# A fast sparse grid based space-time boundary element method for the nonstationary heat equation

H. Harbrecht, J. Tausch

Departement Mathematik und Informatik
Fachbereich Mathematik
Universität Basel
CH-4051 Basel

Preprint No. 2016-32
December 2016

www.math.unibas.ch

# A FAST SPARSE GRID BASED SPACE-TIME BOUNDARY ELEMENT METHOD FOR THE NONSTATIONARY HEAT EQUATION

H. HARBRECHT AND J. TAUSCH

Abstract. This article presents a fast sparse grid based space-time boundary element method for the solution of the nonstationary heat equation. We make an indirect ansatz based on the thermal single layer potential which yields a first kind integral equation. This integral equation is discretized by Galerkin's method with respect to the sparse tensor product of the spatial and temporal ansatz spaces. By employing the $\mathcal{H}$-matrix and Toeplitz structure of the resulting discretized operators, we arrive at an algorithm which computes the approximate solution in a complexity that essentially corresponds to that of the spatial discretization. Nevertheless, the convergence rate is nearly the same as in case of a traditional discretization in full tensor product spaces.

## 1. Introduction

The numerical solution of parabolic evolution problems arises in many applications. In case of the non-stationary heat equation, a boundary reduction by means of boundary integral equations is possible. Provided that the heat equation is homogeneous, only the $n$-dimensional surface $\Gamma := \partial\Omega$ needs to be discretized instead of the spatial domain $\Omega \subset \mathbb{R}^{n+1}$, $n = 1, 2$. If one uses $N^\Gamma$ degrees of freedom for discretizing functions on the surface $\Gamma$ and $N^I$ degrees of freedom for discretizing functions on the time interval $I$, then a traditional Galerkin discretization would have $N^\Gamma \cdot N^I$ degrees of freedom. By "traditional" we mean the discretization of functions on $\Gamma \times I$ in the *full tensor product space*. On the other hand, by using the *sparse tensor product* between the spatial and temporal ansatz space, this number of the degrees of freedom can be considerably reduced to essentially $\max\{N^\Gamma, N^I\}$ degrees of freedom, see e.g. [3, 7, 22]. Here and in the sequel, essentially means that the complexity estimate may be multiplied by (poly-) logarithmic factors. In the context of space-time discretizations, this fact has been exploited in e.g. [8, 17] for finite element methods and in [5] for boundary element methods.

The nonlocality of boundary integral operators results in densely populated system matrices and algorithms that scale at least quadratically in the number of degrees of freedom, unless fast methods are used. Such methods have been developed recently for the layer potentials of the heat equation when using the full tensor product space, see e.g. [18, 19], but for sparse tensor product spaces this is still an open problem.

---

2000 *Mathematics Subject Classification.* 35K20, 65F50, 65M38.

*Key words and phrases.* Boundary element methods, heat equation, sparse grids.

This article presents a fast algorithm which scales essentially linearly in the number of degrees of freedom of the sparse tensor product space. Consequently, we are able to take full advantage of the reduction of the degrees of freedom. For further literature on boundary element methods for sparse grid discretizations, we refer the reader to e.g. [4, 9, 16, 20].

The rest of the article is organized as follows. Section 2 introduces the Dirichlet problem for the heat equation and the indirect boundary integral reformulation using the thermal single layer operator. The traditional Galerkin discretization in full tensor product spaces is discussed in Section 3. The sparse tensor product discretization is then considered in Section 4. In particular, we show that the convergence rate is nearly the same as for the traditional Galerkin discretization provided that the solution offers enough smoothness in terms of Sobolev spaces of dominant mixed derivatives. Section 5 describes the numerical realization of a fast boundary element method which scales essentially linear in the number of unknowns in the sparse tensor product space. One of the key issue that the stiffness matrix is Toeplitz in time. It remains to show that the treatment of the spatial portion of the system matrix can also be applied efficiently. This is the topic of Section 6 while the related error analysis is derived in Section 7. Finally, numerical results obtained with our impementation of the algorithm is presented in Section 8.

## 2. Problem formulation

Let $\Omega \subset \mathbb{R}^{n+1}$, $n = 1, 2$, be a simply connected domain with piecewise smooth boundary $\Gamma := \partial\Omega$ and let $I = (0, T)$ be a time interval for for a given $T > 0$. We consider the following Dirichlet boundary problem for the heat equation: Seek $u \in H^1(\Omega) \otimes L^2(I) \cap H^{-1}(\Omega) \otimes H^1(I)$, such that

$$\partial_t u - \Delta u = 0 \quad \text{in } \Omega \times I \tag{2.1}$$

with boundary condition

$$u = f \quad \text{on } \Gamma \times I \tag{2.2}$$

and initial condition

$$u = 0 \quad \text{on } \Omega \times \{0\}. \tag{2.3}$$

To solve the problem (2.1)–(2.3), we introduce the thermal single layer operator

$$\mathcal{V}g(\mathbf{x}, t) = \int_0^t \int_\Gamma G(\|\mathbf{x} - \mathbf{y}\|, t - \tau) g(\mathbf{y}, \tau) \, \mathrm{d}\sigma_{\mathbf{y}} \, \mathrm{d}\tau \tag{2.4}$$

where $\mathbf{x} \in \Gamma$ and $G(\cdot, \cdot)$ is the heat kernel, given by

$$G(r, t) = \frac{1}{(4\pi t)^{\frac{n+1}{2}}} \exp\left(-\frac{r^2}{4t}\right), \quad t \geq 0 \tag{2.5}$$

and $G(r, t) = 0$ if $t < 0$.

In view of the continuity of the single layer potential operator at the boundary, the ansatz

$$u(\mathbf{x},t) = \int_0^t \int_\Gamma G(\|\mathbf{x}-\mathbf{y}\|, t-\tau) q(\mathbf{y},\tau)\,\mathrm{d}\sigma_{\mathbf{y}}\,\mathrm{d}\tau \tag{2.6}$$

amounts to the boundary integral equation

$$\mathcal{V}q = f \quad \text{on } \Gamma\times I. \tag{2.7}$$

Once (2.7) has been solved for $q$, the solution $u$ of the heat equation (2.1)–(2.3) can be computed for all $(\mathbf{x},t)\in\Omega\times I$ by means of (2.6).

To describe the mapping properties of the boundary integral operator $\mathcal{V}$, let us consider for $r,s\geq 0$ the anisotropic Sobolev spaces of the following form

$$\mathcal{H}^{r,s}(\Gamma\times I) := H^r(\Gamma)\otimes L^2(I)\cap L^2(\Gamma)\otimes H_0^s(I),$$

equipped with the norm

$$\|u\|_{\mathcal{H}^{r,s}(\Gamma\times I)} = \|u\|_{H^r(\Gamma)\otimes L^2(I)} + \|u\|_{L^2(\Gamma)\otimes H^s(I)}.$$

The index 0 indicates that zero initial conditions at $t=0$ are incorporated. Moreover, if $r,s<0$, the space $\mathcal{H}^{r,s}(\Gamma\times I)$ is defined by duality, i.e., $\mathcal{H}^{r,s}(\Gamma\times I) := \big(\mathcal{H}^{-r,-s}(\Gamma\times I)\big)'$. Then, in accordance with [6, 15], the operator $\mathcal{V}$ defines a bilinear form on $\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)$ which is continuous

$$\langle\mathcal{V}p,q\rangle_{L^2(\Gamma\times I)} \lesssim \|p\|_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)}\|q\|_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)} \quad \text{for all } p,q\in\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)$$

and elliptic

$$\langle\mathcal{V}p,p\rangle_{L^2(\Gamma\times I)} \gtrsim \|p\|^2_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)} \quad \text{for all } p\in\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I).$$

Consequently, the boundary integral equation (2.7) is uniquely solvable provided that the right hand side satisfies $f\in\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)$.

## 3. Galerkin discretization

For the Galerkin discretization, we consider two sequences of nested spaces

$$V_0^\Gamma\subset V_1^\Gamma\subset\cdots\subset V_{\ell_s}^\Gamma\subset\cdots\subset L^2(\Gamma), \qquad V_0^I\subset V_1^I\subset\cdots\subset V_{\ell_t}^I\subset\cdots\subset L^2(I).$$

We shall assume that these ansatz spaces are generated by *single-scale bases* $\Phi_{\ell_s}^\Gamma = \{\varphi_{\ell_s,k_s}^\Gamma\}_{k_s\in\Delta_{\ell_s}^\Gamma}$ and $\Phi_{\ell_t}^I = \{\varphi_{\ell_t,k_t}^I\}_{k_t\in\Delta_{\ell_t}^I}$, respectively, that is

$$|\Delta_{\ell_s}^\Gamma| = \dim V_{\ell_s}^\Gamma \sim 2^{\ell_s n}, \quad |\Delta_{\ell_t}^I| = \dim V_{\ell_t}^I \sim 2^{\ell_t}.$$

and

$$V_{\ell_s}^\Gamma = \operatorname{span}\Phi_{\ell_s}^\Gamma, \quad V_{\ell_t}^I = \operatorname{span}\Phi_{\ell_t}^I.$$

We denote the approximation power of the ansatz spaces by $d_s$ and $d_t$, i.e.,

$$\inf_{v_{\ell_s}\in V_{\ell_s}^\Gamma}\|v-v_{\ell_s}\|_{L^2(\Gamma)} \lesssim 2^{-\ell_s d_s}\|v\|_{H^{d_s}(\Gamma)}, \qquad \inf_{v_{\ell_t}\in V_{\ell_t}^I}\|v-v_{\ell_t}\|_{L^2(I)} \lesssim 2^{-\ell_t d_t}\|v\|_{H^{d_t}(I)}.$$

For example, the piecewise constant ($d_s = 1$) or continuous piecewise linear ($d_s = 2$) ansatz functions on a sequence of meshes, obtained by uniform refinement, satisfy our assumptions on the spatial ansatz spaces $V_{\ell_s}^{\Gamma}$.

We shall write $\boldsymbol{L} := (L_s, L_t)$. Then, due to Céa's lemma, a Galerkin scheme for (2.7) in the tensor product space $U_{\boldsymbol{L}}^{\Gamma\times I} := V_{L_s}^{\Gamma} \otimes V_{L_t}^{I}$ leads to the error estimate

$$\|q - q_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac{1}{2},-\frac{1}{4}}(\Gamma\times I)} \lesssim \big(2^{-\frac{L_s}{2}} + 2^{-\frac{L_t}{4}}\big)\big(2^{-L_s d_s} + 2^{-L_t d_t}\big)\|q\|_{\mathcal{H}^{d_s,d_t}(\Gamma\times I)}, \tag{3.8}$$

provided that the boundary $\Gamma$ and the given Dirichlet datum $f$, and thus the solution $q$, are smooth enough, see [6, 15]. As easily seen from (3.8), in case of $d_s = 2d_t$, the optimal choice is $L_t = 2L_s$.

## 4. Sparse tensor product discretization

The tensor product space $U_{\boldsymbol{L}}^{\Gamma\times I} = V_{L_s}^{\Gamma} \otimes V_{L_t}^{I}$ contains $\dim V_{L_s}^{\Gamma} \cdot \dim V_{L_t}^{I} \sim 2^{L_s n} \cdot 2^{L_t}$ degrees of freedom. Compared with this, finite element methods which are based on a sparse grid discretization of the space-time cylinder offer essentially the complexity $\mathcal{O}(2^{L_s(n+1)})$, see e.g. [3, 8, 17] and the references therein. This means, the time discretization comes for free, at least from a complexity point of view. As a consequence, although algorithms are available which solve the heat equation by layer potentials in essentially linear complexity relative to the number of unknowns in the tensor product space $U_{\boldsymbol{L}}^{\Gamma\times I}$ (cf. [13, 14, 18, 19]), there is no gain in the use of boundary integral equations. To overcome this obstruction, as in [5], we shall consider a Galerkin discretization in the *sparse tensor product* of the ansatz spaces $V_{L_s}^{\Gamma}$ and $V_{L_t}^{I}$.

