

3. Our considerations, and thus estimate (3.8), do not apply if $0 \leqslant p \leqslant \min\{n_1, n_2\}/2$. However, it still holds that $\sum_{\ell=0}^{\infty} \lambda_\ell < \infty$ since $\mathcal{K}$ is a Hilbert–Schmidt operator.

3.4 *Smoothness of the eigenfunctions*

Depending on the smoothness of f, we are able to prove the following result on the regularity of the functions in the collections $\{\varphi_\ell\}$ and $\{\psi_\ell\}$. This result will be essential for any numerical computation of the truncated singular value decomposition.

LEMMA 3.7 Let $f \in H^p(\Omega_1 \times \Omega_2)$. Then, the eigenfunctions $\{\varphi_\ell\}$ and $\{\psi_\ell\}$ are in $H^p(\Omega_1)$ and $H^p(\Omega_2)$, respectively, and satisfy

$$\|\varphi_\ell\|_{H^p(\Omega_1)} \lesssim \frac{1}{\sqrt{\lambda_\ell}}, \quad \|\psi_\ell\|_{H^p(\Omega_2)} \lesssim \frac{1}{\sqrt{\lambda_\ell}}, \quad \ell \in \mathbb{N}. \tag{3.11}$$

Proof. In view of (3.5) and Lemma 3.2, we deduce

$$\|\varphi_\ell\|_{H^p(\Omega_1)} = \frac{1}{\sqrt{\lambda_\ell}} \|\mathcal{S}^\star \psi_\ell\|_{H^p(\Omega_1)} \lesssim \frac{1}{\sqrt{\lambda_\ell}} \|\psi_\ell\|_{L^2(\Omega_2)} = \frac{1}{\sqrt{\lambda_\ell}}$$

for all $\ell \in \mathbb{N}$. The second estimate is shown completely analogously. $\square$

3.5 *Numerical approximation*

So far, we have used an *exact* description of the eigenfunctions. However, this does not hold in practice. Instead, the eigenvalues $\{\lambda_\ell\}_{\ell=1}^{M}$ and eigenfunctions $\{\varphi_\ell\}_{\ell=1}^{M}$ and $\{\psi_\ell\}_{\ell=1}^{M}$ need to be approximately computed in the finite element spaces which were introduced in Section 2.

Recall that, on both subdomains, the finite element spaces $V_j^{(i)} \subset L^2(\Omega_i)$ provide the same approximation order r, which results in the same approximation property (2.3), but depending on the subdomains' spatial dimensions n_1, n_2, in different cost complexities (2.4).

According to (2.3) and (3.11), when we use $N_\varphi \sim 2^{jn_1}$ degrees of freedom for an approximation of φ_ℓ in $V_j^{(1)}$ and $N_\psi \sim 2^{jn_2}$ degrees of freedom for an approximation of ψ_ℓ in $V_j^{(2)}$, we conclude

$$
\begin{aligned}
\|\varphi_\ell - \varphi_{\ell,N}\|_{L^2(\Omega_1)} &\lesssim N_\varphi^{-\min\{p,r\}/n_1} \|\varphi_\ell\|_{H^{\min\{p,r\}}(\Omega)} \lesssim \frac{1}{\sqrt{\lambda_\ell}} N_\varphi^{-\min\{p,r\}/n_1}, \\
\|\psi_\ell - \psi_{\ell,N}\|_{L^2(\Omega_2)} &\lesssim N_\psi^{-\min\{p,r\}/n_2} \|\psi_\ell\|_{H^{\min\{p,r\}}(\Omega)} \lesssim \frac{1}{\sqrt{\lambda_\ell}} N_\psi^{-\min\{p,r\}/n_2}.
\end{aligned}
\tag{3.12}
$$

Here, $\varphi_{\ell,N}$ and $\psi_{\ell,N}$ denote the numerical approximations to φ_ℓ and ψ_ℓ, respectively.

According to (3.12), to ensure $\|\varphi_\ell - \varphi_{\ell,N}\|_{L^2(\Omega_1)} \lesssim \varepsilon/\sqrt{\lambda_\ell M}$ (we will later see that this is the proper accuracy), we have to use (cf. (3.10))

$$
N_\varphi \sim \left(\frac{\varepsilon}{\sqrt{M}}\right)^{-n_1/\min\{r,p\}} \sim \varepsilon^{2pn_1/(\min\{n_1,n_2\}-2p)\min\{r,p\}}
$$

unknowns for the representation of $\varphi_{\ell,N}$ and, to ensure $\|\psi_\ell - \psi_{\ell,N}\|_{L^2(\Omega_2)} \lesssim \varepsilon/\sqrt{\lambda_\ell M}$, we have to use

$$
N_\psi \sim \left(\frac{\varepsilon}{\sqrt{M}}\right)^{-n_2/\min\{r,p\}} \sim \varepsilon^{2pn_2/(\min\{n_1,n_2\}-2p)\min\{r,p\}}
$$

unknowns for the representation of $\psi_{\ell,N}$. In what follows, we will always use $N := \max\{N_\varphi, N_\psi\}$ degrees of freedom which does not deteriorate the cost complexity. In particular, N does not depend on ℓ, i.e., all eigenfunctions $\{\varphi_{N,\ell}\}$ and $\{\psi_{N,\ell}\}$ are approximated in the same ansatz spaces.

REMARK 3.8 If $p > r$, then we may even estimate

$$
\|\varphi_\ell\|_{H^r(\Omega)} \lesssim \left(\frac{1}{\sqrt{\lambda_\ell}}\right)^{r/p}, \quad \|\psi_\ell\|_{H^r(\Omega)} \lesssim \left(\frac{1}{\sqrt{\lambda_\ell}}\right)^{r/p}
$$

by using interpolation arguments. Hence, (3.12) can be improved by

$$
\begin{aligned}
\|\varphi_\ell - \varphi_{\ell,N}\|_{L^2(\Omega_1)} &\lesssim \lambda_\ell^{-\min\{r,p\}/2p} N^{-\min\{p,r\}/n_1}, \\
\|\psi_\ell - \psi_{\ell,N}\|_{L^2(\Omega_2)} &\lesssim \lambda_\ell^{-\min\{r,p\}/2p} N^{-\min\{p,r\}/n_2}.
\end{aligned}
$$

As a consequence, if $p > r$, the number of unknowns for approximating the eigenfunctions can be reduced when ℓ increases. However, to exploit this fact for the computation of the truncated singular value decomposition, one needs specific information on the smoothness index p.

We assume that the approximate eigenfunctions are normalized and pairwise orthogonal, i.e., for $1 \leqslant \ell, \ell' \leqslant M$, we have

$$
\int_{\Omega_1} \varphi_{\ell,N}(\mathbf{x}) \varphi_{\ell',N}(\mathbf{x}) \, d\mathbf{x} = \delta_{\ell,\ell'}, \quad \int_{\Omega_2} \psi_{\ell,N}(\mathbf{y}) \psi_{\ell',N}(\mathbf{y}) \, d\mathbf{y} = \delta_{\ell,\ell'}.
$$

With regard to the approximation of the M largest eigenvalues in the space V_N, defined via their Rayleigh quotients

$$\lambda_{\ell,N} = \int_{\Omega_1} \int_{\Omega_1} k(\mathbf{x}, \mathbf{x}') \varphi_{\ell,N}(\mathbf{x}) \varphi_{\ell,N}(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}', \quad \ell = 1, 2, \dots, M,$$

we assume the following estimate:

$$0 \leqslant \lambda_\ell - \lambda_{\ell,N} \lesssim \lambda_\ell \|\varphi_\ell - \varphi_{\ell,N}\|_{L^2(\Omega_1)}^2, \quad \ell = 1, 2, \dots, M. \tag{3.13}$$

We emphasize that these assumptions, in particular (3.13), are all satisfied if one computes the approximation $\{(\lambda_{N,\ell}, \varphi_{N,\ell})\}_{\ell=1}^M$ to the eigenpairs $\{(\lambda_\ell, \varphi_\ell)\}_{\ell=1}^M$ with a Ritz–Galerkin method in the space $V_j^{(1)}$ with $\dim V_j^{(1)} \sim N$, i.e., with $j \sim \log N$. For the details we refer the reader to D'yakonov (1996).

