Graz University of Technology
Institute of Applied Mechanics

Preprint Series

Preprint No 01/2012

Fast Galerkin Method for Parabolic Space-Time Boundary Integral Equations

Michael Messner, Martin Schanz
Institute of Applied Mechanics, Graz University of Technology
Johannes Tausch
Mathematics Department, Southern Methodist University

Published in: *Journal of Computational Physics,* **258** *(2014), 15–30*
DOI: 10.1016/j.jcp.2013.10.029
Latest revision: May 21, 2014

Abstract

An efficient scheme for solving boundary integral equations of the heat equation based on the Galerkin method is introduced. The parabolic fast multipole method (pFMM) is applied to accelerate the evaluation of the thermal layer potentials. In order to remain attractive for a wide range of applications, a key issue is to ensure efficiency for a big variety of temporal to spatial mesh ratios. Within the parabolic Galerkin FMM (pGFMM) it turns out that the temporal nearfield can become very costly. To that end, a modified fast Gauss transform (FGT) is developed. The complexity and convergence behavior of the method are analyzed and numerically investigated on a range of model problems. The results demonstrate that the complexity is nearly optimal in the number of discretization parameters while the convergence rate of the Galerkin method is preserved.

1 Introduction

Solving linear diffusion problems with the boundary element technique is an attractive alternative to the more standard PDE-based approaches, such as Finite Difference- and Finite Element Methods. By now, the boundary integral formulations of the heat equation and the mapping properties of their operators are well understood. In particular, the thermal single layer operator is coercive in an appropriate anisotropic Sobolev space [1, 4]. Likewise, the thermal double layer operator is compact in the proper setting [4, 11]. These results are the background for the analysis of Galerkin methods in both space and time.

While the focus in this work is on the Galerkin method, we note that other possible discretization schemes are feasible, e.g. the collocation method [2] and the Nyström method [18]. These methods can be used both in space and time, while either of them may be combined with the convolution quadrature method [13, 16] for the the temporal discretization. However, Galerkin methods have distinct advantages because of their stability and applicability to non-smooth surfaces. The different discretization options for parabolic integral equations and their theoretical foundations of are surveyed in [5].

Since integral operators are non-local, discretizations lead to dense matrices which makes fast methods imperative to handle large scale problems efficiently. This is a very well studied subject in the elliptic case [14], and has recently attracted considerable interest for parabolic boundary integral equations. A possible approach in this direction is to employ Fourier techniques, which was initiated in [8] and later extended in [7] and [12]. Nevertheless, the focus here is on clustering techniques, because of their success in the elliptic case. Typical examples of such methods are the Fast Multipole Method and H-matrices. Their idea is to agglomerate source- and evaluation panels and approximate admissible (i.e. well separated) interactions by a truncated series expansion.

In contrast to the elliptic case, discretizations of thermal layer potentials involve a temporal in addition to a spatial mesh, hence the clustering must be done in space and time. This scheme was originally described in [17] and [19]. The admissibility condition is that the temporal separation of two spatio-temporal clusters is proportional to their temporal size. Moreover, the scaling of the Green's function dictates that the temporal cluster size is proportional to the square of the spatial cluster size. Since the heat kernel is singular in time, nearby time steps must be evaluated directly.

The cost of the temporal nearfield evaluation is a major factor for the overall efficiency of the method. In our previous work, we considered a specific Nyström discretization method where the temporal and spatial mesh sizes satisfy $h_t \sim h_s^2$. In this case the nearfield can be evaluated in optimal complexity due to the exponential decay of the heat kernel in space. However, one of the reasons to use the Galerkin scheme is to allow for more general mesh ratios. This is often the case in realistic problems with complicated geometries where one would like to have a fine spatial mesh but relatively large time steps. In such a setting many panels in the temporal nearfield have to be considered, which effects the overall complexity of the scheme and thus makes an acceleration of the nearfield important.

The contribution of a single time step can be regarded as a Gauss transform, a computation that arises frequently in many different areas. The original method [9] is based on a Taylor expansion of the Gaussian (sometime better known as Hermite expansion), while more recent versions are

based on exponential [10] and Chebyshev expansions [20]. In the context considered here, the latter approach is more interesting because it has better convergence properties for large spatial clusters.

Unfortunately however, the relevant kernels in the Galerkin method are not Gaussians, but integrated versions, which imposes a number of technical difficulties. First, the kernels are rather complicated and, hence, it is tedious and impractical to derive the discrete translation operators. More importantly, the integrated kernels do not separate into spatial directions and thus the evaluation of translation operators becomes very expensive. To overcome these issues, we replace the integral by a high-order quadrature rule and thus evaluate a series of fast Gauss transforms.

The outline of this paper is as follows. In Sections 2 and 3, we briefly review parabolic boundary integral equations and the parabolic FMM. The purpose of these sections is only to introduce the subject and set the notations; a reader who is not familiar with these issues is encouraged to refer to [17, 19]. Section 4 describes the efficient evaluation of the temporal nearfield and contains the main new contributions of this work. Finally, Section 5 concludes with a number of simple but relevant numerical examples that illustrate the theoretical development of the previous sections.