The sparse space-time tensor Galerkin discretization is based on multilevel decompositions of the ansatz spaces. To that end, we set

$$\begin{aligned} W_{\ell_s}^{\Gamma} &:= V_{\ell_s}^{\Gamma} \ominus V_{\ell_s-1}^{\Gamma}, \quad W_{\ell_s}^{\Gamma} = \operatorname{span} \Psi_{\ell_s}^{\Gamma}, \\ W_{\ell_t}^{I} &:= V_{\ell_t}^{I} \ominus V_{\ell_t-1}^{I}, \quad W_{\ell_t}^{I} = \operatorname{span} \Psi_{\ell_t}^{I}. \end{aligned}$$

The basis functions $\Psi_{\ell_s}^{\Gamma} = \{\psi_{\ell_s,k_s}^{\Gamma}\}_{k_s\in\nabla_{\ell_s}^{\Gamma}}$ and $\Psi_{\ell_t}^{I} = \{\psi_{\ell_t,k_t}^{I}\}_{k_t\in\nabla_{\ell_t}^{I}}$ are hierarchical bases or wavelets. Instead of a discretization in the full tensor product space

$$U_{\boldsymbol{L}} := V_{L_s}^{\Gamma} \otimes V_{L_t}^{I} = \bigoplus_{\frac{\ell_s}{L_s},\,\frac{\ell_t}{L_t}\le 1} W_{\ell_s}^{\Gamma} \otimes W_{\ell_t}^{I},$$

we will consider a discretization in the sparse tensor product space

$$\widehat{U}_{\boldsymbol{L}} := \widehat{V_{L_s}^{\Gamma} \otimes V_{L_t}^{I}} = \bigoplus_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\le 1} W_{\ell_s}^{\Gamma} \otimes W_{\ell_t}^{I}. \tag{4.9}$$

The following lemma has been proven in [7, 8]. It states that the time discretization is essentially free provided that $2L_s \gtrsim L_t$.

**Lemma 4.1.** *For $L_s = \sigma L_t \to \infty$, where $\sigma > 0$ is fixed, the sparse tensor product space* (4.9) *satisfies*

$$\dim \widehat{U}_{\boldsymbol{L}} \sim \begin{cases} 2^{L_s n} + 2^{L_t}, & \text{if } L_s n \neq L_t, \\ L_s 2^{L_s n}, & \text{if } L_s n = L_t. \end{cases}$$

On the other hand, the approximation property in the sparse tensor product space is essentially the same as in the full analogue, provided that we spend some extra smoothness in terms of the mixed Sobolev spaces

$$\mathcal{H}^{r,s}_{\mathrm{mix}}(\Gamma \times I) := H^r(\Gamma) \otimes H^s_0(I).$$

In particular, we find the following result for the best approximation in the energy space under consideration.

**Lemma 4.2.** *For $L_s = \sigma L_t \to \infty$, where $\sigma > 0$ is fixed, there holds*

$$\inf_{\widehat{v}_{\boldsymbol{L}} \in \widehat{U}^{\Gamma \times I}_{\boldsymbol{L}}} \|v - \widehat{v}_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac{1}{2},-\frac{1}{4}}(\Gamma \times I)} \lesssim \sqrt{L_s} 2^{-\frac{L_s L_t}{4L_s + 2L_t}} (2^{-L_s d_s} + 2^{-L_t d_t}) \|v\|_{\mathcal{H}^{d_s,d_t}_{\mathrm{mix}}(\Gamma \times I)}$$

*provided that $L_s d_s \neq L_t d_t$. In case of equality, i.e., $L_s d_s = L_t d_t$, an additional logarithmic factor appears:*

$$\inf_{\widehat{v}_{\boldsymbol{L}} \in \widehat{U}^{\Gamma \times I}_{\boldsymbol{L}}} \|v - \widehat{v}_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac{1}{2},-\frac{1}{4}}(\Gamma \times I)} \lesssim L_s 2^{-\frac{L_s L_t}{4L_s + 2L_t}} 2^{-L_s d_s} \|v\|_{\mathcal{H}^{d_s,d_t}_{\mathrm{mix}}(\Gamma \times I)}.$$

*Proof.* The estimates

$$\inf_{\widehat{v}_{\boldsymbol{L}} \in \widehat{U}^{\Gamma \times I}_{\boldsymbol{L}}} \|v - \widehat{v}_{\boldsymbol{L}}\|_{L^2(\Gamma \times I)} \lesssim \begin{cases} \big(2^{-L_s d_s} + 2^{-L_t d_t}\big) \|v\|_{\mathcal{H}^{d_s,d_t}_{\mathrm{mix}}(\Gamma \times I)}, & \text{if } L_s d_s \neq L_t d_t, \\ \sqrt{L_s} 2^{-L_s d_s} \|v\|_{\mathcal{H}^{d_s,d_t}_{\mathrm{mix}}(\Gamma \times I)}, & \text{if } L_s d_s = L_t d_t. \end{cases} \tag{4.10}$$

are shown in [7]. From the definition of anisotropic Sobolev spaces it follows that

$$\mathcal{H}^{\frac{1}{2},\frac{1}{4}}_{\mathrm{mix}}(\Gamma \times I) \subset \mathcal{H}^{\frac{\lambda}{2},\frac{1-\lambda}{4}}_{\mathrm{mix}}(\Gamma \times I) \text{ for all } \lambda \in [0,1]$$

and, therefore,

$$\begin{aligned} \inf_{\widehat{v}_{\boldsymbol{L}} \in \widehat{U}^{\Gamma \times I}_{\boldsymbol{L}}} \|v - \widehat{v}_{\boldsymbol{L}}\|_{L^2(\Gamma \times I)} &\lesssim (2^{-\frac{\lambda}{2} L_s} + 2^{-\frac{1-\lambda}{4} L_t}) \|v\|_{\mathcal{H}^{\frac{\lambda}{2},\frac{1-\lambda}{4}}_{\mathrm{mix}}(\Gamma \times I)} \\ &\lesssim (2^{-\frac{\lambda}{2} L_s} + 2^{-\frac{1-\lambda}{4} L_t}) \|v\|_{\mathcal{H}^{\frac{1}{2},\frac{1}{4}}_{\mathrm{mix}}(\Gamma \times I)} \end{aligned}$$

if $2\lambda L_s \neq (1-\lambda) L_t$. In the case $2\lambda L_s = (1-\lambda) L_t$, which means that

$$\lambda = \frac{L_t}{2L_s + L_t} = \frac{1}{2\sigma + 1},$$

an additional logarithmic logarithmic factor shows up:

$$\begin{aligned} \inf_{\widehat{v}_{\boldsymbol{L}} \in \widehat{U}^{\Gamma \times I}_{\boldsymbol{L}}} \|v - \widehat{v}_{\boldsymbol{L}}\|_{L^2(\Gamma \times I)} &\lesssim \sqrt{L_s} 2^{-\frac{\lambda}{2} L_s} \|v\|_{\mathcal{H}^{\frac{1}{2},\frac{1}{4}}_{\mathrm{mix}}(\Gamma \times I)} \\ &= \sqrt{L_s} 2^{-\frac{L_s L_t}{4L_s + 2L_t}} \|v\|_{\mathcal{H}^{\frac{1}{2},\frac{1}{4}}_{\mathrm{mix}}(\Gamma \times I)}. \end{aligned} \tag{4.11}$$

This is also the best attainable rate since the two terms $2^{-\frac{\lambda}{2}L_s}$ and $2^{-\frac{1-\lambda}{4}L_t}$ are balanced[1]. We shall next denote by $\widehat{\Pi}_{\boldsymbol{L}} : L^2(\Gamma\times I)\to\widehat{U}_{\boldsymbol{L}}$ the $L^2$-orthogonal projection onto the sparse tensor product space $\widehat{U}_{\boldsymbol{L}}$. Then, from

$$\begin{aligned}\inf_{\widehat{v}_{\boldsymbol{L}}\in\widehat{U}_{\boldsymbol{L}}^{\Gamma\times I}}\|v-\widehat{v}_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)} &= \sup_{u\in\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}\frac{\langle v-\widehat{v}_{\boldsymbol{L}},u\rangle_{L^2(\Gamma\times I)}}{\|u\|_{\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}}\\ &\le \sup_{u\in\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}\frac{\langle v-\widehat{\Pi}_{\boldsymbol{L}}v,u-\widehat{\Pi}_{\boldsymbol{L}}u\rangle_{L^2(\Gamma\times I)}}{\|u\|_{\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}}\\ &\le \|v-\widehat{\Pi}_{\boldsymbol{L}}v\|_{L^2(\Gamma\times I)}\sup_{u\in\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}\frac{\|u-\widehat{\Pi}_{\boldsymbol{L}}u\|_{L^2(\Gamma\times I)}}{\|u\|_{\mathcal{H}^{\frac12,\frac14}(\Gamma\times I)}},\end{aligned}$$

we conclude the assertion by inserting the estimates (4.10) and (4.11). □

**Remark 4.3.** *Along the lines of* [5, 6, 7]*, we can determine the best cost complexity of the tensor product approximation and the sparse tensor product approximation, respectively, as* $L_s=\sigma L_t\to\infty$*. If we consider piecewise linear ansatz function in space, i.e.,* $d_s=2$*, and piecewise constant ansatz function in time, i.e.,* $d_t=1$*, we obtain the best cost complexity for the discretization in the tensor product space* $U_{\boldsymbol{L}}$ *for the choice* $L_s=2L_t$*: When using* $N$ *degrees of freedom for the discretization, it follows*

$$\|q-q_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)}\lesssim\begin{cases}N^{-\frac56}\|q\|_{\mathcal{H}^{2,1}(\Gamma\times I)}, & \text{if } n=1,\\ N^{-\frac58}\|q\|_{\mathcal{H}^{2,1}(\Gamma\times I)}, & \text{if } n=2.\end{cases}$$

*Compared with this, the best cost complexity for the Galerkin discretization with respect to the sparse tensor product space* $\widehat{U}_{\boldsymbol{L}}$ *is given by equilibrating the degrees of freedom in* $V_{L_s}^{\Gamma}$ *and* $V_{L_t}^{I}$*. For* $N$ *degrees of freedom, we find then the estimate*

$$\|q-q_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac12,-\frac14}(\Gamma\times I)}\lesssim\begin{cases}N^{-\frac76}(\log N)^{\frac76+\frac12}\|q\|_{\mathcal{H}^{2,1}_{\text{mix}}(\Gamma\times I)}, & \text{if } n=1 \text{ and } L_s=L_t,\\ N^{-\frac98}(\log N)^{\frac98+1}\|q\|_{\mathcal{H}^{2,1}_{\text{mix}}(\Gamma\times I)}, & \text{if } n=2 \text{ and } 2L_s=L_t.\end{cases}$$

*We see that the cost complexity is nearly doubled when using the sparse tensor product discretization in* $n=2$ *dimensions. Moreover, for* $n=1$ *dimensions, the piecewise linear discretization in space does not pay off since the choice* $d_s=1$ *would essentially give the same cost complexity.*

## 5. Algorithms

5.1. **Fast matrix-vector multiplication.** Throughout the article, the basis in $\widehat{U}_{\boldsymbol{L}}$ will be denoted by

$$\widehat{\Psi}_{\boldsymbol{L}}:=\left\{\widehat{\psi}_{\boldsymbol{\ell},\mathbf{k}}=\psi^{\Gamma}_{\ell_s,k_s}\otimes\psi^{I}_{\ell_t,k_t}:\mathbf{k}=(k_s,k_t)\in\nabla_{\boldsymbol{\ell}}:=\nabla^{\Gamma}_{\ell_s}\times\nabla^{I}_{\ell_t},\ \frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\le 1\right\}.$$

[1]By balancing these terms, we obtain an improvement of the results in [5].