On the basis of (3.8), (3.12) and (3.13), we can now estimate the error of the discretized truncated singular value decomposition.

THEOREM 3.9 Let $f \in H^p(\Omega_1 \times \Omega_2)$ and choose

$$M \sim \varepsilon^{2\min\{n_1,n_2\}/(\min\{n_1,n_2\}-2p)}, \quad N \sim \varepsilon^{2p\max\{n_1,n_2\}/(\min\{n_1,n_2\}-2p)\min\{r,p\}}.$$

Then, the truncated and approximated singular value decomposition satisfies the error estimate

$$\left\| f - \sum_{\ell=0}^M \sqrt{\lambda_{\ell,N}} (\varphi_{\ell,N} \otimes \psi_{\ell,N}) \right\|_{L^2(\Omega_1 \times \Omega_2)} \lesssim \varepsilon,$$

uniformly in $\varepsilon > 0$.

Proof. It holds that

$$
\begin{aligned}
E &:= \left\| \sum_{\ell=0}^\infty \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) - \sum_{\ell=0}^M \sqrt{\lambda_{\ell,N}} (\varphi_{\ell,N} \otimes \psi_{\ell,N}) \right\|_{L^2(\Omega_1 \times \Omega_2)} \\
&\leqslant \left\| \sum_{\ell=M+1}^\infty \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)} + \left\| \sum_{\ell=0}^M \sqrt{\lambda_\ell} (\varphi_\ell \otimes \psi_\ell) - \sqrt{\lambda_{\ell,N}} (\varphi_{\ell,N} \otimes \psi_{\ell,N}) \right\|_{L^2(\Omega_1 \times \Omega_2)}.
\end{aligned}
$$

According to (3.8) and (3.9) the truncation error, i.e., the first term, is bounded by $c\varepsilon$. Moreover, we split the second term into appropriate differences:

$$
\begin{aligned}
\sqrt{\lambda_\ell}(\varphi_\ell \otimes \psi_\ell) - \sqrt{\lambda_{\ell,N}}(\varphi_{\ell,N} \otimes \psi_{\ell,N}) = {}& (\sqrt{\lambda_\ell} - \sqrt{\lambda_{\ell,N}})(\varphi_\ell \otimes \psi_\ell) + \sqrt{\lambda_{\ell,N}}(\varphi_\ell \otimes (\psi_\ell - \psi_{\ell,N})) \\
&+ \sqrt{\lambda_{\ell,N}}((\varphi_\ell - \varphi_{\ell,N}) \otimes \psi_{\ell,N}).
\end{aligned}
$$

Thus, we arrive at

$$E \lesssim \varepsilon + \left\| \sum_{\ell=0}^{M} \left(\sqrt{\lambda_\ell} - \sqrt{\lambda_{\ell,N}} \right) (\varphi_\ell \otimes \psi_\ell) \right\|_{L^2(\Omega_1 \times \Omega_2)} + \left\| \sum_{\ell=0}^{M} \sqrt{\lambda_{\ell,N}} (\varphi_\ell \otimes (\psi_\ell - \psi_{\ell,N})) \right\|_{L^2(\Omega_1 \times \Omega_2)}$$

$$+ \left\| \sum_{\ell=0}^{M} \sqrt{\lambda_{\ell,N}} ((\varphi_\ell - \varphi_{\ell,N}) \otimes \psi_{\ell,N}) \right\|_{L^2(\Omega_1 \times \Omega_2)}.$$

We now estimate the three terms on the right-hand side separately. In view of (3.13), we have $\sqrt{\lambda_\ell} \geqslant \sqrt{\lambda_{\ell,N}}$ and, with (3.12), we obtain

$$0 \leqslant \left| \sqrt{\lambda_\ell} - \sqrt{\lambda_{\ell,N}} \right|^2 \leqslant \lambda_\ell - \lambda_{\ell,N} \lesssim \lambda_\ell \|\varphi_\ell - \varphi_{\ell,N}\|^2_{L^2(\Omega_1)} \lesssim \frac{\varepsilon^2}{M}$$

for all $0 \leqslant \ell \leqslant M$. This yields

$$\left\| \sum_{\ell=0}^{M} \left(\sqrt{\lambda_\ell} - \sqrt{\lambda_{\ell,N}} \right) (\varphi_\ell \otimes \psi_\ell) \right\|^2_{L^2(\Omega_1 \times \Omega_2)} = \sum_{\ell=0}^{M} \left| \sqrt{\lambda_\ell} - \sqrt{\lambda_{\ell,N}} \right|^2 \lesssim \sum_{\ell=0}^{M} \frac{\varepsilon^2}{M} \lesssim \varepsilon^2. \qquad (3.14)$$

Next, with $\lambda_{\ell,N} \leqslant \lambda_\ell$, we have

$$\left\| \sum_{\ell=0}^{M} \sqrt{\lambda_{\ell,N}} (\varphi_\ell \otimes (\psi_\ell - \psi_{\ell,N})) \right\|^2_{L^2(\Omega_1 \times \Omega_2)} = \sum_{\ell=0}^{M} \lambda_{\ell,N} \|\varphi_\ell \otimes (\psi_\ell - \psi_{\ell,N})\|^2_{L^2(\Omega_1 \times \Omega_2)}$$

$$\leqslant \sum_{\ell=0}^{M} \lambda_\ell \|\varphi_\ell\|^2_{L^2(\Omega_1)} \|\psi_\ell - \psi_{\ell,N}\|^2_{L^2(\Omega_2)}$$

$$\lesssim \sum_{\ell=0}^{M} \frac{\varepsilon^2}{M} \lesssim \varepsilon^2, \qquad (3.15)$$

and likewise

$$\left\| \sum_{\ell=0}^{M} \sqrt{\lambda_{\ell,N}} ((\varphi_\ell - \varphi_{\ell,N}) \otimes \psi_{\ell,N}) \right\|^2_{L^2(\Omega_1 \times \Omega_2)} \lesssim \varepsilon^2. \qquad (3.16)$$

Plugging (3.14–3.16) into the above estimate of E finally leads to the desired estimate, $E \lesssim \varepsilon$. $\qquad \square$

Altogether, since we have to deal with M eigenfunctions with N coefficients each, we arrive at the following theorem.

THEOREM 3.10 The number of degrees of freedom needed to approximate a function $f \in H^p(\Omega_1 \times \Omega_2)$ by the singular value decomposition approach (3.1) with (3.12) and (3.13) to a prescribed accuracy ε is

$$\text{dof}_{\text{svd}}(\varepsilon) = M \cdot N \sim \varepsilon^{-(2\min\{r,p\}\min\{n_1,n_2\}+2p\max\{n_1,n_2\})/(2p-\min\{n_1,n_2\})\min\{r,p\}}. \tag{3.17}$$

We emphasize that the estimate (3.17) does not include the work to be spent for *computing* the singular values or the eigenfunctions. Here, a naive approach would result in a cost complexity of order $M \cdot N^2$; the use of fast methods for nonlocal operators would result in an almost linear or even linear complexity per eigenpair. Note that, in any case, at least linear complexity $\mathcal{O}(M \cdot N)$ is required, which is indeed achieved with modern algorithms; see, e.g., Schwab & Todor (2006), Dahmen *et al.* (2008) and Dai *et al.* (2008). Therefore, our analysis is based on the best possible situation for the truncated and approximated singular value decomposition and our later comparison will be fair in this respect.