2 Parabolic BEM

Green's formula relates the Dirichlet and Neumann data on the boundary S of a Lipschitz domain Ω. For the heat equation $\partial_t u = \Delta u$ with homogeneous initial conditions it can be stated as

$$\pm c u(x,t) = \mathcal{K}u(x,t) - \mathcal{V}\frac{\partial u}{\partial n}(x,t), \quad x \in S,\ t > 0. \tag{1}$$

Here, c is the jump term of the double layer potential with the plus sign applying for an exterior and the minus sign for an interior problem. The single- and double layer potentials are given by

$$\begin{aligned} \mathcal{V}g(x,t) &= \int_0^t \int_S G(x-y,t-\tau) g(y,\tau)\, \mathrm{d}s_y \mathrm{d}\tau, \\ \mathcal{K}g(x,t) &= \int_0^t \int_S \frac{\partial}{\partial n_y} G(x-y,t-\tau) g(y,\tau)\, \mathrm{d}s_y \mathrm{d}\tau, \end{aligned}$$

respectively, and the Green's function of the heat equation in $\mathbb{R}^3$ is

$$G(r,\delta) = \frac{1}{(4\pi\delta)^{\frac{3}{2}}} \exp\left(-\frac{|r|^2}{4\delta}\right). \tag{2}$$

For $\delta > 0$ it is a Gaussian in space that becomes increasingly peaked as δ gets smaller and in the limit as $\delta \to 0$, it converges to the delta distribution. Moreover, it exhibits the following scaling property we will often use

$$G(r,\lambda\delta) = \lambda^{-\frac{3}{2}} G\left(\frac{r}{\sqrt{\lambda}}, \delta\right). \tag{3}$$

We consider a Bubnov-Galerkin discretization scheme with space-time tensor product spaces $Q_h^{d_s,d_t} := Q_{h_s}^{d_s} \otimes Q_{h_t}^{d_t}$, where $Q_{h_s}^{d_s}$ is a space of piecewise polynomial functions subject to a conforming triangulation of the boundary with mesh-width h_s. We denote the nodal basis functions by φ_k^s, then

$$Q_{h_s}^{d_s} = \operatorname{span}\left\{ \varphi_1^s, \dots \varphi_{N_s}^s \right\}.$$

Likewise, $Q_{h_t}^{d_t}$ is to be understood as the space of piecewise polynomial ansatz functions on a uniform partition of the time interval $[0,T]$ into N_t time steps with step size $h_t = {}^{T}/_{N_t}$. For simplicity of exposition we only consider piecewise constant functions in the intervals $[t_{i-1}, t_i]$. Then the discretization matrix of the single layer potential is

$$\mathsf{V}_{(k,\ell),(i,j)} = \int_{t_{i-1}}^{t_i} \int_S \int_{t_{j-1}}^{t_j} \int_S G(x-y, t-\tau) \varphi_k^s(x) \varphi_\ell^s(y) \, \mathrm{d}s_y \mathrm{d}\tau \, \mathrm{d}s_x \mathrm{d}t \,, \tag{4}$$

and for the double layer potential accordingly. Note that the order of integration is irrelevant, because Fubini's Theorem applies. Thus we can perform the two time integrals first and the spatial integrals later. This leads to the representation

$$\mathsf{V}_{(k,\ell),(i,j)} = \int_S \int_S V_{i-j}(x-y) \varphi_k^s(x) \varphi_\ell^s(y) \, \mathrm{d}s_y \mathrm{d}s_x \,, \tag{5}$$

where $V_d(x-y)$ is a space-dependent kernel also depending on the difference $d = i - j$. With $r = x - y$ it is given by

$$V_d(r) = \int_{t_{i-1}}^{t_i} \int_{t_{j-1}}^{t_j} G(r, t-\tau) \mathrm{d}\tau \mathrm{d}t \,.$$

We introduce the change of variables $t = h_t(i - \frac{1}{2} + t')$ and $\tau = h_t(j - \frac{1}{2} + \tau')$ to end up with the Galerkin kernel in time [15]

$$\begin{aligned} V_d(\tilde{r}) &= \sqrt{h_t} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} G(\tilde{r}, d + t' - \tau') \mathrm{d}\tau' \mathrm{d}t' \\ &= \sqrt{h_t} \left[G^{(-2)}(\tilde{r}, d+1) - 2G^{(-2)}(\tilde{r}, d) + G^{(-2)}(\tilde{r}, d-1) \right], \end{aligned} \tag{6}$$

where $\tilde{r} = {}^{r}/_{\sqrt{h_t}}$ and $G^{(-2)}$ is the second anti-derivative of the heat kernel with respect to time

$$G^{(-2)}(r, \delta) = \frac{\sqrt{\delta}}{4\pi} \left[\left(\frac{|r|}{\sqrt{4\delta}} + \frac{\sqrt{\delta}}{|r|} \right) \operatorname{erfc}\left(\frac{|r|}{\sqrt{4\delta}} \right) - \frac{1}{\sqrt{\pi}} \exp\left(-\frac{|r^2|}{4\delta} \right) \right].$$

Since $G^{(-2)}(r,0) = 0$ expression (6) is valid for $d \geq 1$, while for $d = 0$ we have

$$V_0(\tilde{r}) = \sqrt{h_t} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{t'} G(\tilde{r}, t' - \tau') \mathrm{d}\tau' \mathrm{d}t' = \sqrt{h_t} G^{(-2)}(\tilde{r}, 1) \,. \tag{7}$$

Thus the discretization of (1) leads to the following time stepping scheme

$$(\mp c\mathsf{M}_0 + \mathsf{K}_0)\,\mathsf{u}_i - \mathsf{V}_0\mathsf{v}_i = \sum_{d=1}^{i} (\mathsf{V}_d\mathsf{v}_{i-d} - \mathsf{K}_d\mathsf{u}_{i-d}) \tag{8}$$

where $\mathsf{V}_d := \mathsf{V}_{(\cdot,\cdot),(i,j)}$ is the $N_s \times N_s$-matrix whose coefficients are given by (5), K_d is the analogous matrix for the double layer operator and M_0 is the mass matrix. The vectors v_j and u_j contain the expansion coefficients of the solution in the j-th time step. Thus the i-th time step involves evaluating the right hand side, which is a convolution over the time evolution, and the solution of a linear system for either u_i or v_i, depending on whether a Dirichlet or Neumann problem is solved.