Then, the Galerkin matrix $\widehat{\mathbf{V}}_{\boldsymbol{L}} = \langle \widehat{\Psi}_{\boldsymbol{L}}, \widehat{\Psi}_{\boldsymbol{L}} \rangle_{L^2(\Gamma\times I)}$ consists of the block matrices

$$\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'} := \langle V(\Psi^{\Gamma}_{\ell'_s} \otimes \Psi^{I}_{\ell'_t}), \Psi^{\Gamma}_{\ell_s} \otimes \Psi^{I}_{\ell_t} \rangle_{L^2(\Gamma\times I)} \tag{5.12}$$

where $\frac{\ell_s}{L_s} + \frac{\ell_t}{L_t}, \frac{\ell'_s}{L_s} + \frac{\ell'_t}{L_t} \leq 1$. Here, the block $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ has asymptotically the dimension $2^{\ell_s n+\ell_t} \times 2^{\ell'_s n+\ell'_t}$. Obviously, by writing $\widehat{\mathbf{u}}_{\boldsymbol{L}} = [\mathbf{u}_{\boldsymbol{\ell}}]_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\leq 1}$, the matrix-vector multiplication $\widehat{\mathbf{w}}_{\boldsymbol{L}} = \widehat{\mathbf{V}}_{\boldsymbol{L}} \widehat{\mathbf{u}}_{\boldsymbol{L}}$ can be block wise computed by

$$\widehat{\mathbf{w}}_{\boldsymbol{L}} = [\mathbf{w}_{\boldsymbol{\ell}}]_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\leq 1} = \left[ \sum_{\frac{\ell'_s}{L_s}+\frac{\ell'_t}{L_t}\leq 1} \mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'} \mathbf{u}_{\boldsymbol{\ell}'} \right]_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\leq 1} = \widehat{\mathbf{V}}_{\boldsymbol{L}} \widehat{\mathbf{u}}_{\boldsymbol{L}}. \tag{5.13}$$

**Lemma 5.1.** *Assume that the block matrix-vector product $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}\mathbf{u}_{\boldsymbol{\ell}'}$ is computable in complexity $\mathcal{O}\big(M \cdot 2^{\max\{\ell_s n+\ell_t, \ell'_s n+\ell'_t\}}\big)$. Then, the matrix-vector product $\widehat{\mathbf{w}}_{\boldsymbol{L}} = \widehat{\mathbf{V}}_{\boldsymbol{L}} \widehat{\mathbf{u}}_{\boldsymbol{L}}$ is of complexity $\mathcal{O}\big(ML_sL_t \dim(\widehat{U}_{\boldsymbol{L}})\big)$.*

*Proof.* The assertion follows immediately from (5.13) and

$$\begin{aligned}
&\sum_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}, \frac{\ell'_s}{L_s}+\frac{\ell'_t}{L_t}\leq 1} M \cdot 2^{\max\{\ell_s n+\ell_t, \ell'_s n+\ell'_t\}} \\
&\quad = \sum_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\leq 1} M \cdot \Bigg( \sum_{\substack{\frac{\ell'_s}{L_s}+\frac{\ell'_t}{L_t}\leq 1 \\ \ell_s n+\ell_t \leq \ell'_s n+\ell'_t}} 2^{\ell'_s n+\ell'_t} + \sum_{\substack{\frac{\ell'_s}{L_s}+\frac{\ell'_t}{L_t}\leq 1 \\ \ell_s n+\ell_t > \ell'_s n+\ell'_t}} 2^{\ell_s n+\ell_t} \Bigg) \\
&\quad \lesssim \sum_{\frac{\ell_s}{L_s}+\frac{\ell_t}{L_t}\leq 1} M \cdot \Big( \dim(\widehat{U}_{\boldsymbol{L}}) + 2^{\ell_s n+\ell_t} L_s L_t \Big) \\
&\quad \lesssim ML_sL_t \dim(\widehat{U}_{\boldsymbol{L}}).
\end{aligned}$$

□

5.2. **Restrictions and prolongations.** Since it is algorithmically difficult to compute matrices in wavelet coordinates and with ansatz and test functions on different levels, we use restrictions and prolongations to realize matrix vector products with $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ in single-scale spaces.

Because $W^{\Gamma}_{\ell_s} \subset V^{\Gamma}_{\ell'_s}$ for any $\ell_s \leq \ell'_s$, we can represent a given function $u_{\ell_s} \in W^{\Gamma}_{\ell_s}$ in the space $V^{\Gamma}_{\ell'_s}$. Such a prolongation will be denoted by $J^{\ell'_s}_{\ell_s}$. Its discrete counterpart $\mathbf{J}^{\ell'_s}_{\ell_s}$ can obviously be applied to a given vector $\mathbf{u}_{\ell_s}$ in complexity $\mathcal{O}\big(2^{\ell'_s n}\big)$. Vice versa, a function $u_{\ell'_s}$ in $V^{\Gamma}_{\ell'_s}$ can be restricted to the space $W^{\Gamma}_{\ell_s}$ which we denote by $J^{\ell_s}_{\ell'_s}$. The cost of the corresponding discrete operation $\mathbf{J}^{\ell'_s}_{\ell_s}\mathbf{u}_{\ell'_s}$ is of the order $\mathcal{O}\big(2^{\ell'_s n}\big)$. Note that $\big(\mathbf{J}^{\ell_s}_{\ell'_s}\big)^T = \mathbf{J}^{\ell'_s}_{\ell_s}$.

Likewise, due to $W^{I}_{\ell_t} \subset V^{I}_{\ell'_t}$ for any $\ell_t \leq \ell'_t$, corresponding operators $I^{\ell'_t}_{\ell_t}$ and $I^{\ell_t}_{\ell'_t}$ exist with respect to the time. Their discrete counterparts are denoted by $\mathbf{I}^{\ell'_t}_{\ell_t}$ and $\mathbf{I}^{\ell_t}_{\ell'_t}$, where the application to a vector costs $\mathcal{O}\big(2^{\ell'_t}\big)$ operations.

In the following, we will use the notational convention

$$\tilde{\ell}_s := \max\{\ell_s, \ell'_s\} \quad \text{and} \quad \tilde{\ell}_t := \max\{\ell_t, \ell'_t\}.$$

Thus, we obtain the representation in the single-scale spaces

$$\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'} = \Big(\mathbf{I}_{\tilde{\ell}_t}^{\ell_t} \otimes \mathbf{J}_{\tilde{\ell}_s}^{\ell_s}\Big)\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}\Big(\mathbf{I}_{\ell'_t}^{\tilde{\ell}_t} \otimes \mathbf{J}_{\ell'_s}^{\tilde{\ell}_s}\Big) \tag{5.14}$$

where $\tilde{\boldsymbol{\ell}} = (\tilde{\ell}_s, \tilde{\ell}_t)$ and

$$\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}} := \langle V(\Phi_{\tilde{\ell}_s}^{\Gamma} \otimes \Phi_{\tilde{\ell}_t}^{I}), \Phi_{\tilde{\ell}_s}^{\Gamma} \otimes \Phi_{\tilde{\ell}_t}^{I}\rangle_{L^2(\Gamma\times I)} \tag{5.15}$$

**Remark 5.2.** *The dimension of the matrix* $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$ *is asymptotically* $2^{\max\{\ell_t,\ell'_t\}n+\max\{\ell_s,\ell'_s\}}$ *which is, in general, larger than the dimensions of* $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$. *In fact, it turns out that it is not possible to compute a matrix-vector product with* $\widehat{\mathbf{V}}_{\boldsymbol{L}}$ *in the desired* $\mathcal{O}\big(ML_sL_t\dim(\widehat{U}_{\boldsymbol{L}})\big)$ *complexity, if the factors in are evaluated in the sequence suggested by* (5.14), *even if the application of* $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$ *has linear complexity. However, we will show below that* $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$ *can be approximated by a sum of Kronecker products, which will lead to an algorithm with log-linear complexity in* $\dim(\widehat{U}_{\boldsymbol{L}})$.

5.3. **Block matrix-vector multiplication.** To get a guideline for the realization of an essentially optimal block matrix-vector multiplication, let us assume from now on that $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ is approximated by a sum of tensor products

$$\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'} \approx \sum_{i=1}^{M} \mathbf{A}_{\ell_t,\ell'_t}^{(i)} \otimes \mathbf{B}_{\ell_s,\ell'_s}^{(i)}. \tag{5.16}$$

Such a representation is also called *low-rank approximation.* Provided that for all $i = 1,\ldots,M$ the application of the matrices $\mathbf{A}_{\ell_t,\ell'_t}^{(i)}$ and $\mathbf{B}_{\ell_s,\ell'_s}^{(i)}$ to a vector can be evaluated in $\mathcal{O}\big(2^{\max\{\ell_t,\ell'_t\}}\big)$ and $\mathcal{O}\big(2^{\max\{\ell_s,\ell'_s\}n}\big)$ operations, respectively, then the matrix-vector product

$$\mathbf{w}_{\boldsymbol{\ell}} = \mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}\mathbf{u}_{\boldsymbol{\ell}'} \approx \sum_{i=1}^{M} \big(\mathbf{A}_{\ell_t,\ell'_t}^{(i)} \otimes \mathbf{B}_{\ell_s,\ell'_s}^{(i)}\big)\mathbf{u}_{\boldsymbol{\ell}'}$$

is computable within the complexity $\mathcal{O}\big(M \cdot 2^{\max\{\ell_s n+\ell_t,\ell'_s n+\ell'_t\}}\big)$. This is seen as follows.

From the identity

$$\operatorname{vec}(\mathbf{w}_{\boldsymbol{\ell}}^{(i)}) = (\mathbf{A}_{\ell_t,\ell'_t}^{(i)} \otimes \mathbf{B}_{\ell_s,\ell'_s}^{(i)})\operatorname{vec}(\mathbf{u}_{\boldsymbol{\ell}'}) \Longleftrightarrow \mathbf{w}_{\boldsymbol{\ell}}^{(i)} = \mathbf{B}_{\ell_s,\ell'_s}^{(i)}\mathbf{u}_{\boldsymbol{\ell}'}\big(\mathbf{A}_{\ell_t,\ell'_t}^{(i)}\big)^T \tag{5.17}$$

we conclude that, for $\ell_s n + \ell'_t \leq \ell'_s n + \ell_t$, it is cheaper to compute the vector $\mathbf{w}_{\boldsymbol{\ell}}^{(i)}$ in the order

$$\mathbf{z} = \mathbf{B}_{\ell_s,\ell'_s}^{(i)}\mathbf{u}_{\boldsymbol{\ell}'}, \quad \mathbf{w}_{\boldsymbol{\ell}}^{(i)} = \big(\mathbf{A}_{\ell_t,\ell'_t}^{(i)}\mathbf{z}^T\big)^T \tag{5.18}$$

(we refer to Fig. 5.1 for a corresponding visualization). Here, the evaluation of $\mathbf{z}$ is of complexity $\mathcal{O}\big(2^{\ell'_t} \cdot 2^{\max\{\ell_s,\ell'_s\}n}\big)$ and thus the complexity for computing $\mathbf{w}_{\boldsymbol{\ell}}^{(i)}$ via (5.18) is

$$\mathcal{O}\big(2^{\ell'_t} \cdot 2^{\max\{\ell_s,\ell'_s\}n} + 2^{\ell_s n} \cdot 2^{\max\{\ell_t,\ell'_t\}}\big) = \mathcal{O}\big(2^{\max\{\ell_s n+\ell_t,\ell'_s n+\ell'_t,\ell_s n+\ell'_t\}}\big).$$

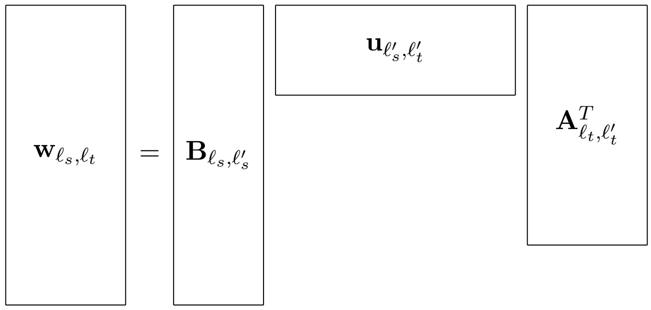


FIGURE 5.1. Visualization of the matrix-vector product: Here, it is cheaper to perform first the multiplication $\mathbf{u}_{\ell_s',\ell_t'}\mathbf{A}^T_{\ell_t,\ell_t'}$ and then the multiplication of the result with $\mathbf{B}_{\ell_s,\ell_s'}$.