4. Sparse grids

On the basis of the multiscale analyses (2.5) on each individual subdomain, one naturally obtains a second method to approximate functions in tensor-product spaces: by choosing complementary spaces

$$W_j^{(i)} = \text{span}\{\xi_{j,k}^{(i)} : k \in \nabla_j^{(i)} := \Delta_j^{(i)} \setminus \Delta_{j-1}^{(i)}\}, \quad i = 1, 2$$

such that

$$V_j^{(i)} = W_j^{(i)} \oplus V_{j-1}^{(i)}, \quad V_0^{(i)} = W_0^{(i)},$$

we can define the so-called *general sparse grid space* (see Bungartz & Griebel, 2004 and Griebel & Harbrecht, 2013),

$$\hat{\mathbf{V}}_J^\sigma := \bigoplus_{j_1\sigma + j_2/\sigma \leqslant J} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)}, \tag{4.1}$$

where $\sigma > 0$ is a given parameter. Thus, a function $\hat{f}_J \in \hat{\mathbf{V}}_J^\sigma$ is represented as

$$\hat{f}_J(\mathbf{x}, \mathbf{y}) = \sum_{j_1\sigma + j_2/\sigma \leqslant J} \sum_{k_1 \in \nabla_{j_1}^{(1)}} \sum_{k_2 \in \nabla_{j_2}^{(2)}} \beta_{(j_1,k_1),(j_2,k_2)} \xi_{j_1,k_1}^{(1)}(\mathbf{x}) \xi_{j_2,k_2}^{(2)}(\mathbf{y}). \tag{4.2}$$

Sparse grids can be constructed via hierarchical bases, interpolets and wavelet-like bases (see, e.g., Zenger, 1991; DeVore *et al.*, 1998; Strömberg, 1998; Griebel & Knapek, 2009; Griebel *et al.*, 1990) or even directly by finite elements in terms of frames (see, e.g., Griebel, 1994; Griebel & Oswald, 1994; Harbrecht *et al.*, 2008b). For a survey on sparse grids we refer the reader to Bungartz & Griebel (2004) and the references therein.

The dimension of the general sparse grid space $\hat{\mathbf{V}}_J^\sigma$ is essentially equal to the dimension of the finest univariate finite element spaces which enter its construction, i.e., it is essentially equal to the value of

$\max\{\dim V_{J/\sigma}^{(1)}, \dim V_{J\sigma}^{(2)}\}$. Nevertheless, by considering smoothness in terms of mixed Sobolev spaces, its approximation power is essentially the same as in the full tensor-product space. Precisely, we have the following theorem.

THEOREM 4.1 (Griebel & Harbrecht, 2013) The sparse grid space $\hat{\mathbf{V}}_J^\sigma$ possesses

$$\dim \hat{\mathbf{V}}_J^\sigma \sim \begin{cases} 2^{J \max\{n_1/\sigma, n_2\sigma\}} & \text{if } n_1/\sigma \neq n_2\sigma, \\ 2^{Jn_2\sigma} J & \text{if } n_1/\sigma = n_2\sigma \end{cases}$$

degrees of freedom. Moreover, for a given function $f \in H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2)$ with $0 < s_1, s_2 \leqslant r$, there holds the approximation estimate

$$\inf_{\hat{f}_J \in \hat{\mathbf{V}}_J^\sigma} \|f - \hat{f}_J\|_{L^2(\Omega_1 \times \Omega_2)} \lesssim \begin{cases} 2^{-J \min\{s_1/\sigma, s_2\sigma\}} \|f\|_{H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2)} & \text{if } s_1/\sigma \neq s_2\sigma, \\ 2^{-Js_1/\sigma} \sqrt{J} \|f\|_{H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2)} & \text{if } s_1/\sigma = s_2\sigma. \end{cases}$$

In this theorem, the convergence rate is given for a function $f \in H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2)$. In the following, however, we are interested in the convergence rate if the smoothness of f is measured in the isotropic Sobolev space $H^p(\Omega_1 \times \Omega_2)$. Since for all $s_1 + s_2 = p$ it holds that $H_{\text{mix}}^{s_1,s_2}(\Omega_1 \times \Omega_2) \subset H^p(\Omega_1 \times \Omega_2)$, we have to balance s_1 and s_2 carefully if we want to compute the convergence rate of an $f \in H^p(\Omega_1 \times \Omega_2)$ in the sparse grid space $\hat{\mathbf{V}}_J^\sigma$. The analysis given in Griebel & Harbrecht (2013) shows that the best space for approximating functions in isotropic Sobolev spaces is obtained for the choice $\sigma = \sqrt{n_1/n_2}$. With this choice, the cost complexity to approximate a function $f \in H^p(\Omega_1 \times \Omega_2)$ is given as follows.

THEOREM 4.2 (Griebel & Harbrecht, 2013) The number of degrees of freedom needed to approximate a function $f \in H^p(\Omega_1 \times \Omega_2)$ in the sparse grid space $\hat{\mathbf{V}}_J^\sigma$ with $\sigma = \sqrt{n_1/n_2}$ to a prescribed accuracy ε is essentially

$$\text{dof}_{\text{sg}}(\varepsilon) = N \sim \varepsilon^{-\max\{(n_1+n_2)/p, n_1/r, n_2/r\}}. \tag{4.3}$$

REMARK 4.3 The sparse grid space $\hat{\mathbf{V}}_J^\sigma$ has substantially fewer unknowns than the corresponding full tensor-product space

$$\mathbf{V}_J^\sigma := \bigoplus_{j_1\sigma, j_2/\sigma \leqslant J} W_{j_1}^{(1)} \otimes W_{j_2}^{(2)}.$$

In this space, a function $f \in H^p(\Omega_1 \times \Omega_2)$ is approximated by the rate

$$\inf_{f_J \in \mathbf{V}_J^\sigma} \|f - f_J\|_{L^2(\Omega_1 \times \Omega_2)} \lesssim \{2^{-J \min\{p,r\}/\sigma} + 2^{-J \min\{p,r\}\sigma}\} \|f\|_{H^{\min\{p,r\}}(\Omega_1 \times \Omega_2)}$$

at the cost $\dim \mathbf{V}_J^\sigma \sim 2^{J(n_1/\sigma + n_2\sigma)}$. The choice for $\sigma = 1$ yields thus the cost complexity $N^{-\min\{p,r\}/(n_1+n_2)}$, i.e., $N \lesssim \varepsilon^{-(n_1+n_2)/\min\{p,r\}}$, which is known to be Kolmogorov's n-width for Sobolev balls in the space $H^{\min\{p,r\}}(\Omega_1 \times \Omega_2)$; see Kolmogorov (1936). Indeed, up to logarithmic terms, the sparse grid space $\hat{\mathbf{V}}_J^\sigma$ with $\sigma = \sqrt{n_1/n_2}$ also achieves this rate. Moreover, in the case $(n_1 + n_2)/p < \max\{n_1, n_2\}/r$, it achieves an even higher rate.

5. Comparison of the two approximations

Now, for a given and fixed approximation order r and spatial dimensions n_1 and n_2, we will compare the two different approximation schemes when approximating a function $f \in H^p(\Omega_1 \times \Omega_2)$. We are interested to know for which values of p the sparse grid approach is asymptotically superior to the truncated and approximated singular value decomposition. To this end, we shall distinguish three regimes of the smoothness parameter p where one should have in mind the condition $p \geqslant \min\{n_1, n_2\}$. The subsequent discussion has to be carefully interpreted since we deal with upper bounds only; cf. Remark 3.6.