Since all operators in (8) are dense, the total cost of the time stepping method is at least of order $N_t^2 N_s^2$ if the matrix-vector products are evaluated naively. Fortunately, the complexity can be reduced with the parabolic Fast Multipole Method. Here, the first few terms of the right hand side (temporal nearfield) are evaluated directly, exploiting the fact that the matrices are sparse due to the exponential decay of the kernel. The remaining terms are evaluated by clustering and truncated kernel expansions. The method was originally developed for Nyström discretizations of parabolic layer potentials [17, 19], but the modifications for the Galerkin method are rather straightforward.

However, the cost of evaluating the temporal nearfield depends strongly on the ratio of the temporal to the spatial mesh-width. In fact, in the following discussion we will show that order $N_t N_s$ complexity (up to logarithmic factors) can be obtained only if $h_t \sim h_s^2$. If the spatial mesh is finer, then the cost of the temporal nearfield will become dominant. To overcome this difficulty we will introduce a variant of the fast Gauss transform to obtain optimal complexity for arbitrary mesh ratios.

3 The Parabolic FMM

The key idea of the Fast Multipole Method is to agglomerate source- and evaluation panels and to use a suitable expansion of the kernel to accelerate the computation of the cluster interactions. For parabolic problems, the clustering is done in space and time. We briefly review this construction only to the extent necessary for the discussion of the temporal nearfield. The description of all the details in the parabolic FMM can be found in [17, 19].

3.1 Space-Time Clustering

Interacting space-time clusters must satisfy an admissibility condition to ensure that the expansion converges sufficiently fast for source- and evaluation points. The heat kernel is smooth in space unless $t = \tau$. When τ approaches t, it becomes more peaked and eventually approaches the delta distribution in $\mathbb{R}^3$. The proper admissibility condition can be derived from the space-time scaling expressed in equation (3).

Space and time are clustered using uniform hierarchic subdivisions in cubes and intervals. The root level cube contains the entire surface and is L_s times uniformly refined. Note that in a Galerkin scheme the supports of test- and trial functions need to be fully contained within

h_t ...	temporal mesh width
h_s ...	spatial mesh width
$h_t^{(\ell_t)}$ ...	half length of cubes in level ℓ_t of temporal tree
$h_s^{(\ell_s)}$ ...	half length of cubes in level ℓ_s of spatial tree
L_t ...	temporal leaf level of the pGFMM
L_s ...	spatial leaf level of the pGFMM
ρ_L ...	cluster ratio at the leaf level of the pGFMM
n_t ...	number of time steps in the leaves of the temporal tree
n_f ...	number of neighbor cubes in a linear direction

Table 1: Characteristic variables in the pGFMM.

the cluster they belong to. Hence we assign test- and trial functions to cubes according to their center position and enlarge their half side length by r_s^{max} such that the supports of the functions are fully contained within the cluster they belong to. Thus we generate overlapping cubes with half a side length in level ℓ_s given by

$$h_s^{(\ell_s)} = 2^{-\ell_s} h_s^{(0)} + r_s^{max}$$

and for reasonable spatial discretization parameters, i.e. $r_s^{max} \ll h_s^{(\ell_s)} \ll h_s^{(0)}$, we conclude

$$h_s^{(\ell_s)} \sim 2^{-\ell_s} h_s^{(0)} \,. \tag{9}$$

Here, $A \sim B$ means that there is a constant $c > 1$ such that $A/c \leq B \leq cA$, further $A \gtrsim B$ and $A \lesssim B$ mean $A \leq cB$ and $A \geq B/c$, respectively. The characteristic variables of the algorithm are summarized in Table 1.

In order to preserve uniformity of the time discretization (see Section 2), we group n_t time steps into leaf clusters with half length

$$h_t^{(L_t)} = \frac{n_t h_t}{2} \,. \tag{10}$$

Under this assumption we extend the time interval of interest $[0,T]$ to the root level interval $[0,T_0] \supset [0,T]$ with

$$T_0 = 2^{L_t} n_t h_t = 2h_t^{(0)} \geq T \tag{11}$$

leading to a temporal tree exceeding the original time interval. Then we compute the depth of this tree from (11)

$$L_t = \left\lceil \log_2 \left(\frac{T}{n_t h_t} \right) \right\rceil$$

and similar to (9) the ℓ_t^{th} level's interval half length

$$h_t^{(\ell_t)} = 2^{-\ell_t} h_t^{(0)} \,. \tag{12}$$

We now consider the expansion of the heat kernel in two space-time clusters. We assume that the source- and evaluation clusters are in the same space-time levels (ℓ_s, ℓ_t). Further, the space-time centers of the source and evaluation clusters are $(\bar{y}, \bar{\tau})$ and $(\bar{x}, \bar{t})$, respectively. We introduce local variables with the transformation

$$\begin{aligned} x &= \bar{x} + x' h_s^{(\ell_s)}, \quad y = \bar{y} + y' h_s^{(\ell_s)}, \quad -1 \le x_i', y_i' \le 1, \quad i \in \{1,2,3\} \\ t &= \bar{t} + t' h_t^{(\ell_t)}, \quad \tau = \bar{\tau} + \tau' h_t^{(\ell_t)}, \quad -1 \le t', \tau' \le 1. \end{aligned} \tag{13}$$