Due to the supposition $\ell_s n + \ell_t' \le \ell_s' n + \ell_t$, we have

$$\ell_s n + \ell_t' \le (\ell_s' n + \ell_t') - \ell_t' + (\ell_s n + \ell_t) - \ell_s n$$

and thus

$$2(\ell_s n + \ell_t') \le (\ell_s' n + \ell_t') + (\ell_s n + \ell_t) \le 2\max\{\ell_s n + \ell_t, \ell_s' n + \ell_t'\}.$$

Therefore, the complexity for the matrix-vector multiplication (5.18) is of complexity $\mathcal{O}\big(2^{\max\{\ell_s n+\ell_t,\ell_s' n+\ell_t'\}}\big)$ which is order optimal.

Whereas, if $\ell_s n + \ell_t' > \ell_s' n + \ell_t$, we should compute the matrix product in the order $\mathbf{B}^{(i)}_{\ell_s,\ell_s'}\big(\mathbf{A}^{(i)}_{\ell_t,\ell_t'}\mathbf{u}^T_{\boldsymbol{\ell}'}\big)^T$.

If $\ell_s n + \ell_t' > \ell_s' n + \ell_t$, we change the order of multiplication in (5.17) and compute

$$\mathbf{z} = \mathbf{A}^{(i)}_{\ell_t,\ell_t'}\mathbf{u}^T_{\boldsymbol{\ell}'}, \quad \mathbf{w}^{(i)}_{\boldsymbol{\ell}} = \mathbf{B}^{(i)}_{\ell_s,\ell_s'}\mathbf{z}^T. \tag{5.19}$$

By using arguments analogous to above, one readily infers that the complexity of computing $\mathbf{w}^{(i)}_{\boldsymbol{\ell}}$ via (5.19) is also of order optimal complexity $\mathcal{O}\big(2^{\max\{\ell_s n+\ell_t,\ell_s' n+\ell_t'\}}\big)$.

**Remark 5.3.** *One logarithmic factor in the cost complexity of the matrix-vector product described here can be removed by using the unidirectional principle, see e.g.* [1, 2, 21]. *Nevertheless, we have not exploited this approach for sake of simplicity in representation.*

5.4. **Tensor product representation of $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$.** In this section, we show how to compute the low-rank approximation (5.16) using the factorization in (5.14). To keep the technical level of the discussion at a minimum, we assume that the temporal spaces $V^I_{\ell_t}$ consist of piecewise constant ansatz functions on a uniform subdivision of $I = (0, T)$ into $2^{\ell_t} n_t$ intervals, where $n_t$ is a small integer. Thus the temporal basisfunctions $\varphi^I_{\ell_t,k_t}$ are scaled and translated versions of the box function.

We begin by introducing an $\mathcal{H}$-matrix pattern of the matrix $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$ in time, see Fig. 5.2 for a visualization. Here, the blocks become larger with increasing distance to the diagonal. Specifically, the pattern is obtained by equidistantly subdividing the interval $I = (0, T)$ into $2^m$ sub-intervals $I_{m,k} := 2^{-m}T(k, k+1)$, $k = 0, 2, \ldots, 2^m - 1$, $m = 0, 1, \ldots, \tilde{\ell}_t$. The

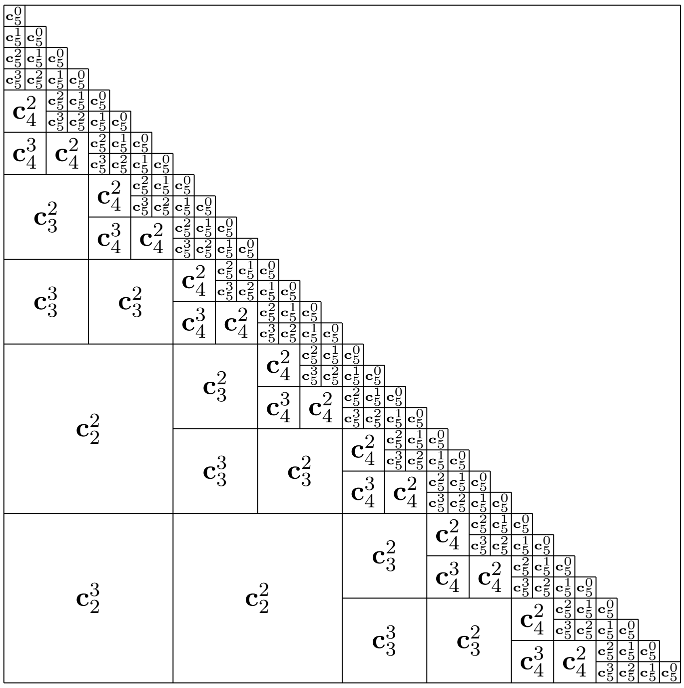

FIGURE 5.2. Partitioning of $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$ for the case that $\tilde{\ell}_t = 5$.

block $I_{m,k} \times I_{m,k'}$ is called *admissible* if $d := k - k' \geq 2$ (mind that $k \geq k'$, because of the causality of the thermal layer potentials). Starting with the coarsest blocks and collecting all blocks, one recursively obtains the pattern shown in Fig. 5.2, see [10, 11] for details.

Here, the blocks $\mathbf{c}_m^d$ are square matrices of size $2^{-m}\left|\Delta_{\tilde{\ell}_t}^I\right| \cdot \left|\Delta_{\tilde{\ell}_s}^\Gamma\right|$ whose components are given by

$$\left[\mathbf{c}_m^d\right]_{(k_s,k_t),(k_s',k_t')} = \int_0^T \int_0^T \left\{ \int_\Gamma \int_\Gamma G(\|\mathbf{x}-\mathbf{y}\|, t-\tau)\Phi_{\tilde{\ell}_s,k_s}^\Gamma(\mathbf{x})\Phi_{\tilde{\ell}_s,k_s'}^\Gamma(\mathbf{y})\,\mathrm{d}\sigma_{\mathbf{y}}\,\mathrm{d}\sigma_{\mathbf{x}} \right\} \times \Phi_{\tilde{\ell}_t,k_t}^I(t)\Phi_{\tilde{\ell}_t,k_t'}^I(\tau)\,\mathrm{d}\tau\,\mathrm{d}t, \tag{5.20}$$

where the functions $\Phi_{\tilde{\ell}_t,k_t}^I$ and $\Phi_{\tilde{\ell}_t,k_t'}^I$ are supported in $I_{m,d}$ and $I_{m,0}$, respectively.

This partitioning suggests to write the $\mathcal{H}$-matrix as a sum of $2\tilde{\ell}_t$ block-Toeplitz matrices that contain the identical blocks. To that end, define the $(2^m \times 2^m)$-matrices

$$\mathbf{H}_m^0 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \quad \mathbf{H}_m^1 = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}, \quad \mathbf{H}_m^2 = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 0 \end{bmatrix} \tag{5.21}$$

and

$$\mathbf{H}_m^3 = \begin{bmatrix} 0 & & & & & & & \\ 0 & 0 & & & & & & \\ 0 & 0 & 0 & & & & & \\ 1 & 0 & 0 & 0 & & & & \\ & 0 & 0 & 0 & 0 & & & \\ & & 1 & 0 & 0 & 0 & & \\ & & & 0 & 0 & 0 & 0 & \\ & & & & \ddots & \ddots & \ddots & \ddots \\ & & & & & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{5.22}$$

Note that the ones and zeros in the third subdiagonal of $\mathbf{H}_m^3$ alternate because of the pattern in which the blocks $\mathbf{c}_m^3$ appear in the matrix $\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}}$.

With these notations, one obtains the decomposition

$$\mathbf{V}_{\tilde{\boldsymbol{\ell}},\tilde{\boldsymbol{\ell}}} = \sum_{d\in\{0,1\}} \mathbf{H}_{\tilde{\ell}_t}^d \otimes \mathbf{c}_{\tilde{\ell}_t}^d \quad + \sum_{\substack{m\in\{2,\dots,\tilde{\ell}_t\}\\ d\in\{2,3\}}} \mathbf{H}_m^d \otimes \mathbf{c}_m^d \tag{5.23}$$

The matrices $\mathbf{c}_{\tilde{\ell}_t}^0$ and $\mathbf{c}_{\tilde{\ell}_t}^1$ contain the temporal near-field and appear only in the finest-level $\tilde{\ell}_t$, whereas the matrices $\mathbf{c}_m^2$ and $\mathbf{c}_m^3$, $m \in \{2, \dots, \tilde{\ell}_t\}$ contain well-separated interactions of the temporal variable.

*Temporal far-field.* Consider the block $\mathbf{c}_m^d$ in the temporal far-field where the ansatz- and test functions $\Phi_{\tilde{\ell}_t,k_t}^I$ and $\Phi_{\tilde{\ell}_t,k_t'}^I$ have support inside $I_{m,d}$ and $I_{m,0}$, respectively. Since $d \in \{2,3\}$, the kernel is smooth and can be well approximated by a degenerate kernel expansion. Such an expansion can be obtained, for instance, by interpolation. This is most conveniently achieved in the local coordinates $t'$, $\tau'$ of the respective intervals

$$t = T2^{-m}(d + t'), \quad \tau = T2^{-m}\tau', \quad 0 \le \tau', t' \le 1, \tag{5.24}$$

Thus, setting $\widetilde{\mathbf{r}} = \mathbf{r}/\sqrt{T2^{-m}}$, it follows that

$$\begin{aligned} G\Big(\|\mathbf{r}\|, t-\tau\Big) &= (T2^{-m})^{-\frac{n+1}{2}} G\Big(\|\widetilde{\mathbf{r}}\|, d + t' - \tau'\Big), \\ &= (T2^{-m})^{-\frac{n+1}{2}} \left\{ \sum_{i,i'=0}^{p_t} G\Big(\|\widetilde{\mathbf{r}}\|, d + \omega^{(i)} - \omega^{(i')}\Big) L_i(t') L_{i'}(\tau') + E_{p_t}\big(\|\widetilde{\mathbf{r}}\|\big) \right\} \\ &= \sum_{i,i'=0}^{p_t} G\Big(\|\mathbf{r}\|, t^{(i)} - \tau^{(i')}\Big) L_i(t') L_{i'}(\tau') + (T2^{-m})^{-\frac{n+1}{2}} E_{p_t}\big(\|\widetilde{\mathbf{r}}\|\big). \end{aligned} \tag{5.25}$$

Here, $\omega^{(i)}$ and $\omega^{(i')}$ are interpolation nodes in $[0,1]$, $t^{(i)} = T2^{-m}(d+\omega^{(i)})$, $\tau^{(i')} = T2^{-m}\omega^{(i')}$, $L_i$ are Lagrange polynomials and $p_t$ is the interpolation order. For interpolation at the Chebyshev nodes it can be shown that the error $E_{p_t}(r)$ decays exponentially in $p_t$, at a rate that is bounded independently of $r$ or $m$. Hence we obtain a bound of the form $2^{m\frac{n+1}{2}}\eta^{p_t}$.