5.1 *The case $p \geqslant r((n_1 + n_2)/\max\{n_1, n_2\})$*

In this case, according to Theorem 4.2, it holds that the sparse grid approach produces essentially the highest possible approximation rate $N^{-r/\min\{n_1,n_2\}}$, whereas the truncated and approximated singular value decomposition reaches this rate only if f is analytical. The latter follows from the fact that $M \sim |\log(\varepsilon)|^{\min\{n_1,n_2\}}$ dominating singular values are needed; see Schwab & Todor (2006). Thus the sparse grid approach is clearly superior.

5.2 *The case $r((n_1 + n_2)/\max\{n_1, n_2\}) > p \geqslant r$*

According to Theorems 3.10 and 4.2, the truncated and approximated singular value decomposition has the complexity

$$\mathrm{dof}_{\mathrm{svd}}(\varepsilon) \sim \varepsilon^{-(2r\min\{n_1,n_2\}+2p\max\{n_1,n_2\})/(2p-\min\{n_1,n_2\})r}$$

and the sparse grid approach has the complexity

$$\mathrm{dof}_{\mathrm{sg}}(\varepsilon) \sim \varepsilon^{-(n_1+n_2)/p}.$$

The sparse grid approximation is asymptotically superior to the truncated and approximated singular value decomposition if $\mathrm{dof}_{\mathrm{sg}}(\varepsilon) \lesssim \mathrm{dof}_{\mathrm{svd}}(\varepsilon)$, which holds if

$$\frac{n_1 + n_2}{p} \leqslant \frac{2r\min\{n_1, n_2\} + 2p\max\{n_1, n_2\}}{(2p - \min\{n_1, n_2\})r}.$$

One readily infers that this inequality is equivalent to

$$0 \leqslant p^2 - rp + r\frac{\min\{n_1, n_2\}(n_1 + n_2)}{2\max\{n_1, n_2\}} =: g(p).$$

The polynomial $g(p)$ might be bounded from below by

$$g(p) \geqslant rp - rp + r\frac{\min\{n_1, n_2\}(n_1 + n_2)}{2\max\{n_1, n_2\}} \geqslant 0$$

due to $p \geqslant r$. Thus, the sparse grid approach exhibits a higher rate of convergence.

5.3   *The case $r > p \geqslant 0$*

Theorems 3.10 and 4.2 lead to the complexity

$$\text{dof}_{\text{svd}}(\varepsilon) \sim \varepsilon^{-2(n_1+n_2)/(2p-\min\{n_1,n_2\})}$$

for the truncated and approximated singular value decomposition and to the complexity

$$\text{dof}_{\text{sg}}(\varepsilon) \sim \varepsilon^{-(n_1+n_2)/p}$$

for the sparse grid approach. Now, it holds that $\text{dof}_{\text{sg}}(\varepsilon) \lesssim \text{dof}_{\text{svd}}(\varepsilon)$ if the inequality

$$\frac{n_1+n_2}{p} \leqslant \frac{2(n_1+n_2)}{2p - \min\{n_1,n_2\}}$$

is fulfilled. As one can easily check, this inequality is indeed always satisfied. Thus, the sparse grid approach again exhibits a higher rate of convergence.

Let us remark here that, in the situation $p \leqslant r$, an approximation in the *full* tensor-product space $\mathbf{V}_J^\sigma$ with $\sigma = \sqrt{n_1/n_2}$ is of the same complexity as the approximation by the sparse grid space $\hat{\mathbf{V}}_J^\sigma$.

Altogether, we clearly see from these three cases that the general sparse grid approach is approximately at least as good as the truncated and approximated singular value decomposition and is better in most situations (i.e., for an f that is less regular).

6. Numerical results

First, we present an analytical example where the sparse grid approach is superior to the truncated and approximated singular value decomposition. Afterwards, we give the results of two numerical experiments for the exponential kernel and Gaussian kernel which are important two-point correlation kernels from stochastic applications. In these examples, we thus consider the special situation where f is a symmetric function (i.e., $f(\mathbf{x},\mathbf{y}) = f(\mathbf{y},\mathbf{x})$ and $\Omega := \Omega_1 = \Omega_2 \subset \mathbb{R}^n$ with $n := n_1 = n_2$). Then, the singular value decomposition is simply the spectral decomposition.

6.1   *An analytic example*

Let f be defined as

$$f(\mathbf{x},\mathbf{y}) = \sum_{j=0}^{\infty} \sum_{k \in \nabla_j} a_{j,k} \psi_{j,k}(\mathbf{x}) \psi_{j,k}(\mathbf{y}),$$

where $\{\psi_{j,k}\}$ denotes an L^2-normalized orthogonal wavelet basis on the domain Ω. Owing to the inverse estimate we have

$$\|\psi_{j,k}\|_{H^p(\Omega)} \sim 2^{jp} \quad \text{for all } p < \gamma, \tag{6.1}$$

where γ denotes the regularity of the wavelets. Consequently, the function f is in $H^p(\Omega \times \Omega)$ with $0 \leqslant p < \gamma$, provided that

$$\sum_{j=0}^{\infty} \sum_{k \in \nabla_j} 2^{2jp} a_{j,k}^2 < \infty. \tag{6.2}$$

The kernel (3.3) of the integral operator $\mathcal{K}$ is given by

$$k(\mathbf{x}, \mathbf{x}') = \sum_{j,j'=0}^{\infty} \sum_{k \in \nabla_j} \sum_{k' \in \nabla_{j'}} a_{j,k} a_{j',k'} \psi_{j,k}(\mathbf{x}) \left(\int_{\Omega} \psi_{j,k}(\mathbf{y}) \psi_{j',k'}(\mathbf{y}) \, d\mathbf{y} \right) \psi_{j',k'}(\mathbf{x}')$$

$$= \sum_{j=0}^{\infty} \sum_{k \in \nabla_j} a_{j,k}^2 \psi_{j,k}(\mathbf{x}) \psi_{j,k}(\mathbf{x}').$$

Hence, the associated eigenvalues and eigenfunctions of $\mathcal{K}$ are $\{a_{j,k}^2\}$ and $\{\psi_{j,k}\}$, respectively.

Estimate (6.2) implies

$$2^{2jp} a_{j,k}^2 < c 2^{-jn}, \quad k \in \nabla_j, \ j \in \mathbb{N}_0,$$

since otherwise we would have

$$\sum_{j=0}^{\infty} \sum_{k \in \nabla_j} 2^{2jp} a_{j,k}^2 \gtrsim \sum_{j=0}^{\infty} 2^{jn} 2^{-jn} = \infty$$

due to $\#(\nabla_j) \sim 2^{jn}$. Consequently, the eigenvalues $\{a_{j,k}^2\}$ of $\mathcal{K}$ decay essentially like

$$a_{j,k}^2 < c 2^{-j(2p+n)}, \quad k \in \nabla_j, \ j \in \mathbb{N}_0,$$

which is faster than

$$a_{j,k}^2 < c 2^{-2jp}, \quad k \in \nabla_j, \ j \in \mathbb{N}_0$$

as predicted by Theorem 3.3. To this end, we exploited

$$2^{-jn(2p/n)} \gtrsim \ell^{-2p/n} \gtrsim 2^{-(j+1)n(2p/n)}.$$

As a consequence, the eigenfunctions oscillate less strongly than shown in Lemma 3.7. In fact, we have (6.1) instead of $\|\psi_{j,k}\|_{H^p(\Omega)} \lesssim 1/\sqrt{\lambda_{j,k}} \sim 2^{j(p+n/2)}$ (cf. (3.11)). However, by choosing $p \to \gamma$, we see that the estimates in Section 3 related to the eigenvalues and eigenfunctions are sharp except for the factor $n/2$ in the exponent.