In the local variables, the heat kernel is

$$\begin{aligned} G_{loc}(x', y', t', \tau') &= (4\pi(t-\tau))^{-\frac{3}{2}} \exp\left(-\frac{|x-y|^2}{4(t-\tau)}\right) \\ &= \left(4\pi h_t^{(\ell_t)}(d + t' - \tau')\right)^{-\frac{3}{2}} \exp\left(-\rho \frac{|r + x' - y'|^2}{d + t' - \tau'}\right), \end{aligned} \tag{14}$$

and the constants $d > 0$, $r \in \mathbb{R}^3$ and $\rho > 0$ are given by

$$d = \frac{(\bar{t} - \bar{\tau})}{h_t^{(\ell_t)}}, \tag{15}$$

$$r = \frac{(\bar{x} - \bar{y})}{h_s^{(\ell_s)}}, \tag{16}$$

$$\rho = \frac{\left(h_s^{(\ell_s)}\right)^2}{4 h_t^{(\ell_t)}}. \tag{17}$$

The fast evaluation of potentials is based on separation of variables of the local heat kernel, which is accomplished by series expansion in the local variables [17, 19]. The series will converge rapidly only if the kernel is smooth and not too peaked. The form of (14) suggests that two space-time clusters are admissible if they are separated by either one or two time intervals in the same level in which case $d \in \{4, 6\}$. Moreover, the rate of convergence depends on the magnitude of the factor ρ, because it determines how peaked the local kernel is. To ensure the same accuracy of all cluster interactions, the space and time levels are selected such that ρ is uniformly bounded. From (9) and (17) it follows that

$$\rho \sim \frac{\left(2^{L_s - \ell_s} h_s^{(L_s)}\right)^2}{2^{L_t - \ell_t} 4 h_t^{(L_t)}} = 2^{2(L_s - \ell_s) - (L_t - \ell_t)} \rho_L. \tag{18}$$

where $\rho_L = {(h_s^{(L_s)})^2}/{4 h_t^{(L_t)}}$. Thus the factor ρ is bounded for all clusters as long as $L_s - \ell_s$ is smaller than half of $L_t - \ell_t$. The precise relationship is

$$\ell_s = \max\left(L_s - \left\lfloor \frac{L_t - \ell_t}{2} \right\rfloor, 0\right). \tag{19}$$

From (18) and (19) it follows that the value of ρ alternates between the values $\sim \rho_L$ and $\sim \rho_L/2$ when ℓ_t decreases until $\ell_s = 0$. After that, the value of ρ decreases and the geometry is only contained within one cube.

The factor ρ_L controls the accuracy of the parabolic FMM in two ways. First, ρ_L determines the convergence speed of the expansion of (14) in the local variables. For a fixed expansion order a smaller value of ρ_L will result in a smaller error. On the other hand, the heat kernel decays exponentially with increasing spatial distance. Therefore at most n_f neighboring clusters in each linear direction have to be included in a potential calculation. The resulting truncation error can be bounded as follows

$$\exp\left(-\rho\frac{|r+x'-y'|^2}{d+t'-\tau'}\right) \gtrsim \exp\left(-\frac{\rho_L}{2}\frac{(2n_f)^2}{d+2}\right) \geq \exp\left(-\rho_L\frac{n_f^2}{4}\right) . \tag{20}$$

We conclude with the following observations:

1. By increasing the factors n_f or ρ_L we control the spatial farfield truncation error.

2. Because of the scaling in (14) the expansion orders of the heat kernel must be increased when ρ_L is increased.

3. The kernels in the d^{th} step of the temporal nearfield, c.f. (8), decay in space like $\exp(-r^2/4dh_t)$. Distant interactions may be truncated if the resulting error is of the same magnitude as the bound given in (20). This implies that the cut-off radius satisfies

$$r = \sqrt{\rho_L n_f^2 d h_t} . \tag{21}$$

   If h_s decreases faster than $\sqrt{h_t}$, the number of panels per leaf level cube increases with mesh refinements affecting the complexity of the overall algorithm. In this case fast methods for the temporal nearfield must be employed, which will be described in the following section.

4 The FGT in the Temporal Nearfield

In the original parabolic FMM the interaction of well separated temporal clusters is evaluated using Chebyshev expansion of (14) in space and time, while the contribution of the first two leaf time clusters is computed directly, taking only nearby interactions into account. However, (21) implies that this is only efficient if the spatial mesh-width is not too small.

In applications, one is often interested in complicated geometries, which have to be resolved with a fine mesh, while still maintaining a relatively large time step size. In this case the direct calculation of the temporal nearfield becomes the dominant factor. To remedy this situation, we consider using a fast method for the temporal nearfield. Since the kernels (6) are integrated quantities of a Gaussian, it should be possible to use some modified form of the fast Gauss transform. However, there are two difficulties we encounter. First, the form of V_d is rather complicated, thus the translation operators of fast algorithms are difficult to derive and implement. Second, and more importantly, the kernel does not separate into its spatial directions as it is the case for the

heat kernel, which makes translating series expansion very expensive if a high spatial expansion order q is desired.

To resolve both issues, we do not use the analytic kernel (6) for the time steps in the temporal nearfield, but instead replace it by a kernel that is derived from a numerical quadrature rule. Thus we obtain a sequence of Gauss kernels for which one of the well known FGT algorithms can easily be adopted. The FGT [9] enables a fast computation of such Gauss transforms by evaluating the spatial influence of clustered sources onto clustered targets. Since the Gauss kernel is smooth everywhere and exponentially decaying, only a constant number of such close by cluster interactions within one level need to be evaluated. It is important to observe that due to the proposed quadrature in time, the spatial interactions in the FGT benefit form the fact that the variables of the kernel still separate into spatial directions, which enables an efficient implementation of M2L translations in tensor product form [20]. However, since the integrands leading to (6) are singular, or nearly singular in time, it is important that the quadrature method is chosen appropriately, which will be discussed in this section.