If we let $p_t \sim L_t$ then in the worst case $m = L_t$ the error decays exponentially in $L_t$ and the number terms in the series (5.25) is order $(L_t + 1)^2$.

Neglecting the interpolation error and substituting the series of (5.25) in (5.20) results in a decomposition into Kronecker products. It follows that

$$\mathbf{c}_m^d \approx \sum_{i,i'=0}^{p_t} \mathbf{a}^{(m,i)} \left(\mathbf{a}^{(m,i')}\right)^T \otimes \mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}, \tag{5.26}$$

where

$$\begin{aligned} \left[\mathbf{a}^{(m,i)}\right]_{k_t} &= \int_{I_{m,0}} L_i(\tau')\Phi^I_{\tilde{\ell}_t,k_t}(\tau)d\tau\,, \\ \left[\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}\right]_{k_s,k_s'} &= \int_\Gamma \int_\Gamma G(\|\mathbf{x}-\mathbf{y}\|, t^{(i)} - \tau^{(i')})\Phi^\Gamma_{\tilde{\ell}_s,k_s}(\mathbf{x})\Phi^\Gamma_{\tilde{\ell}_s,k_s'}(\mathbf{y})\,\mathrm{d}\sigma_\mathbf{y}\,\mathrm{d}\sigma_\mathbf{x}\,. \end{aligned}$$

Note that $\mathbf{a}^{(m,i)}$ is a vector of length $2^{-m}\left|\Delta_{\tilde{\ell}_t}\right|$ and $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$ is a square matrix of size $\left|\Delta_{\tilde{\ell}_s}\right|$. Since the interpolation points and ansatz functions in $I_{m,d}$ are obtained by shifting $2^{-m}Td$ units from the interval $I_m^0$ the vector $\mathbf{a}^{(m,i)}$ is the same for $t$- and the $\tau$-variable.

*Temporal near-field.* For uniform time discretization, the matrices $\mathbf{c}^d_{\tilde{\ell}_t}$, $d \in \{0,1\}$, in (5.23) have the block-Toeplitz structure

$$\mathbf{c}^0_{\tilde{\ell}_t} = \begin{bmatrix} \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,0)} & & & \\ \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,1)} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t-1)} & \cdots & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,1)} & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,0)} \end{bmatrix} \quad \text{and} \quad \mathbf{c}^1_{\tilde{\ell}_t} = \begin{bmatrix} \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t)} & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t-1)} & \cdots & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,1)} \\ \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t+1)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t+1)} \\ \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,2n_t-1)} & \cdots & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t+1)} & \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,n_t)} \end{bmatrix}$$

where $n_t = \dim V_0^I$ and

$$\left[\mathbf{b}_{\tilde{\ell}_s}^{(\tilde{l}_t,i)}\right]_{k_s,k_s'} = \int_\Gamma \int_\Gamma G_{\tilde{\ell}_t,i}(\|\mathbf{x}-\mathbf{y}\|)\Phi^\Gamma_{\tilde{\ell}_s,k_s}(\mathbf{x})\Phi^\Gamma_{\tilde{\ell}_s,k_s'}(\mathbf{y})\,\mathrm{d}\sigma_\mathbf{y}\,\mathrm{d}\sigma_\mathbf{x}\,. \tag{5.27}$$

Here, the kernel contains integration with the ansatz functions in time

$$G_{\tilde{\ell}_t,i}(\|\mathbf{r}\|) = \int_0^T \int_0^T G(\|\mathbf{r}\|, t-\tau)\Phi^I_{\tilde{\ell}_t,0}(\tau)\Phi^I_{\tilde{\ell}_t,i}(t)\;\mathrm{d}\tau\,\mathrm{d}t.$$

The kernel can be expressed in closed form. For the case $i = 0$, the kernel has a $\mathcal{O}(1/\|\mathbf{r}\|)$ singularity, for $i = 1$ the singularity is $\mathcal{O}(\|\mathbf{r}\|)$, and for $i \geq 2$ the kernel is smooth. For the singular cases the spatial integration of the coefficients of (5.27) can be computed with generalized Duffy transforms, similar to the those used for elliptic boundary integral operators, see [13].

Define the shift-matrices

$$\mathbf{s}_n^{(i)} = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ 1 & & \ddots & & \\ & \ddots & & \ddots & \\ & & 1 & & 0 \end{bmatrix},$$

where $n$ indicates the dimension and $i$ the position of the sub-diagonal. Moreover, define

$$\mathbf{S}_{\tilde{\ell}_t}^{(i)} = \begin{cases} \mathbf{s}_{n_t 2^{\tilde{\ell}_t}}^{(i)}, & 0 \le i \le n_t - 1, \\ \mathbf{H}_{\tilde{\ell}_t}^1 \otimes \mathbf{s}_{n_t}^{(i-n_t)}, & n_t \le i \le 2n_t - 1. \end{cases}$$

Then, the near-field in (5.23) can be written as

$$\sum_{d \in \{0,1\}} \mathbf{H}_{\tilde{\ell}_t}^d \otimes \mathbf{c}_{\tilde{\ell}_t}^d = \sum_{i=0}^{2n_t-1} \mathbf{S}_{\tilde{\ell}_t}^{(i)} \otimes \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t, i)}. \tag{5.28}$$

*Tensor product form of* $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$. The approximation of $\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ in the form of (5.16) can now be obtained by combining (5.14), (5.23), (5.26) and (5.28). Using the multiplication rules of the Kronecker product, we conclude that

$$\mathbf{V}_{\boldsymbol{\ell},\boldsymbol{\ell}'} \approx \sum_{i=0}^{2n_t-1} \mathbf{A}_{\ell_t,\ell_t'}^{(i)} \otimes \mathbf{B}_{\ell_s,\ell_s'}^{(\tilde{\ell}_t,i)} + \sum_{\substack{m \in \{2,\dots,\tilde{\ell}_t\} \\ d \in \{2,3\} \\ i,i' \in \{0,\dots,p_t\}}} \mathbf{A}_{\ell_t,\ell_t'}^{(m,d,i,i')} \otimes \mathbf{B}_{\ell_s,\ell_s'}^{(m,d,i,i')}, \tag{5.29}$$

where

$$\begin{aligned} \mathbf{A}_{\ell_t,\ell_t'}^{(i)} &= \mathbf{I}_{\tilde{\ell}_t}^{\ell_t} \mathbf{S}_{\tilde{\ell}_t}^{(i)} \mathbf{I}_{\ell_t'}^{\tilde{\ell}_t}, \\ \mathbf{A}_{\ell_t,\ell_t'}^{(m,d,i,i')} &= \mathbf{I}_{\tilde{\ell}_t}^{\ell_t} \left( \mathbf{H}_m^d \otimes \mathbf{a}_m^{(i)} \left( \mathbf{a}_m^{(i)} \right)^T \right) \mathbf{I}_{\ell_t'}^{\tilde{\ell}_t}, \\ \mathbf{B}_{\ell_s,\ell_s'}^{(\tilde{\ell}_t,i)} &= \mathbf{J}_{\tilde{\ell}_s}^{\ell_s} \mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,i)} \mathbf{J}_{\ell_s'}^{\tilde{\ell}_s}, \\ \mathbf{B}_{\ell_s,\ell_s'}^{(m,d,i,i')} &= \mathbf{J}_{\tilde{\ell}_s}^{\ell_s} \mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')} \mathbf{J}_{\ell_s'}^{\tilde{\ell}_s}. \end{aligned}$$

Clearly, the matrices $\mathbf{A}_{\ell_t,\ell_t'}^{(i)}$ and $\mathbf{A}_{\ell_t,\ell_t'}^{(m,d,i,i')}$ can be applied with order $2^{\tilde{\ell}_t}$ operations. Note that the order in which the Kronecker product in the second matrix is evaluated is irrelevant, because both factors are square. In the following section, we will show that the matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\ell_t,i)}$ and $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$ can be applied with order $L_s^7 2^{n\tilde{\ell}_s}$ complexity. Then it follows easily that $\mathbf{B}_{\ell_s,\ell_s'}^{(i)}$ and $\mathbf{B}_{\ell_s,\ell_s'}^{(m,d,i,i')}$ can be applied with the same order of operations.

This, together with Lemma 5.1 and the fact that $p_t \sim L_t$ in (5.25) implies that the matrix $\widehat{\mathbf{V}}_{\mathbf{L}}$ can be applied with $\mathcal{O}\big(L_s^8 L_t^3 \dim(\widehat{U}_{\mathbf{L}})\big)$ cost. Thus the complexity of the algorithm described in Section 5 is log-linear in $\dim(\widehat{U}_{\mathbf{L}})$.

## 6. Fast evaluation of the Matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,i)}$ and $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$

In this section, we show that the spatial matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,i)}$ and $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$ are $\mathcal{H}$-matrices and describe an algorithm compute matrix vector products in $\mathcal{O}\big(L_s^7 2^{2\tilde{\ell}_s}\big)$ complexity. To simplify the discussion we restrict ourselves to the more important case of a two dimensional surface in three-space, that is, $n = 2$ in (2.5). The modifications for the case $n = 1$ are trivial and will result in lower powers of $L_s$ in the complexity estimate.

Since the calculus with $\mathcal{H}$-matrices is well known, see [10, 11], we limit ourselves to a high-level description of the algorithm mainly to set the stage for the ensuing error analysis. There, we will show how the parameters of the algorithm can be selected such that error and complexity bounds can be obtained that are independent of the parameters $\tilde{\ell}_t$, $m$ and $d, i, i'$.

We first give more detail on how the spatial finite element spaces $V_{\ell_s}^{\Gamma}$ are generated. To that end, assume that the surface $\Gamma$ is given by a number of parameterizations of the reference triangle $\hat{\sigma} = \{(\hat{x}_1, \hat{x}_2) : 0 \le \hat{x}_2 \le \hat{x}_1 \le 1\}$

$$x_\nu : \hat{\sigma} \to \Gamma_\nu, \quad \nu \in \mathcal{P}(0)\,,$$

where $\mathcal{P}(0)$ is an index set for the initial triangular patches. We assume that the interiors of $\Gamma_\nu$ are disjoint and that common sides of two adjacent $\Gamma_\nu$'s are parametrized in a consistent manner.

The coarsest space $V_0^{\Gamma}$ consists of functions whose preimage in $\hat{\sigma}$ is a polynomial. The spaces $V_{\ell_s}^{\Gamma}$ consist of functions whose preimages are piecewise polynomials on the $\ell_s$-th uniform refinement of $\hat{\sigma}$. Every $\ell_s$-th level refined triangle parameterizes a triangular patch $\Gamma_\nu$, $\nu \in \mathcal{P}(\ell_s)$ which in turn generates a sequence of triangularizations of $\Gamma$

$$\Gamma = \bigcup_{\nu \in \mathcal{P}(\ell_s)} \Gamma_\nu.$$

The uniform refinement implies a tree structure in the sense that every triangular patch $\Gamma_\nu$, $\nu \in \mathcal{P}(\ell_s)$ is the union of four triangular patches in level $\ell_s + 1$, denoted as the four children $\mathcal{K}(\nu)$ of $\nu$

$$\Gamma_\nu = \bigcup_{\nu' \in \mathcal{K}(\nu)} \Gamma_{\nu'}.$$

Moreover, every patch $\nu$ in level $\ell_s > 0$ has a parent $\pi(\nu)$ in level $\ell_s - 1$.