According to (6.2), we deduce

$$\sum_{j=J}^{\infty}\sum_{k\in\nabla_j}a_{j,k}^2 \lesssim 2^{-2Jp}\sum_{j=J}^{\infty}\sum_{k\in\nabla_j}2^{2Jp}a_{j,k}^2 \lesssim 2^{-2Jp}\sum_{j=0}^{\infty}\sum_{k\in\nabla_j}2^{2jp}a_{j,k}^2 \lesssim 2^{-2Jp},$$

which implies that the approximated singular value decomposition based on the first 2^{Jn} singular values approximates f with the error $\varepsilon_J \sim 2^{-Jp}$ in the L^2 norm.

We apply properly chosen ansatz functions of approximation power r to approximate the continuous eigenfunctions $\{\psi_{j,k}\}$ by their discrete counterparts $\{\psi_{j,k,N}\}$. Thus, in view of (2.3) and by using (6.1) with $p = \min\{\gamma, r\}$, we have

$$\|\psi_{j,k} - \psi_{j,k,N}\|_{L^2(\Omega)} \lesssim N_j^{-\min\{\gamma,r\}/n}2^{j\min\{\gamma,r\}}.$$

Note that we use here an ansatz which using the optimal number of unknowns to approximate the eigenfunction $\{\psi_{j,k}\}$, i.e., we employ N_j degrees of freedom to approximate the function $\{\psi_{j,k}\}$ instead of a fixed number N which is identical for all eigenfunctions. This is different from Section 3 and even favours the truncated and approximated singular value decomposition over the sparse grid approach; cf. Remark 3.8.

To ensure the necessary accuracy $\varepsilon_J / \sqrt{M_J \lambda_{j,k}}$, we have to balance

$$\|\psi_{j,k} - \psi_{j,k,N}\|_{L^2(\Omega)} \overset{!}{\sim} \frac{\varepsilon_J}{\sqrt{M_J \lambda_{j,k}}} \sim \frac{2^{-Jp}}{2^{Jn/2}2^{-j(p+n/2)}};$$

cf. Section 3. To this end, we essentially have to use

$$N_j \sim \left(\frac{2^{-(J-j)p}}{2^{(J-j)n/2}2^{j\min\{\gamma,r\}}} \right)^{-n/\min\{\gamma,r\}} \sim 2^{(J-j)(p+n/2)n/\min\{\gamma,r\}}2^{jn}$$

unknowns. Therefore, to represent the first 2^{Jn} eigenfunctions, we need at least (up to a constant)

$$\sum_{j=0}^{J} N_j 2^{jn} = \sum_{j=0}^{J} 2^{(J-j)(p+n/2)n/\min\{\gamma,r\}}2^{2jn} = 2^{2Jn}\sum_{j=0}^{J} 2^{(J-j)(p+n/2-2\min\{\gamma,r\})n/\min\{\gamma,r\}} \tag{6.3}$$

degrees of freedom. Altogether, for our special example, this shows that the truncated and approximated singular value decomposition involves the cost complexity (6.3) to obtain the error accuracy $\varepsilon_J \sim 2^{-Jp}$.

In contrast to that, the general sparse grid approach with $\sigma = \sqrt{n_1/n_2} = 1$ produces essentially the same error rate $\varepsilon_J \sim 2^{-Jp}$ with 2^{2Jn} unknowns. If $p + n/2 > 2\min\{\gamma, r\}$, the cost complexity (6.3) is significantly larger than 2^{2Jn}. For example, choosing $p \to \gamma$ and $r > \gamma$, we obtain

$$n > 2\gamma,$$

which describes a situation where the sparse grid approach is asymptotically superior to the corresponding truncated and approximated singular value decomposition. Otherwise, the truncated and approximated singular value decomposition is superior to the corresponding sparse grid approach.
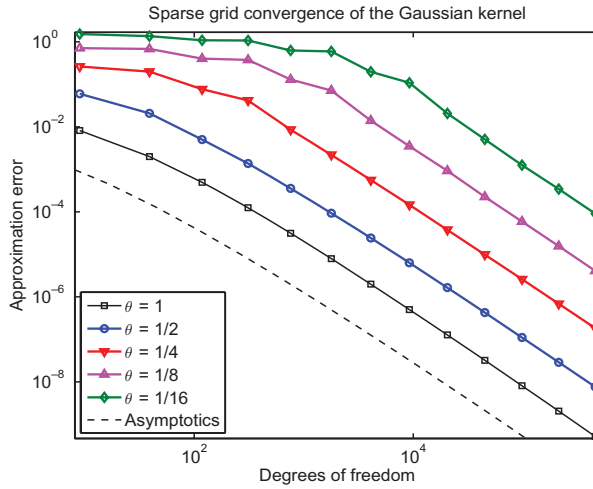
FIG. 2. The L^2 error of the sparse grid approximation of the Gaussian kernel converges essentially like N^{-2}.

In the case of the truncated and approximated singular value decomposition, we have two contradictory effects. On the one hand, the eigenfunctions can be arbitrarily smooth as, for example, in the case of Fourier series. Then, the smoothness is limited by the decay rate of the singular values. On the other hand, when the eigenfunctions are not smooth, also the eigenvalue decay can be arbitrarily fast. A simple example is functions of finite rank but with eigenfunctions of low regularity as in the above example. Both properties affect the cost complexity of the truncated and approximated singular value decomposition differently. Compared with this, the convergence of the general sparse grid approach rather depends on the smoothness of the function to be approximated.

6.2 *Gaussian kernel*

In the second example let Ω be the unit interval $(0, 1)$ (i.e., $n = 1$). On level j, we subdivide $(0, 1)$ into 2^j intervals of length 2^{-j}, which leads to $2^j + 1$ ansatz functions. The approximation spaces V_j under consideration are then generated by continuous piecewise linear polynomials on an equidistant partition of the interval $(0, 1)$ (i.e., $r = 2$).

We discuss the approximation of the Gaussian kernel

$$f(x, y) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{|x - y|^2}{\theta^2}\right), \quad \theta > 0.$$

Since f is known to be analytical and thus arbitrarily smooth, in particular, $f \in H_{\mathrm{mix}}^{2,2}((0, 1) \times (0, 1))$, the general sparse grid approach with $\sigma = \sqrt{n_1/n_2} = 1$ converges with the optimal rate $h_j^2\sqrt{j} \sim 4^{-j}\sqrt{j}$. This is also observed by our numerical results; see Fig. 2. Of course, the smaller the value of θ, the larger becomes the constant which appears in front of the complexity estimate.
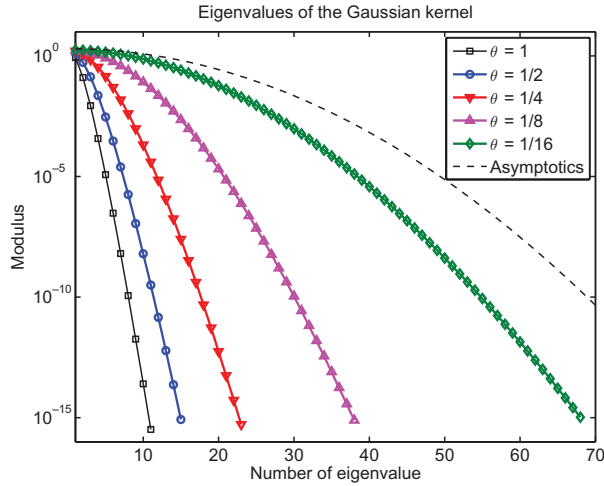
FIG. 3. The eigenvalues of the Gaussian kernel decay, i.e., a decay like $\exp(x^{-2})$ as $x \to \infty$.