4.1 Numerical quadrature in time

This section discusses how the integrals (6) are evaluated by numerical quadrature. Since the integrand only depends on the difference $t'-\tau'$, it is possible to reduce the integration to one dimension. Moreover, the smoothness of the integrand depends on the difference $d=i-j$, hence, the quadrature rule will be adjusted accordingly. With the transformation

$$\xi = t'-\tau', \quad \eta = t'+\tau'$$

integral (6) becomes

$$V_d(\tilde{r}) = \frac{\sqrt{h_t}}{2}\int_A G(\tilde{r}, d+\xi)\mathrm{d}A\,, \quad d\geq 1\,,$$

where A is a diamond-shaped polygon in the ξ,η-plane with vertexes $(\pm 1,0)$ and $(0,\pm 1)$. For the case $d=0$ the same transformation as applied to (7) leads to

$$V_0(\tilde{r}) = \frac{\sqrt{h_t}}{2}\int_{A_0} G(\tilde{r}, \xi)\mathrm{d}A_0\,,$$

where A_0 is the triangle in the ξ,η-plane with vertexes $(1,0)$ and $(0,\pm 1)$. Since the η-integrals are trivial, they are evaluated analytically. It follows that

$$V_d(\tilde{r}) = \begin{cases} V_{d-1}^-(\tilde{r}) + V_d^+(\tilde{r})\,, & d\geq 1\,,\\ V_0^+(\tilde{r})\,, & d=0\,, \end{cases}$$

where

$$V_d^\pm(\tilde{r}) = \sqrt{h_t}\int_0^1 G(\tilde{r}, d^\pm+\xi)\varphi^\pm(\xi)\mathrm{d}\xi, \tag{22}$$

with $\varphi^+(\xi) := 1-\xi$, $\varphi^-(\xi) := \xi$, $d^+ := d$ and $d^- := d-1$. With these definitions the following two cases need to be distinguished.

The case $d^{\pm} \geq 1$ The integrand in (22) is smooth and a g^{th} order Gauss-Legendre rule can be applied. We transform integral (22) to the standard interval $[-1,1]$ and use the scaling (3) with $\lambda = 1/2$. Then

$$\begin{aligned} V_d^{\pm}(\hat{r}) &= \sqrt{2h_t} \int_{-1}^{1} G(\hat{r}, \hat{d}+x)\hat{\varphi}^{\pm}(x)\,\mathrm{d}x \\ &= \sqrt{2h_t} \sum_{j=0}^{g-1} G(\hat{r}, \hat{d}+x_j)\hat{\varphi}^{\pm}(x_j)w_j + \sqrt{2h_t}E_g^{\pm}(\hat{r}, \hat{d}) \end{aligned} \tag{23}$$

where x_j and w_j are the Gauss-Legendre quadrature points and weights, $E_g^{\pm}$ is the quadrature error and

$$\hat{r} = \sqrt{2}\tilde{r}, \quad \hat{d} = 2d^{\pm} + 1, \quad \text{and} \quad \hat{\varphi}^{\pm}(x) = \frac{1}{2}(1 \mp x).$$

Thus the kernel in (23) is expressed as a sum of kernels with variables separating into spatial directions.

The case $d^{\pm} = 0$ The integrands are singular and a composite quadrature rule must be applied. To this end, we introduce a new parameter $0 < \mu < 1$ and consider the following dyadic splitting of the interval

$$[0,1] = \bigcup_{m=0}^{M} I_m \tag{24}$$

where

$$I_M = [0, \mu^M] \quad \text{and} \quad I_m = \left[\mu^{m+1}, \mu^m\right], \quad m = 0, \ldots M-1.$$

The kernels $V_0^{\pm}(\tilde{r})$ can now be written as a sum of the kernels

$$V_{0,m}^{\pm}(\tilde{r}) := \sqrt{h_t} \int_{I_m} G(\tilde{r}, x)\varphi^{\pm}(x)\,\mathrm{d}x, \quad m = 0, \ldots M. \tag{25}$$

This splitting is motivated by the following considerations. The factor M is selected such that the kernel $V_{0,M}^{\pm}(\tilde{r})$ is sufficiently local in space and hence, the layer potentials with this kernel can be evaluated directly with $O(N_s)$ cost because only nearby interactions must be computed. The integrands of the remaining kernels $V_{0,m}^{\pm}(\tilde{r})$ for $m < M$ are smooth, and approximated by a Gauss-Legendre quadrature rule, similar to (23). Transforming I_m to the standard interval leads to

$$\begin{aligned} V_{0,m}^{\pm}(\hat{r}) &= \frac{\sqrt{2h_t}}{\sqrt{\mu^m - \mu^{m+1}}} \int_{-1}^{1} G(\hat{r}, \hat{d}+x)\hat{\varphi}^{\pm}(x)\,\mathrm{d}x \\ &= \frac{\sqrt{2h_t}}{\sqrt{\mu^m - \mu^{m+1}}} \sum_{j=0}^{g-1} G(\hat{r}, \hat{d}+x_j)\hat{\varphi}^{\pm}(x_j)w_j + \frac{\sqrt{2h_t}}{\sqrt{\mu^m - \mu^{m+1}}} E_{m,g}^{\pm}(\hat{r}, \hat{d}), \end{aligned} \tag{26}$$

where

$$\hat{r} = \frac{\sqrt{2}\tilde{r}}{\sqrt{\mu^m - \mu^{m+1}}}, \quad \hat{d} = \frac{1+\mu}{1-\mu}$$

and

$$\begin{aligned}\hat{\varphi}^{+}(x) &= 1-\left(\frac{\mu^m+\mu^{m+1}}{2}+\frac{\mu^m-\mu^{m+1}}{2}x\right),\\ \hat{\varphi}^{-}(x) &= \frac{\mu^m+\mu^{m+1}}{2}+\frac{\mu^m-\mu^{m+1}}{2}x.\end{aligned}$$

The accuracy of the quadrature is controlled by the parameters μ, $\hat{d}$, M and g. To find reasonable choices we must understand how they influence the quadrature error. This will be done next.