The neighbors $\mathcal{N}(\nu)$ of a patch $\nu \in \mathcal{P}(\ell_s)$ are given by

$$\mathcal{N}(\nu) = \left\{ \nu' \in \mathcal{P}(\ell_s) : \min_{\substack{\mathbf{x}\in\Gamma_\nu \\ \mathbf{y}\in\Gamma_{\nu'}}} \|\mathbf{x} - \mathbf{y}\| \le S L_s^{\frac{1}{2}} 2^{-\ell_s} \right\}. \tag{6.30}$$

Here, $S > 0$ is a predetermined constant. The factor $L_s^{\frac{1}{2}}$ implies that the neighbor list is expanded as the mesh is refined and is necessary to ensure convergence of the method. We assume that the constants are such that all patches in level zero are neighbors of each other. The interaction list $\mathcal{I}(\nu)$ of a patch $\nu \in \mathcal{P}(\ell_s)$ is the set of patches whose parents are neighbors, but who are not neighbors themselves:

$$\mathcal{I}(\nu) = \big\{\nu' \in \mathcal{P}(\ell_s) : \pi(\nu') \in \mathcal{N}\big(\pi(\nu)\big) \text{ and } \nu' \notin \mathcal{N}(\nu)\big\}.$$

Because of the uniform subdivision, the number of neighbors and the number of patches in interactions list are $\mathcal{O}(L_s)$.

The definition of neighbors and interaction lists implies the subdivision

$$\Gamma \times \Gamma = \bigcup_{\substack{\nu \in \mathcal{P}(\tilde{\ell}_s)\\ \nu' \in \mathcal{N}(\nu)}} \Gamma_\nu \times \Gamma_{\nu'} \quad \cup \quad \bigcup_{\ell_s=0}^{\tilde{\ell}_s} \bigcup_{\substack{\nu \in \mathcal{P}(\ell_s)\\ \nu' \in \mathcal{I}(\nu)}} \Gamma_\nu \times \Gamma_{\nu'}, \tag{6.31}$$

where the number of terms is $\mathcal{O}\big(L_s 2^{2\tilde{\ell}_s}\big)$.

Let $\mathbf{b}_{\tilde{\ell}_s}$ be one of the spatial matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\ell_t,i)}$ or $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$ and let $G(\cdot)$ denote its kernel. Since we will introduce additional superscipts below, we omit the kernel identifying superscripts for notational convenience. From the subdivision (6.31), we obtain the decomposition

$$\mathbf{b}_{\tilde{\ell}_s} = \mathbf{b}_{\tilde{\ell}_s}^{\text{near}} + \sum_{\ell_s=0}^{\tilde{\ell}_s} \mathbf{b}_{\tilde{\ell}_s}^{(\ell_s)}, \tag{6.32}$$

where $k_s, k_s' \in \Delta_{\tilde{\ell}_s}$ and

$$\begin{aligned}
\Big[\mathbf{b}_{\tilde{\ell}_s}^{\text{near}}\Big]_{k_s,k_s'} &= \sum_{\substack{\nu \in \mathcal{P}(\tilde{\ell}_s)\\ \nu \in \mathcal{N}(\nu)}} \int_{\Gamma_\nu} \int_{\Gamma_{\nu'}} G(\|\mathbf{x}-\mathbf{y}\|) \Phi_{\tilde{\ell}_s,k_s}^\Gamma(\mathbf{x}) \Phi_{\tilde{\ell}_s,k_s'}^\Gamma(\mathbf{y})\,\mathrm{d}\sigma_{\mathbf{y}}\,\mathrm{d}\sigma_{\mathbf{x}}\,,\\
\Big[\mathbf{b}_{\tilde{\ell}_s}^{(\ell_s)}\Big]_{k_s,k_s'} &= \sum_{\substack{\nu \in \mathcal{P}(\ell_s)\\ \nu' \in \mathcal{I}(\nu)}} \int_{\Gamma_\nu} \int_{\Gamma_{\nu'}} G(\|\mathbf{x}-\mathbf{y}\|) \Phi_{\tilde{\ell}_s,k_s}^\Gamma(\mathbf{x}) \Phi_{\tilde{\ell}_s,k_s'}^\Gamma(\mathbf{y})\,\mathrm{d}\sigma_{\mathbf{y}}\,\mathrm{d}\sigma_{\mathbf{x}}\,.
\end{aligned}$$

Since the number of basis functions in level $\tilde{\ell}_s$ that overlap with a patch in level $\tilde{\ell}_s$ are bounded, the matrix $\mathbf{b}_{\tilde{\ell}_s}^{\text{near}}$ has $\mathcal{O}\big(L_s 2^{2\tilde{\ell}_s}\big)$ nonvanishing entries. Of course, the matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\ell_s)}$ become increasingly populated as the level $\ell_s$ decreases, but since the integrals are over patches in interaction lists, the kernels are smooth functions. Thus, we can approximate the kernel by a degenerate expansion which will lead to a factorization that can be evaluated with $\mathcal{O}\big(L_s 2^{2\tilde{\ell}_s}\big)$ complexity.

To that end, we enclose every patch $\Gamma_\nu$ in $\mathcal{P}(\ell_s)$ by an axiparallel cube with sidelength $2S_1 2^{-\ell_s}$ and center $\mathbf{x}_\nu$. The constant $S_1$ is chosen such that the cubes will contain the patch $\Gamma_\nu$ tightly which is possible because of the uniform refinement scheme. Then any point in the enclosing cube has local coordinates in $[-1,1]^3$, that is,

$$\mathbf{x} = \mathbf{x}_\nu + 2^{-\ell_s} S_1 \hat{\mathbf{x}}, \quad \text{where} \quad \hat{\mathbf{x}} \in [-1,1]^3. \tag{6.33}$$

Now, consider two points $\mathbf{x} \in \Gamma_\nu$, $\mathbf{y} \in \Gamma_{\nu'}$, where $\nu \in \mathcal{P}(\ell_s)$ and $\nu' \in \mathcal{I}(\nu)$, with corresponding local coordinates $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. The kernel is now expanded into a Chebyshev series in the local coordiates, that is,

$$G(\|\mathbf{x}-\mathbf{y}\|) \approx \sum_{\substack{|\boldsymbol{\alpha}|\le p_s\\ |\boldsymbol{\beta}|\le p_s}} E^{\nu,\nu'}_{\boldsymbol{\alpha},\boldsymbol{\beta}} T_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}) T_{\boldsymbol{\beta}}(\hat{\mathbf{y}}) \tag{6.34}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are multiindices and $T_{\boldsymbol{\alpha}}(\cdot)$ are the Chebyshev polynomials. We assume that the expansion order is sufficiently large such that the error can be neglected. Then replacing the kernel by the expansion leads to

$$\left[\mathbf{b}^{(\ell_s)}_{\tilde{\ell}_s}\right]_{k_s,k_s'} \approx \sum_{\substack{\nu\in\mathcal{P}(\ell_s)\\ \nu'\in\mathcal{I}(\nu)}} \sum_{\substack{|\boldsymbol{\alpha}|\le p_s\\ |\boldsymbol{\beta}|\le p_s}} E^{\nu,\nu'}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \int_{\Gamma_\nu} T_{\boldsymbol{\alpha}}(\hat{\mathbf{x}}) \Phi^{\Gamma}_{\tilde{\ell}_s,k_s}(\mathbf{x}) \,\mathrm{d}\sigma_{\mathbf{x}} \int_{\Gamma_{\nu'}} T_{\boldsymbol{\beta}}(\hat{\mathbf{y}}) \Phi^{\Gamma}_{\tilde{\ell}_s,k_s'}(\mathbf{y}) \,\mathrm{d}\sigma_{\mathbf{y}}\,. \tag{6.35}$$

In matrix form, this can be expressed as the factorization

$$\mathbf{b}^{(\ell_s)}_{\tilde{\ell}_s} \approx \widetilde{\mathbf{b}}^{(\ell_s)}_{\tilde{\ell}_s} = \left(\mathbf{M}^{(\ell_s)}_{\tilde{\ell}_s}\right)^T \mathbf{E}^{(\ell_s)} \mathbf{M}^{(\ell_s)}_{\tilde{\ell}_s}$$

where the the matrices $\mathbf{M}^{(\ell_s)}_{\tilde{\ell}_s}$ contain the moments, i.e., the integrals in (6.35), and the matrices $\mathbf{E}^{(\ell_s)}$ contain the expansion coefficients $E^{\nu,\nu'}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. It is not hard to see that these matrices can be evaluated with $\mathcal{O}\big(L_s p_s^3 2^{2\tilde{\ell}_s}\big)$ and $\mathcal{O}\big(L_s p_s^6 2^{2\ell_s}\big)$ complexity.

Finally, we note that all kernels $G(\cdot)$ decay exponentially at infinity. Since interaction lists in the coarser levels contain increasingly distant pairs of patches, it is not necessary to evaluate all terms in the sum (6.32). Instead, we select a minimal level $\bar{\ell}_s$ and evaluate the approximation

$$\mathbf{b}_{\tilde{\ell}_s} \approx \widetilde{\mathbf{b}}_{\tilde{\ell}_s} = \mathbf{b}^{\mathrm{near}}_{\tilde{\ell}_s} + \sum_{\ell_s=\bar{\ell}_s}^{\tilde{\ell}_s} \widetilde{\mathbf{b}}^{(\ell_s)}_{\tilde{\ell}_s}. \tag{6.36}$$

In the following, we will show that the choice of parameters

$$p_s \sim L_s \quad \text{and} \quad \bar{\ell}_s = \begin{cases} \frac{\tilde{\ell}_t}{2} & \text{when } \mathbf{b}_{\tilde{\ell}_s} = \mathbf{b}^{(\tilde{\ell}_t,i)}_{\tilde{\ell}_s}, \\ \frac{m}{2} & \text{when } \mathbf{b}_{\tilde{\ell}_s} = \mathbf{b}^{(m,d,i,i')}_{\tilde{\ell}_s}, \end{cases} \tag{6.37}$$

will be sufficent to ensure that the approximation error does not affect the asymptotic convergence of the discretization error. Thus the complexity of a matrix vector product of $\mathbf{b}_{\tilde{\ell}_s}$ using the approximation (6.36) is $\mathcal{O}\big(L_s^7 2^{n\tilde{\ell}_s}\big)$.

Note that the introduction of the minimal level $\bar{\ell}_s$ does not reduce the asymptotic cost of the matrix-vector product, but ensures the accuracy of the degenerate kernel expansion (6.34). This will become clear in the following error analysis.

## 7. Error Analysis of the Fast Evaluation of the Spatial Matrices

For points $\mathbf{x} \in \Gamma_\nu$ and $\mathbf{y} \in \Gamma_{\nu'}$ on the patches in the subdivision (6.31) the kernel of the matrix $\mathbf{b}_{\tilde{\ell}_s}$ in (6.36) is given by

$$\widetilde{G}(\mathbf{x},\mathbf{y}) = \begin{cases} G(\|\mathbf{x}-\mathbf{y}\|), & \nu \in \mathcal{P}(\tilde{\ell}_s),\ \nu' \in \mathcal{N}(\nu), \\ G_p(\mathbf{x},\mathbf{y}), & \nu \in \mathcal{P}(\ell_s),\ \nu' \in \mathcal{I}(\nu),\ \bar{\ell}_s \le \ell_s \le \tilde{\ell}_s, \\ 0, & \nu \in \mathcal{P}(\ell_s),\ \nu' \in \mathcal{I}(\nu),\ 0 \le \ell_s < \bar{\ell}_s, \end{cases}$$

where $G_{p_s}$ is the truncated series expansion in (6.34). In this section we will prove the following result.

**Lemma 7.1.** *For $p_s$ and $\bar{\ell}_s$ given by* (6.37)*, there are constants $C > 0$, $\eta > 1$, independent of $\tilde{\ell}_s$, $\tilde{\ell}_t$, $m$, $d$, $i$ and $i'$, such that*

$$\left| G(\|\mathbf{x}-\mathbf{y}\|) - \widetilde{G}(\mathbf{x},\mathbf{y}) \right| \le C\eta^{-L_s}\,. \tag{7.38}$$

The lemma asserts exponential decay in $L_s$. From the Strang lemma it then follows that replacing $\mathbf{b}_{\tilde{\ell}_s}$ by $\widetilde{\mathbf{b}}_{\tilde{\ell}_s}$ results in an exponentially small error of the solution. Since the convergence of the discretization method is algebraic, the discretization error dominates the error of the fast method.