According to Schwab & Todor (2006), the singular values of an analytical kernel decay exponentially (if $n_1 = n_2 = 1$). In fact, in the case of the Gaussian kernel the eigenvalues decay even double exponentially (cf. Fig. 3) where the decay is faster the larger θ is. Since the eigenfunctions are also analytical, it holds that

$$\|\varphi_\ell\|_{H^2(0,1)} \lesssim \lambda_\ell^{-s} \quad \text{for any } s > 0.$$

Therefore, it suffices to compute the eigenfunctions with accuracy $\mathcal{O}(h_j^2)$ (see Schwab & Todor, 2006 for the details) which leads to a quadratic rate of convergence with the cost $\mathcal{O}(N \log^{1/2}(N))$. This rate is also validated by Fig. 4. Here, we computed the L^2 error of the truncated and approximated singular value decomposition for all discretization levels $j = 6, 7, \ldots, 15$ and truncation lengths $M = 2, 4, \ldots, 40$. We observe that an increase of the truncation length improves the approximation quality double exponentially until a certain limit is reached which is induced by the approximation error of the discrete eigenfunctions. Conversely, the increase of the discretization level gives a convergence rate $N^{-2} \sim 4^{-j}$ until a certain limit is reached which is induced by the truncation error. We thus obtain the quadratic rate $N^{-2} \sim 4^{-j}$ by proportionally increasing $\sqrt{M}$ and $N = 2^j + 1$, which yields the above predicted cost complexity.

Altogether, both approaches converge in our setting for the Gaussian kernel with essentially the same rate. As seen in Fig. 5, the particular constants which are involved in the estimates of the cost complexities seem to be comparable as well. We plotted the cost complexities of the sparse grid approach (blue lines) and the truncated and approximated singular value decomposition (green lines) for different values of ε and θ. The resulting graphs are quite similar. Moreover, it is observed that the cost complexities increase for both approaches by about the same factors if θ decreases. Note that the slightly different slopes of the graphs result from the fact that the sparse grid approach produces the rate $N^{-2}\sqrt{\log N}$ at the cost $N \log N$, whereas the truncated and approximated singular value decomposition produces the rate N^{-2} at the cost $N\sqrt{\log N}$.
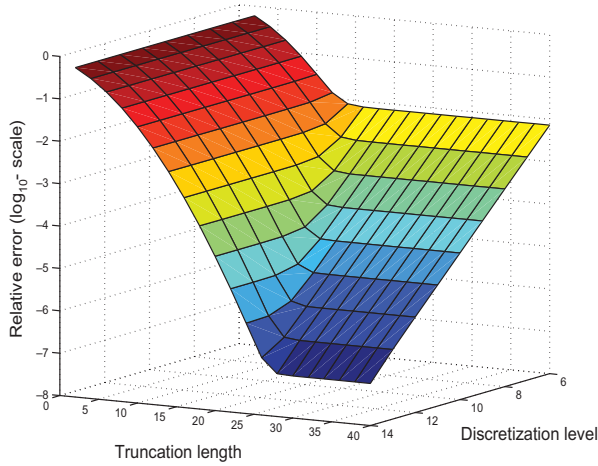
FIG. 4. Truncation length M and discretization level j versus accuracy in the case of the Gaussian kernel.
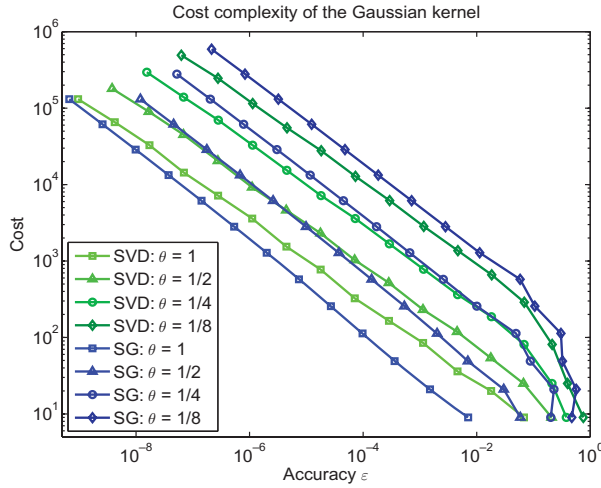

FIG. 5. Comparison of the cost complexities of the sparse grid (SG) and the singular value decomposition (SVD) in the case of the Gaussian kernel.

6.3 *Exponential kernel*

Again, let $\Omega = (0, 1)$ and consider piecewise linear ansatz functions. Our third example is concerned with the approximation of the exponential kernel

$$f(x, y) = \exp(-|x - y|).$$

Since f is only Lipschitz continuous at the diagonal $x = y$, it follows that $f \in H^{3/2-\delta}((0, 1) \times (0, 1))$ for any $\delta > 0$. Therefore, according to Theorem 3.10, we can essentially guarantee the rate $\mathrm{dof}_{svd}(\varepsilon) \sim \varepsilon^{-2}$. Nonetheless, as already mentioned in Remark 3.6, the singular values decay like $\sqrt{\lambda_\ell} \sim \ell^{-2}$, which is
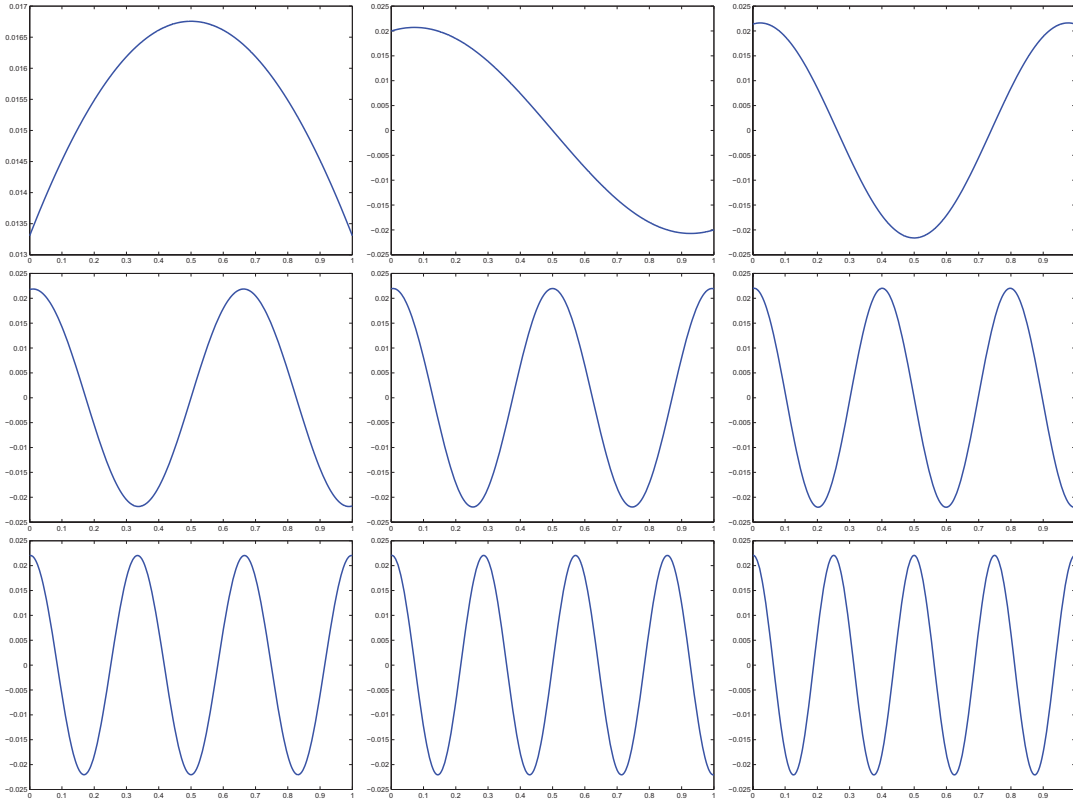
FIG. 6. The first nine eigenfunctions of the exponential kernel.

faster than predicted. In addition, it turned out in our numerical tests that also the eigenfunctions $\{\varphi_\ell\}$ are more regular than expected, i.e., they satisfy $\|\varphi_\ell\|_{H^2(0,1)} \sim 1/\sqrt{\lambda_\ell} \sim \ell^2$. The first nine eigenfunctions are depicted in Fig. 6, which clearly shows the smoothness of the eigenfunctions. By repeating our analysis with these settings (i.e., $p = r = 2$ and $n_1 = n_2 = 1$), one can show that

$$\mathrm{dof}_{\mathrm{svd}}(\varepsilon) \sim \varepsilon^{-4/3}, \tag{6.4}$$

which is much better than predicted. In particular, it is the same rate as produced by an approximation on a uniform grid on $(0,1) \times (0,1)$. This fact is also seen in the left plot of Fig. 7 where we display the L^2 error of the truncated and approximated singular value decomposition for all discretization levels $j = 6, 7, \ldots, 15$ and truncation lengths $M = 2^k$, $k = 0, 1, \ldots, j$. The best cost complexity is offered by choosing $N \sim M \sim \varepsilon^{-2/3}$, which results in the rate (6.4).