4.2 Gauss-Legendre quadrature

Estimates of the error of the Gauss-Legendre quadrature applied to functions with analytic extension into the complex plane have been derived in [3]. Unfortunately, the results there assume that either the quadrature order or the region of analyticity are sufficiently large and are thus not directly applicable when these values are specified, as needed in this work. However, the methodology can be modified to derive a similar result which is more suitable for the following discussion.

For a function $f:[-1,1]\to\mathbb{R}$ that has an analytic extension into the complex plane the error $E_g(f)$ of the Gauss-Legendre rule

$$\int_{-1}^{1} f(x)\mathrm{d}x = \sum_{j=0}^{g-1} f(x_j)w_j + E_g(f)$$

can be expressed in terms of an integral over a simply closed contour γ that encloses the interval $[-1,1]$ and is contained in a region of the complex plane where f is analytic

$$E_g(f) = \frac{1}{\pi i}\int_{\gamma}\frac{Q_g(z)}{P_g(z)}f(z)\mathrm{d}z \tag{27}$$

and $Q_g(z)$ is the Legendre function of the second kind, defined by

$$Q_g(z) = \frac{1}{2}\int_{-1}^{1}\frac{P_g(x)}{z-x}\mathrm{d}x, \quad z\notin[-1,1]\,.$$

Because of the properties of the Legendre polynomials $P_g(x)$ it is convenient to let the contour γ be the ellipse ε_ρ that is the image of the circle of radius ρ in the complex ζ-plane under the transformation $z = 1/2(\zeta+1/\zeta)$, i.e.

$$\varepsilon_\rho = \left\{\frac{1}{2}\left(\zeta+\frac{1}{\zeta}\right) : |\zeta|=\rho\right\}. \tag{28}$$

Lemma 4.1. *If f is analytic in a domain that contains ε_ρ for some $\rho > \sqrt{2}$, then the error of the Gauss-Legendre quadrature can be bounded by*

$$|E_g(f)| \le C_0 \sqrt{g} \frac{\rho^2}{\rho^2 - 2} M(\rho) \rho^{-2g}. \tag{29}$$

where $M(\rho) = \max |f(z)|$ on ε_ρ and $C_0 = \frac{3}{2}\sqrt{\pi}\exp\left(\frac{1}{6}\right) = 3.1408\ldots$.

Proof. It is well known that the Legendre polynomials have the representation

$$P_g(z) = \sum_{j=0}^{g} a_{g-j} a_j \zeta^{g-2j}, \quad a_j = \frac{(2j)!}{(j!)^2 4^j},$$

moreover, $a_{g-j} a_j \le a_g$, see [6, Section 12.4]. Using Stirling's formula, we can estimate

$$\frac{1}{\sqrt{\pi j}} e^{\frac{-1}{6j}} \le a_j \le \frac{1}{\sqrt{\pi j}} e^{\frac{1}{24j}}, \quad j \ge 1.$$

We have

$$P_g(z) = a_g \zeta^g \left(1 + \sum_{j=1}^{g} \frac{a_{g-j} a_j}{a_g} \zeta^{-2j} \right),$$

and thus for $z \in \varepsilon_\rho$ and $\rho > \sqrt{2}$

$$|P_g(z)| \ge a_g \rho^g \left(1 - \sum_{j=1}^{g} \rho^{-2j} \right) \ge a_g \rho^g \frac{\rho^2 - 2}{\rho^2 - 1}.$$

An estimate of the Legendre functions of the second kind is given in [3, eq. (15)]

$$|Q_g(z)| \le \frac{2\rho}{\rho^2 - 1} \rho^{-g}, \quad z \in \varepsilon_\rho.$$

From the contour integral formula (27) it follows with the length $\ell(\varepsilon_\rho)$ of ε_ρ that

$$|E_g(f)| \le \frac{\ell(\varepsilon_\rho)}{\pi} \frac{2}{a_g} \frac{\rho}{\rho^2 - 2} M(\rho) \rho^{-2g} = C_\rho^g \sqrt{g} \frac{\rho^2}{\rho^2 - 2} M(\rho) \rho^{-2g}$$

where

$$C_\rho^g = \frac{\ell(\varepsilon_\rho)}{\pi\rho} \frac{2}{\sqrt{g} a_g}.$$

Clearly, the factor C_ρ^g is uniformly bounded in g and ρ. The given upper bound C_0 can be derived with elementary arguments. □

Lemma 4.2. *Let $\hat{d} \ge 2$, $g \ge 1$ and $\varphi(x) = \alpha + \beta x$ be a linear function with coefficients $|\alpha| \le 1$ and $|\beta| \le 1$. Then for the function $f(x) = G(r, \hat{d} + x)\varphi(x)$ the estimate*

$$|E_g(f)| \le C_1 g^2 \hat{d} \left(\hat{d} + \sqrt{\hat{d}^2 - 1} \right)^{-2g}$$

holds, where $C_1 = \frac{4^5}{9} C_0$.