The two error sources are the far-field truncation, i.e., replacing the kernel by zero in levels $\ell_s < \bar{\ell}_s$, and the Chebyshev approximation in levels $\bar{\ell}_s \le \ell_s \le \tilde{\ell}_s$. The estimate of the latter error is based on the following result.

**Lemma 7.2.** *If $\hat{x} \mapsto f(r,\hat{x})$ is a function that for all $r \ge r_0 > 0$ is analytic in the same neighborhood of the interval $[-1,1]$ in the complex plane, then there are constants $C > 0$ and $\eta > 1$ such that for all $\rho \in [0,1], r \ge r_0$ the approximation error of the truncated Chebyshev series satisfies*

$$\max_{-1\le\hat{x}\le 1} \left| f(r,\rho\hat{x}) - \sum_{n=0}^{p} f_n(r,\rho) T_n(\hat{x}) \right| \le C\eta^{-p}.$$

*Here, $T_n$ are the Chebyshev polynomials and*

$$f_n(r,\rho) = \int_{-1}^{1} f(r,\rho\hat{x}) \frac{T_n(\hat{x})}{\sqrt{1-\hat{x}^2}}\,\mathrm{d}\hat{x}.$$

The proof of Lemma 7.2 for fixed $r,\rho$ is standard and the uniformity of $C$ and $\eta$ in $\rho \in [0,1], r \ge r_0$ follows easily from the proof. We omit the details. The analogous result also holds for multivariate functions, when $\hat{\mathbf{x}} \in [-1,1]^n$.

*Proof of Lemma 7.1.* We begin with the far-field truncation for $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$. The kernel of the matrix is

$$G(\|\mathbf{x}-\mathbf{y}\|) = \exp\left( -\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2^{-m}\delta} \right),$$

where $\delta = d+\omega_i-\omega_{i'}$ is in the interval $[1,5]$. For the points $\mathbf{x}\in\Gamma_\nu$ and $\mathbf{y}\in\Gamma_{\nu'}$, $\nu\in\mathcal{P}(\ell_s)$ and $\nu'\in\mathcal{I}(\nu)$ the distance satisfies $\|\mathbf{x}-\mathbf{y}\|\geq SL_s^{\frac{1}{2}}2^{-\ell_s}$ because $\nu'$ and $\nu$ are not neighbors. Thus the estimate

$$G(\|\mathbf{x}-\mathbf{y}\|)\leq \exp\left(-2^{m-2\ell_s}L_s\frac{S^2}{\delta}\right)\leq \exp\left(-L_s\frac{S^2}{\delta}\right)$$

holds when $\ell_s<\frac{m}{2}$ and the bound in (7.38) is established for $\eta=\exp\left(\frac{S^2}{\delta}\right)$.

We now consider the far-field truncation error for the matrices $\mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,s)}$. A simple change of variables shows that the kernel is

$$G(\|\mathbf{x}-\mathbf{y}\|)=\int_0^{h_t}\int_0^t\frac{1}{(t-\tau)^{\frac{3}{2}}}\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{4(t-\tau)}\right)\mathrm{d}\tau\,\mathrm{d}t = h_t^{\frac{1}{2}}g_d\left(\frac{\|\mathbf{x}-\mathbf{y}\|}{\sqrt{h_t}}\right).$$

where $g_d$ is given by

$$g_d(r)=\begin{cases}\displaystyle\int_0^1\int_0^t\frac{1}{(t-\tau)^{\frac{3}{2}}}\exp\left(-\frac{r^2}{4(t-\tau)}\right)\mathrm{d}\tau\,\mathrm{d}t, & d=0,\\ \displaystyle\int_0^1\int_0^1\frac{1}{(d+t-\tau)^{\frac{3}{2}}}\exp\left(-\frac{r^2}{4(d+t-\tau)}\right)\mathrm{d}\tau\,\mathrm{d}t, & 0<d<\dim(V_0^I).\end{cases}$$

These functions can be expressed in closed form using incomplete gamma functions and satisfy the estmate $g_d(r)\leq\frac{C}{r^2}\exp\left(-\frac{r^2}{d+1}\right)$. As before, it follows for $\mathbf{x}\in\Gamma_\nu$ and $\mathbf{y}\in\Gamma_{\nu'}$, where $\nu\in\mathcal{P}(\ell_s)$, $\nu'\in\mathcal{I}(\nu)$ and $\tilde{\ell}_s<\frac{\tilde{\ell}_t}{2}$, that $G(\|\mathbf{x}-\mathbf{y}\|)\leq C\exp\left(-L_s\frac{S^2}{d+1}\right)$ holds. This is the bound in (7.38).

We now turn to the Chebyshev approximation error of the matrices $\mathbf{b}_{\tilde{\ell}_s}^{(m,d,i,i')}$.

The kernel is

$$G(\|\mathbf{x}-\mathbf{y}\|)=\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2^{-m}\delta}\right)=\exp\left(-2^{m-2\ell_s}\frac{S_1^2}{\delta}\|\mathbf{r}_{\nu,\nu'}+\hat{\mathbf{x}}-\hat{\mathbf{y}}\|^2\right),$$

where $\mathbf{x}\in\Gamma_\nu$, $\mathbf{y}\in\Gamma_{\nu'}$, $\nu\in\mathcal{P}(\ell_s)$, $\nu'\in\mathcal{I}(\nu)$, $\delta=d+\omega_i-\omega_{i'}$ and the constant $S_1$ is from (6.33). To estimate the truncation error use Lemma 7.2. Here, the factor $2^{m-2\ell_s}$ plays the role of the parameter $\rho$ and $\mathbf{r}_{\nu,\nu'}$ that of $r$. Since the summation in (6.36) is over the levels $\ell_s\geq\bar{\ell}_s=\frac{m}{2}$, it follows that indeed $0<\rho\leq 1$. Likewise, the scaling of the enclosing cubes and the definition of the neighbors in (6.30) implies that $\|\mathbf{r}_{\nu,\nu'}\|\geq 3$ if $S$ and $L_s$ are sufficiently large. Thus the lemma implies that the error decays exponentially in $p_s$ and since $p_s\sim L_s$ the bound (7.38) follows.

It remains to estimate the truncation error in $\mathbf{b}_{\tilde{\ell}_s}^{(\tilde{\ell}_t,d)}$. The argument is based on a similar scaling. In local coordinates, the kernel is

$$G(\|\mathbf{x}-\mathbf{y}\|)=\sqrt{T}2^{\frac{\tilde{\ell}_t}{2}}g_d\left(\frac{\|\mathbf{x}-\mathbf{y}\|}{\sqrt{T}2^{\frac{\tilde{\ell}_t}{2}}}\right)=\sqrt{T}2^{\frac{\tilde{\ell}_t}{2}}g_d\left(2^{\frac{\tilde{\ell}_t}{2}-\ell_s}\frac{S_1}{\sqrt{T}}\|\mathbf{r}_{\nu,\nu'}+\hat{\mathbf{x}}-\hat{\mathbf{y}}\|\right)$$

First note that this is an analytical function in $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ because the difference $\|\mathbf{r}_{\nu,\nu'}+\hat{\mathbf{x}}-\hat{\mathbf{y}}\|$ is uniformly bounded away from zero for patches in the interaction lists. The role of $\rho$ in Lemma 7.2 is played by the factor $2^{\frac{\tilde{\ell}_t}{2}-\ell_s}$. Because of $\ell_s\geq\bar{\ell}_s=\frac{\ell_t}{2}$, it follows that $0<\rho\leq 1$

and, thus the lemma guarantees exponential decay in $p_s$ and hence in $L_s$, which establishes (7.38). □

## 8. A Numerical Example

To illustrate the theory presented in this work, we discuss numerical results obtained with an implementation of the method. We solve the indirect integral formulation (2.7) where $\Gamma$ is the unit sphere and $I = [0,1]$. The right hand side $f(\mathbf{x},t)$ is chosen such that the solution is given by $g(\mathbf{x},t) = t^2(3x_3^2 - 1)$. The spaces $V_{\ell_s}^{\Gamma}$ are the continuous, piecewise linear functions (i.e., $d_s = 2$), subject to a triangulation of the sphere. The coarsest triangulation is obtained by radial projection of the tetrahedron onto the sphere. The spaces $V_{\ell_t}^{I}$ are the piecewise constants (i.e., $d_t = 1$), subject to a uniform subdivision of the unit interval, where initial space has five intervals. The relationship between the finest spatial and temporal resolution is $L_t = 2L_s$.

In Section 6 we have described how matrix vector products with the spatial matrices in (6.32) can be computed efficiently using $\mathcal{H}$-matrix calculus. For a fully discrete algorithm, the coefficients of the matices $\mathbf{b}_{\tilde{\ell}_s}^{\text{near}}$ must be computed by numerical quadrature. Since the kernels have in the worst case a $\mathcal{O}\left(\frac{1}{r}\right)$-singularity, one can use the singularity removing transformations of [12] combined with Gauss quadrature. However, some care must be applied because of the scaling of the kernel for different combinations of $\tilde{\ell}_s$ and $\tilde{\ell}_t$ or $m$. Therefore, this method is combined with an adaptive space refinement. Further, one can exploit the fact that computations for given values of $\tilde{\ell}_t$ and $m$ can be re-used for different values of $\tilde{\ell}_s$. This algorithm introduces additional logarithmic factors in the complexity estimate of the method.

Table 8.1 displays the dimensions of the full and sparse spaces as well as the $L_2$-error $\|g - \widehat{g}_{\boldsymbol{L}}\|_{L_2(\Gamma\times I)}$ of the solution. The expected convergence order in this norm is not $\mathcal{O}(2^{-2L_s})$ as in case of the full Galerkin method. This can be explained as follows. We have $L_s d_s = L_t d_t$, so that in view of Lemma 4.2 the convergence rate with respect to the energy norm is

$$\|g - \widehat{g}_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac{1}{2},-\frac{1}{4}}(\Gamma\times I)} \lesssim L_s 2^{-\frac{L_s L_t}{4L_s+2L_t}} 2^{-L_s d_s} \|g\|_{\mathcal{H}_{\text{mix}}^{d_s,d_t}(\Gamma\times I)}.$$

Hence, inserting the $L^2$-orthogonal projection $\widehat{P}_{\boldsymbol{L}}$ onto the space $\widehat{U}_{\boldsymbol{L}}$, we find by the inverse inequality

$$\begin{aligned}
\|g - \widehat{g}_{\boldsymbol{L}}\|_{L_2(\Gamma\times I)} &\le \|(I - \widehat{P}_{\boldsymbol{L}})g\|_{L_2(\Gamma\times I)} + \|\widehat{P}_{\boldsymbol{L}} g - \widehat{g}_{\boldsymbol{L}}\|_{L_2(\Gamma\times I)} \\
&= \sqrt{L_s} 2^{-L_s d_s} \|g\|_{\mathcal{H}_{\text{mix}}^{d_s,d_t}(\Gamma\times I)} + (2^{L_s/2} + 2^{L_t/4}) \|\widehat{P}_{\boldsymbol{L}} g - \widehat{g}_{\boldsymbol{L}}\|_{\mathcal{H}^{-\frac{1}{2},-\frac{1}{4}}(\Gamma\times I)} \\
&\lesssim L_s (2^{L_s/2} + 2^{L_t/4}) 2^{-\frac{L_s L_t}{4L_s+2L_t}} 2^{-L_s d_s} \|g\|_{\mathcal{H}_{\text{mix}}^{d_s,d_t}(\Gamma\times I)}.
\end{aligned}$$

If we insert $d_s = 2$, $d_t = 1$, and thus $2L_s = L_t$, then we obtain

$$\|g - \widehat{g}_{\boldsymbol{L}}\|_{L_2(\Gamma\times I)} \lesssim L_s 2^{-L_s(d_s - 1/4)} \|g\|_{\mathcal{H}_{\text{mix}}^{d_s,d_t}(\Gamma\times I)} \sim L_s 2^{-\frac{7}{4}L_s} \|g\|_{\mathcal{H}_{\text{mix}}^{2,1}(\Gamma\times I)}.$$

| $L_s$ | $L_t$ | $\dim U_{\boldsymbol{L}}$ | fac | $\dim \widehat{U}_{\boldsymbol{L}}$ | fac | error | fac |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 2.00e+2 | | 1.10e+2 | | 3.77e-1 | |
| 2 | 4 | 2.72e+3 | 13.6 | 5.60e+2 | 5.09 | 2.87e-1 | 0.762 |
| 3 | 6 | 4.16e+4 | 15.2 | 2.72e+3 | 4.86 | 6.96e-2 | 0.242 |
| 4 | 8 | 6.58e+5 | 15.8 | 1.28e+5 | 4.71 | 1.82e-2 | 0.261 |
| 5 | 10 | 1.05e+7 | 16.0 | 5.89e+5 | 4.60 | 4.81e-3 | 0.264 |
| 6 | 12 | 1.68e+8 | 16.0 | 2.66e+6 | 4.52 | 1.38e-3 | 0.286 |

TABLE 8.1. Numerical results obtained with the implementation.