On the other hand, in accordance with Theorem 4.2, the sparse grid approach with $\sigma = \sqrt{n_1/n_2} = 1$ realizes (essentially) the same rate: $\mathrm{dof}_{\mathrm{sg}}(\varepsilon) \sim \varepsilon^{-4/3}$. Indeed, this is validated by our computations (see the left plot of Fig. 8).

In the left plot of Fig. 9, we compare the cost complexities of the sparse grid approach (blue line) and the truncated and approximated singular value decomposition (red line). We observe that both
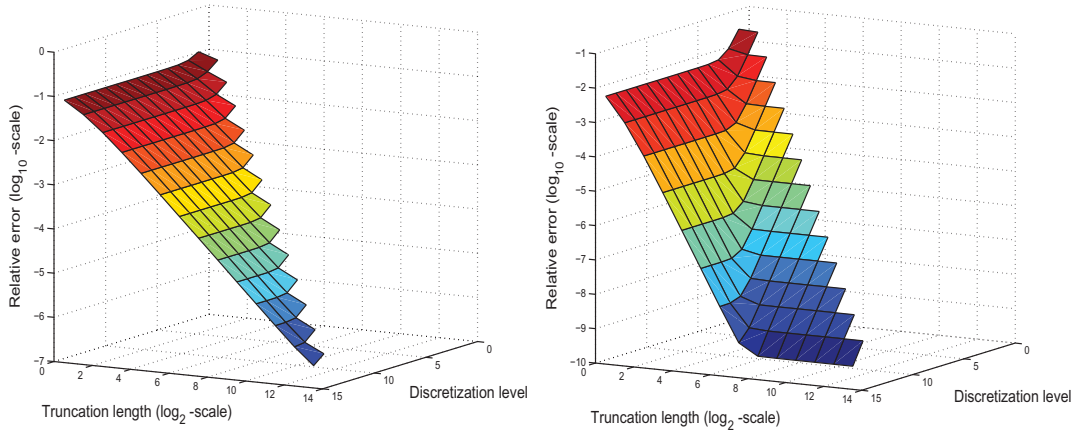
FIG. 7. Truncation length M and discretization level j versus accuracy in the case of the exponential kernel (left) and the Matérn kernel (right).
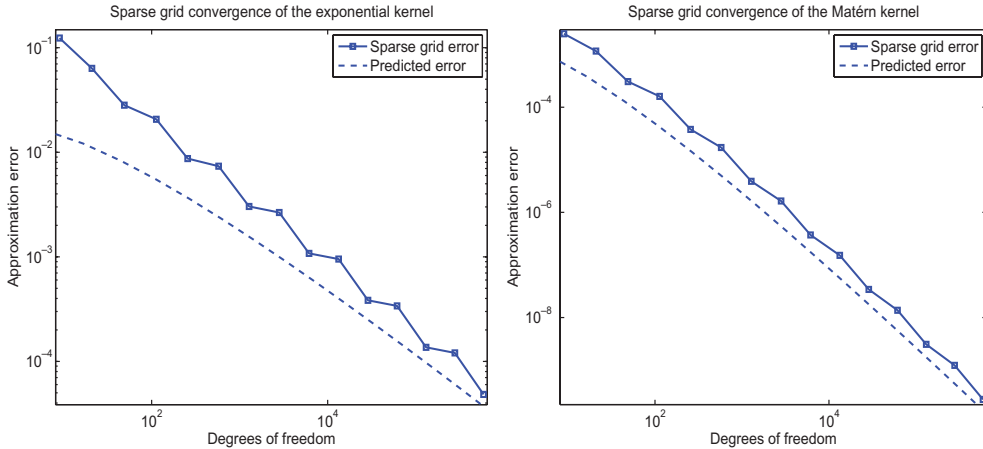

FIG. 8. The sparse grid approximations of the exponential kernel (left) and the Matérn kernel (right) converge essentially like $N^{-3/4}$ and $N^{-7/4}$, respectively.

approaches possess essentially the rate $\varepsilon^{-4/3}$ (dashed black line). The cost complexity of the truncated and approximated singular value decomposition is identical to the cost complexity of the full tensor-product approximation. Hence, it is somewhat better than the cost complexity of the sparse grid approach which involves an additional logarithmic factor.

Note here that only a conventional general sparse grid is involved and no locally adapted sparse grid is used; see, e.g., Bungartz & Griebel (2004) and the references therein. For this approach, we expect a doubling of the convergence rate, i.e., essentially the rate $N^{-3/2}$. The analysis of such a nonlinear approximation scheme is, however, beyond of the scope of this paper.
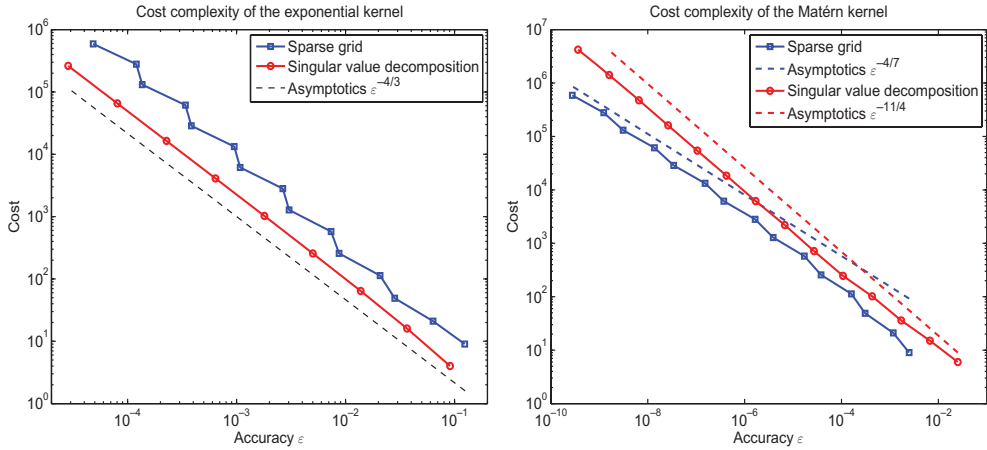
FIG. 9. The cost complexities of the exponential kernel (left) and the Matérn kernel (right).

6.4 *Matérn kernel*

For our last example, we choose the Matérn kernel

$$f(x, y) = (1 + |x - y|) \exp(-|x - y|)$$

and use the same numerical set-up as before. The function f is C^2 smooth with second derivatives which are Lipschitz continuous at the diagonal $x = y$. Hence, it follows that $f \in H^{7/2-\delta}((0,1) \times (0,1))$ for any $\delta > 0$. As can be seen in the right plot of Fig. 1, the singular values decay now like $\sqrt{\lambda_\ell} \sim \ell^{-4}$ which is again, by a factor $1/2$, better than predicted in Theorem 3.3. Moreover, from the right plot of Fig. 7, by fixing the truncation length M, it can be inferred that the eigenfunctions are approximated at the rate N^{-2}. This is slightly better than to be expected from (3.12) but compare here to Remark 3.8. Altogether, for the truncated and approximated singular value decomposition, we get the cost complexity $\varepsilon^{-11/14}$ by balancing $N^{-2} \sim M^{-7/2} \sim \varepsilon$. Compared with this, Theorem 3.10 only gives the rate $\varepsilon^{-4/3}$.