Proof. Since $f(x)$ has a singularity for $x = -\hat{d}$ the value of ρ for the ellipse ε_ρ in Lemma 4.1 must be in the interval $\rho \in (\sqrt{2}, \hat{d} + \sqrt{\hat{d}^2 - 1})$. To obtain a value that leads to a tight error bound we choose ρ with the goal to minimize $M(\rho)\rho^{-2g}$. To that end, we note that for $z \in \varepsilon_\rho$ the estimate

$$\left| \frac{1}{(\hat{d}+z)^{\frac{3}{2}}} \exp\left(-\frac{r^2}{\hat{d}+z}\right) \right| \leq \left(\hat{d} - \frac{1}{2}\left(\rho + \frac{1}{\rho}\right)\right)^{-\frac{3}{2}}$$

holds, for any r, hence we minimize $(\hat{d} - \frac{1}{2}(\rho + \frac{1}{\rho}))^{-\frac{3}{2}}\rho^{-2g}$. Simple calculus shows that the optimal ρ is $\rho_0(x_0)$ defined by

$$\rho_0(x) = \frac{\hat{d} + \sqrt{\hat{d}^2 - 1 + x^2}}{1+x} \quad \text{and} \quad x_0 = \frac{3}{4g}.$$

It is easy to see that the function $\rho_0(x)$ is monotonically decreasing for positive x, hence $\rho_0(x_0)$ is in the aforementioned interval of ρ. Moreover $\rho_0(x)$ is concave up

$$\rho_0(x) \geq \left(\hat{d} + \sqrt{\hat{d}^2 - 1}\right)(1 - x),$$

where the right hand side is the linearization of $\rho_0(x)$ at $x = 0$. Thus we can estimate

$$\rho_0^{-2g}(x_0) \leq \left(\hat{d} + \sqrt{\hat{d}^2 - 1}\right)^{-2g} \left(1 - \frac{3}{4g}\right)^{-2g}. \tag{30}$$

The second term is a monotonically decreasing function of g and hence can be bounded by 16 when $g \geq 1$.

Since $t \to \frac{1}{2}(t + \frac{1}{t})$ is increasing and concave up for $t > 1$ and $\rho_0(x)$ is concave up, it follows that the composition $d_0(x) := \frac{1}{2}(\rho_0(x) + \frac{1}{\rho_0(x)})$ is of also concave up, hence the estimate

$$d_0(x) \leq d_0(0) + (d_0(1) - d_0(0))x = \hat{d} - \frac{1}{2}\left(\hat{d} - \frac{1}{\hat{d}}\right)x$$

follows, which implies that

$$\left(\hat{d} - d_0(x_0)\right)^{-\frac{3}{2}} \leq x_0^{-\frac{3}{2}} \left(\frac{1}{2}\left(\hat{d} - \frac{1}{\hat{d}}\right)\right)^{-\frac{3}{2}} = g^{\frac{3}{2}} \left(\frac{3}{8}\left(\hat{d} - \frac{1}{\hat{d}}\right)\right)^{-\frac{3}{2}}. \tag{31}$$

Since we assume that $\hat{d} \geq 2$ the second term is bounded by $(4/3)^3$.

With $\rho_0(x_0) > \rho_0(1) = \hat{d}$ and $\hat{d} \geq 2$ we also have the estimate

$$\frac{\rho_0^2(x_0)}{\rho_0^2(x_0) - 2} \leq \frac{\hat{d}^2}{\hat{d}^2 - 2} \leq 2. \tag{32}$$

Moreover, for $z \in \varepsilon_\rho$ we have

$$\begin{aligned} |\varphi(z)| &\leq 1 + |z| \leq 1 + \frac{1}{2}\left(\rho_0(x_0) + \frac{1}{\rho_0(x_0)}\right) = 1 + d_0(x_0) \\ &\leq 1 + d_0(0) = 1 + \hat{d} = \hat{d}\left(1 + \frac{1}{\hat{d}}\right), \end{aligned} \tag{33}$$

where the second term can be estimated by $3/2$. Finally, the assertion follows from Lemma 4.1 and the estimates (30)-(33). □

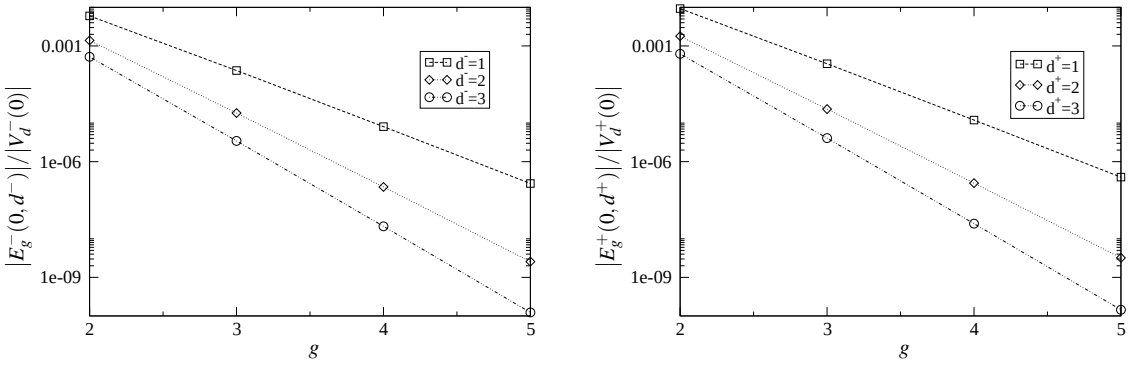
The integrals for the kernels $V_d^{\pm}(\tilde{r})$ and $V_{0,m}^{\pm}(\tilde{r})$ are of the form of Lemma 4.2. Note that the estimates suggest that the largest error occurs for $\tilde{r} = 0$, which we also observe numerically. We briefly illustrate the behavior of the actually computed error below.

The case $d^{\pm} \geq 1$ Note that for we have $\hat{d} = 2d^{\pm} + 1 = 3, 5, 7 \ldots$ and thus Lemma 4.2 directly applies as is confirmed by the semi-logarithmic plots in Figure 1.