In Table 8.1, it can be seen that the error indeed closely reproduces the $\mathcal{O}(L_s 2^{-\frac{7}{4}L_s})$ convergence. Also, the dimensions of the sparse tensor product spaces $\dim \widehat{U}_{\boldsymbol{L}}$ reproduce the $\mathcal{O}(L_s 2^{2L_s})$ estimate of Lemma 4.1 well. Note that for the finer meshes the dimensions of the sparse spaces are dramatically smaller than the full tensor product spaces.

Table 8.2 displays complexity results with our implementation. Our code precomputes the matrices $\mathbf{b}_{\tilde{\ell}_s}^{\text{near}}$ in (6.32) and the coefficients $E_{\boldsymbol{\alpha},\boldsymbol{\beta}}^{\nu,\nu'}$ in (6.34) and store them in memory. We have parallelized this aspect in OpenMP using 16 treads and the timings are reported as setup time. The major cost of the iterative solver is in the computation of the matrix vector product. This aspect of the code is run in serial on a single thread and reported as the apply time. The table also displays the number of stored matrix- and translation coefficients.

From the shown data it is apparent that in most cases the magnification factors obtained are significantly smaller than 16. This shows that the sparse grid method has an improved complexity over any method that is based on the full grid discretization, even if that method has optimal complexity in $\dim U_{\boldsymbol{L}}$, such as the methods of [18] and [14].

However, for the smaller values of $L_s$ the observed memory allocation and cpu-times for our implementation grow much faster than the theroretical $\dim \widehat{U}_{\boldsymbol{L}}$ rate. The reason is that most of the computing resources are consumed by the many $\mathbf{b}_{\tilde{\ell}_s}$-matrices in (6.32). Since these matrices are relatively small for the values of $\tilde{\ell}_s$ that we computed, the $\mathcal{H}$-format does not yield high compression rates, because the asymptotic rates of Section 6 have not been reached. Only for the largest number of refinements the complexity curves level out and suggest that a nearly $\dim \widehat{U}_{\boldsymbol{L}}$ complexity is indeed possible.

## REFERENCES

[1] R. Balder and C. Zenger. The solution of multidimensional real Helmholtz equations on sparse grids. *SIAM J. Sci. Comput.* **17**(3) (1996) 631–646.

[2] H.J. Bungartz. A multigrid algorithm for higher order finite elements on sparse grids. *ETNA Electron. Trans. Numer. Anal.* **6** (1997) 63–77.

[3] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.* **13** (2004) 147–269.

[4] A. Chernov and D. Pham. Sparse spectral BEM for elliptic problems with random input data on a spheroid. *Adv. Comput. Math.* **41**(1) (2015) 77–104.

| $L_s$ | setup(s) | fac | apply(s) | fac | coeffs | fac |
|---|---|---|---|---|---|---|
| 1 | | | 6.00e-4 | | 4.68e+03 | |
| 2 | 3.00e+0 | | 7.99e-3 | 13.3 | 1.27e+05 | 27.1 |
| 3 | 3.60e+1 | 12.0 | 1.69e-1 | 21.2 | 2.57e+06 | 20.2 |
| 4 | 5.57e+2 | 15.5 | 3.56e+0 | 21.0 | 4.61e+07 | 17.9 |
| 5 | 1.19e+4 | 21.3 | 7.81e+1 | 21.9 | 8.20e+08 | 17.8 |
| 6 | 1.48e+5 | 12.4 | 1.24e+3 | 15.9 | 9.24e+09 | 11.3 |

TABLE 8.2. Timings in seconds and number of stored coefficients.


[5] A. Chernov and C. Schwab. Sparse space-time Galerkin BEM for the nonstationary heat equation. *ZAMM Z. Angew. Math. Mech.* **93** (2013) 403–413.

[6] M. Costabel. Boundary integral operators for the heat equation. *Integral Equations Operator Theory* **13** (1990) 498–552.

[7] M. Griebel and H. Harbrecht. On the construction of sparse tensor product spaces. *Math. Comput.* **82** (2013) 975–994.

[8] M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing* **81** (2007) 1–34.

[9] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numer. Mathematik* **83**(2) (1999) 279–312.

[10] W. Hackbusch. A sparse matrix arithmetic based on $\mathcal{H}$-matrices. Part I: Introduction to $\mathcal{H}$-matrices. *Computing* **64** (1999) 89–108.

[11] W. Hackbusch. *Hierarchical Matrices. Algorithms and Analysis.* Springer, Berlin-Heidelberg, 2015.

[12] S. Erichsen and S. Sauter. Efficient automatic quadrature in 3-d Galerkin BEM. *Comp. Meth. Appl. Mech. Engrg.* **157** (1998), 215-224.

[13] M. Messner, M. Schanz, and J. Tausch. A fast Galerkin method for parabolic space-time boundary integral equations. *J. Comp. Phys.* **258** (2014) 15–30.

[14] M. Messner, M. Schanz and J. Tausch. An efficient Galerkin boundary element method for the transient heat equation. *SIAM J. Sci. Comput.* 37-3 (2015), pp. A1554-A1576

[15] P. J. Noon. The Single Layer Heat Potential and Galerkin Boundary Element Methods for the Heat Equation. PhD thesis, University of Maryland, 1988.

[16] N. Reich. Wavelet compression of anisotropic integrodifferential operators on sparse tensor product spaces. *M2AN Math. Model. Numer. Anal.* **44**(1) (2010) 33–73.

[17] R. Stevenson and C. Schwab. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comput.* **78** (2009) 1293–1318.

[18] J. Tausch. A fast method for solving the heat equation by layer potentials. *J. Comp. Phys.* **224** (2007) 956–969.

[19] J. Tausch. Fast Nyström Methods for Parabolic Boundary Integral Equations. In U. Langer, M. Schanz und W.L. Wendland, editors, *Fast Boundary Element Methods in Engineering and Industrial Applications*, volume 63 of Lecture Notes in Computational Mechanics, pages 185–219, Springer, Berlin-Heidelberg, 2012.

[20] C. Winter. Wavelet Galerkin schemes for multidimensional anisotropic integrodifferential operators. *SIAM J. Sci. Comput.* **32**(3) (2010) 1545–1566.

[21] A. Zeiser. Fast matrix-vector multiplication in the sparse-grid Galerkin method. *J. Sci. Comput.* **47**(3) (2011) 328–346.

[22] C. Zenger. Sparse grids. In *Parallel algorithms for partial differential equations (Kiel, 1990)*, volume 31 of *Notes Numer. Fluid Mech.*, pages 241–251. Vieweg, Braunschweig, 1991.

Helmut Harbrecht, Departement Mathematik und Informatik, Universität Basel, Basel, Schweiz.

*E-mail address*: helmut.harbrecht@unibas.ch

Johannes Tausch, Department of Mathematics, Southern Methodist University, Dallas, Texas, USA.

*E-mail address*: tausch@smu.edu

## LATEST PREPRINTS

| No. | **Author:** *Title* |
|---|---|
| 2016-11 | **I. Hedén, S. Zimmermann**<br>*The decomposition group of a line in the plane* |
| 2016-12 | **J. Ballani, D. Kressner, M. Peters**<br>*Multilevel tensor approximation of PDEs with random data* |
| 2016-13 | **M. J. Grote, M. Kray, U. Nahum**<br>*Adaptive eigenspace method for inverse scattering problems in the frequency domain* |
| 2016-14 | **H. Harbrecht, M. Peters, M. Schmidlin**<br>*Uncertainty quantification for PDEs with anisotropic random diffusion* |
| 2016-15 | **F. Da Lio, L. Martinazzi**<br>*The nonlocal Liouville-type equation in $R$ and conformal immersions of the disk with boundary singularities* |
| 2016-16 | **A. Hyder**<br>*Conformally Euclidean metrics on $R^n$ with arbitrary total Q-curvature* |
| 2016-17 | **G. Mancini, L. Martinazzi**<br>*The Moser-Trudinger inequality and its extremals on a disk via energy estimates* |
| 2016-18 | **R. N. Gantner, M. D. Peters**<br>*Higher order quasi-Monte Carlo for Bayesian shape inversion* |
| 2016-19 | **C. Urech**<br>*Remarks on the degree growth of birational transformations* |
| 2016-20 | **S. Dahlke, H. Harbrecht, M. Utzinger, M. Weimar**<br>*Adaptive wavelet BEM for boundary integral equations: Theory and numerical experiments* |
| 2016-21 | **A. Hyder, S. Iula, L. Martinazzi**<br>*Large blow-up sets for the prescribed Q-curvature equation in the Euclidean space* |
| 2016-22 | **P. Habegger**<br>*The norm of Gaussian periods* |
| 2016-23 | **P. Habegger**<br>*Diophantine approximations on definable sets* |

---

Preprints are available under *https://math.unibas.ch/research/publications*

## LATEST PREPRINTS

| No. | **Author:** *Title* |
|---|---|
| 2016-24 | **F. Amoroso, D. Masser**<br>*Lower bounds for the height in Galois extensions* |
| 2016-25 | **W. D. Brownawell, D. W. Masser**<br>*Zero estimates with moving targets* |
| 2016-26 | **H. Derksen, D. Masser**<br>*Linear equations over multiplicative groups, recurrences, and mixing III* |
| 2016-27 | **D. Bertrand, D. Masser, A. Pillay, U. Zannier**<br>*Relative Manin-Mumford for semi-abelian surfaces* |
| 2016-28 | **L. Capuano, D. Masser, J. Pila, U. Zannier**<br>*Rational points on Grassmannians and unlikely intersections in tori* |
| 2016-29 | **C. Nobili, F. Otto**<br>*Limitations of the background field method applied to Rayleigh-Bénard convection* |
| 2016-30 | **W. D. Brownawell, D. W. Masser**<br>*Unlikely intersections for curves in additive groups over positive characteristic* |
| 2016-31 | **M. Dambrine, H. Harbrecht, M. D. Peters, B. Puig**<br>*On Bernoulli's free boundary problem with a random boundary* |
| 2016-32 | **H. Harbrecht, J. Tausch**<br>*A fast sparse grid based space-time boundary element method for the nonstationary heat equation* |

---

Preprints are available under *https://math.unibas.ch/research/publications*