In comparison, the sparse grid approach with N degrees of freedom converges essentially like $N^{-7/4}$; see the right plot of Fig. 8. This is in agreement with Theorem 4.2 which states the cost complexity $\mathrm{dof}_{\mathrm{sg}}(\varepsilon) \sim \varepsilon^{-4/7}$. Consequently, for this example, the sparse grid approach is indeed superior to the truncated and approximated singular value decomposition. This is also confirmed by the right plot of Fig. 9. The cost complexity of the sparse grid approach (blue line) admits a better rate than the cost complexity of the truncated and approximated singular value decomposition (red line).

7. Concluding remarks

In the present paper, we compared the cost complexities of the truncated and approximated singular value decomposition and the general sparse grid approach. We have shown that the sparse grid provides an efficient tool to approximate bi-variate functions. Its cost complexity is at least equal to the truncated and approximated singular value decomposition. In many situations it is even superior.

In the case of the sparse grid approach we envision further improvements by the use of local adaptivity, which would further increase its performance.

In the case of the truncated and approximated singular value decomposition the truncation length is determined by the smoothness of the function under consideration and is thus fixed. Therefore, improvements for the truncated and approximated singular value decomposition can only be achieved by a more efficient representation of the eigenfunctions.

Funding

REFERENCES

BALESCU, R. (1997) *Statistical Dynamics, Matter Out of Equilibrium*. Imperial College, London: Imperial College Press.

BARRETT, J. W., KNEZEVIC, D. & SÜLI, E. (2009) Kinetic models of dilute polymers: analysis, approximation and computation. *11th School on Mathematical Theory in Fluid Mechanics*, 22–29 May 2009, Kacov, Czech Republic. Prague: Necas Center for Mathematical Modeling.

BEBENDORF, M. (2011) Adaptive cross approximation of multivariate functions. *Constr. Approx.*, **34**, 149–179.

BUNGARTZ, H. J. & GRIEBEL, M. (2004) Sparse grids. *Acta Numer.*, **13**, 147–269.

CIORANESCU, D., DAMLAMIAN, A. & GRISO, G. (2008) The periodic unfolding method in homogenization. *SIAM J. Appl. Math.*, **40**, 1585–1620.

DAHMEN, W., ROHWEDDER, T., SCHNEIDER, R. & ZEISER, A. (2008) Adaptive eigenvalue computation: complexity estimates. *Numer. Math.*, **110**, 277–312.

DAI, X., XU, J. & ZHOU, A. (2008) Convergence and optimal complexity of adaptive finite element eigenvalue computations. *Numer. Math.*, **110**, 313–355.

DEVORE, R., KONYAGIN, S. & TEMLYAKOV, V. (1998) Hyperbolic wavelet approximation. *Constr. Approx.*, **14**, 1–26.

D'YAKONOV, E. (1996) *Optimization in Solving Elliptic Problems*. Boca Raton: CRC Press.

GHANEM, R. G. & SPANOS, P. D. (1991) *Stochastic Finite Elements: A Spectral Approach*. Berlin: Springer.

GRASEDYCK, L. (2010) Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.*, **31**, 2029–2054.

GRIEBEL, M. (1994) *Multilevelmethoden als Iterationsverfahren über Erzeugendensystemen.* Teubner Skripten zur Numerik. Stuttgart: B.G. Teubner.

GRIEBEL, M. & HARBRECHT, H. (2013) On the construction of sparse tensor product spaces. *Math. Comput.*, **82**, 975–994.

GRIEBEL, M. & KNAPEK, S. (2009) Optimized general sparse grid approximation spaces for operator equations. *Math. Comput.*, **78**, 2223–2257.

GRIEBEL, M. & OELTZ, D. (2007) A sparse grid space-time discretization scheme for parabolic problems. *Computing*, **81**, 1–34.

GRIEBEL, M. & OSWALD, P. (1994) On additive Schwarz preconditioners for sparse grid discretizations. *Numer. Math.*, **66**, 449–463.

GRIEBEL, M., OSWALD, P. & SCHIEKOFER, T. (1999) Sparse grids for boundary integral equations. *Numer. Math.*, **83**, 279–312.

HACKBUSCH, W. (2012) *Tensor Spaces and Numerical Tensor Calculus*. Berlin–Heidelberg: Springer.

HACKBUSCH, W. & KÜHN, S. (2009) A new scheme for the tensor representation. *J. Fourier Anal. Appl.*, **15**, 706–722.

HARBRECHT, H. (2010) A finite element method for elliptic problems with stochastic input data. *Appl. Numer. Math.*, **60**, 227–244.

HARBRECHT, H., SCHNEIDER, R. & SCHWAB, C. (2008a) Sparse second moment analysis for elliptic problems in stochastic domains. *Numer. Math.*, **109**, 167–188.

HARBRECHT, H., SCHNEIDER, R. & SCHWAB, C. (2008b) Multilevel frames for sparse tensor product spaces. *Numer. Math.*, **110**, 199–220.

HOANG, V. H. & SCHWAB, C. (2005) High-dimensional finite elements for elliptic problems with multiple scales. *SIAM Multiscale Model. Simul.*, **3**, 168–194.

KOLMOGOROV, A. (1936) Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Annals Math.*, **37**, 107–110.

LE BRIS, C. & LELIÈVRE, T. (2009) Multiscale modelling of complex fluids: a mathematical initiation. *Multiscale Modeling and Simulation in Science*. Lecture Notes in Computational Science and Engineering, vol. 66. Berlin: Springer, pp. 49–138.

LÒEVE, M. (1978) *Probability Theory*, vols. I + II. New York: Springer.

LOZINSKI, A., OWENS, R. G. & PHILLIPS, T. N. (2010) The Langevin and Fokker–Planck equations in polymer rheology. *Handbook of Numerical Analysis XVI/XVII* (R. Glowinski ed.). North-Holland: Elsevier.

MATACHE, A. (2002) Sparse two-scale FEM for homogenization problems. *SIAM J. Sci. Comput.*, **17**, 709–720.

OSELEDETS, I. V. & TYRTYSHNIKOV, E. E. (2009) Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM J. Sci. Comput.*, **31**, 3744–3759.

SCHWAB, C. & TODOR, R. A. (2003a) Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.*, **95**, 707–734.

SCHWAB, C. & TODOR, R. A. (2003b) Sparse finite elements for stochastic elliptic problems. Higher order moments. *Computing*, **71**, 43–63.

SCHWAB, C. & TODOR, R. A. (2006) Karhunen–Loéve approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, **217**, 100–122.

STEVENSON, R. & SCHWAB, C. (2009) Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comput.*, **78**, 1293–1318.

STRÖMBERG, J. (1998) Computation with wavelets in higher dimensions. *Proceedings of the International Congress of Mathematicians, Vol. III, Berlin, 1998*. Documenta Mathematica, Extra Vol. III, Universität Bielefeld, Fakultät für Mathematik, pp. 523–532.

WIDMER, G., HIPTMAIR, R. & SCHWAB, C. (2008) Sparse adaptive finite elements for radiative transfer. *J. Comp. Phys.*, **227**, 6071–6105.

ZENGER, C. (1991) Sparse grids. *Parallel Algorithms for Partial Differential Equations. Proceedings of the 6th GAMM-Seminar, Kiel/Germany, 1990*. Notes on Numerical Fluid Mechanics, vol. 31. Braunschweig: Vieweg, pp. 241–251.