The case $d^{\pm} = 0$ In order to reproduce the error of the kernels $V_d^{\pm}(\hat{r})$ in the worst case we must have $\hat{d} = 3$, thus $\mu = 1/2$. However, we also have to account for the factor $(\mu^m - \mu^{m+1})^{-\frac{1}{2}}$ in (26), which we do by investigating the error of the composite quadrature rule given by

$$\sqrt{2h_t} \sum_{m=0}^{M-1} \left(\mu^m - \mu^{m+1}\right)^{-\frac{1}{2}} E_{m,g}^{\pm}(\hat{r}, \hat{d}) \,.$$

In Table 2 we observe that the error decays exponentially in g as predicted by Lemma 4.2, while it grows only very slowly with M.



(a) $\left|E_g^-(0,d^-)\right|/\left|V_d^-(0)\right|$ vs. g (b) $\left|E_g^+(0,d^+)\right|/\left|V_d^+(0)\right|$ vs. g

Figure 1: Maximal quadrature error at $\hat{r} = 0$ vs. quadrature order.

	$M=1$	$M=4$	$M=7$	$M=10$	$M=13$
$g=2$	9.70×10^{-5}	2.48×10^{-4}	3.02×10^{-4}	3.21×10^{-4}	3.27×10^{-4}
$g=3$	2.42×10^{-6}	6.21×10^{-6}	7.55×10^{-6}	8.02×10^{-6}	8.19×10^{-6}
$g=4$	6.31×10^{-8}	1.61×10^{-7}	1.96×10^{-7}	2.08×10^{-7}	2.13×10^{-7}
$g=5$	1.68×10^{-9}	4.31×10^{-9}	5.24×10^{-9}	5.57×10^{-9}	5.69×10^{-9}

Table 2: $\sqrt{2h_t}\sum_{m=0}^{M-1}\left(\mu^m-\mu^{m+1}\right)^{-\frac{1}{2}}\left|E_{g,m}^-(0,\hat{d})\right|/\left|V_{0,m}^-(0)\right|$ for $\mu = 1/2$.

Choosing M The operator $V_{0,M}^{\pm}(\tilde{r})$ is evaluated directly, by adding panel contributions instead of resorting to series expansions. From (25) it can be seen that the decay in space is like $\exp(-r^2/4h_t\mu^M)$. To obtain linear complexity in N_s only contributions in a neighborhood of radius proportional to h_s should be considered. The resulting truncation error should be of the same order as the farfield error in the coarser levels. A comparison with (20) shows that, if the nearfield radius is $n_f h_s$, then M should satisfy

$$\frac{(n_f h_s)^2}{4h_t\mu^M} \sim \rho_L \frac{n_f^2}{4}.$$

This leads to

$$M \sim \log_\mu \left(\frac{h_s^2}{h_t \rho_L} \right). \tag{34}$$

4.3 Spatial level in FGT

From the previous discussion it follows that the kernel $V_d(\tilde{r})$ can be approximated by some composite of the Gauss-Legendre quadrature rule. We write this as

$$V_d(\hat{r}) \approx \begin{cases} \sum_{j=0}^{g-1} G(\hat{r}, 2d-1+x_j)\hat{w}_j^- + \sum_{j=0}^{g-1} G(\hat{r}, 2d+1+x_j)\hat{w}_j^+, & d \geq 2, \\ V_{0,M}^-(\hat{r}) + \sum_{m=0}^{M-1}\sum_{j=0}^{g-1} G(\hat{r}, \frac{1+\mu}{1-\mu}+x_j)\hat{w}_{j,m}^- + \sum_{j=0}^{g-1} G(\hat{r}, 3+x_j)\hat{w}_j^+, & d = 1, \\ V_{0,M}^+(\hat{r}) + \sum_{m=0}^{M-1}\sum_{j=0}^{g-1} G(\hat{r}, \frac{1+\mu}{1-\mu}+x_j)\hat{w}_{j,m}^+, & d = 0. \end{cases}$$

where $x_j \in [-1,1]$ are the Gauss points and $\hat{w}_j^{\pm}$ and $\hat{w}_{j,m}^{\pm}$ are a combination of the Gauss weights, $\hat{\varphi}^{\pm}(x_j)$ and the pre-factors in (23) and (26). The layer operators with singular kernels $V_{0,M}^{\pm}(\hat{r})$ are local and evaluated directly, while the kernels $G(\hat{r}, d+x_j)$ are Gaussians and evaluated using the FGT.

We now determine the spatial level in which the fast Gauss transform with a given variance δ should be computed. Since nearby temporal interactions are more peaked, the interactions have to be computed in a finer spatial level than the smooth part of the pGFMM. Therefore we introduce uniform refinements above level L_s, where the half length of a cube follows from (9), too.

The appropriate level for a given variance δ is again determined from the scaling in the exponential function. We have

$$\exp\left(-\frac{|x-y|^2}{\delta} \right) = \exp\left(-\frac{\left(h_s^{(\ell_s)}\right)^2}{\delta} \left| r' + x' - y' \right|^2 \right)$$

and as in the pGFMM we drop cube interactions in the FGT that are separated by more than n_f cube lengths. The truncation error for the argument of the exponential function can then be bounded from below by

$$\frac{|x-y|^2}{\delta} \geq 2^{-2\ell_s} \frac{\left(h_s^{(0)}\right)^2}{\delta} (2n_f)^2.$$

This value should be larger than the corresponding value for the pGFMM in (20). From (10) and (18) it follows that the spatial level L_s for the FGT is

$$\mathsf{L}_s \leq L_s + \log_4\left(\frac{32 n_t h_t}{\delta}\right). \tag{35}$$

By setting

$$\delta = \begin{cases} 4h_t(d+1), & d \geq 1 \\ 4h_t\mu^m, & d = 0 \end{cases}$$

in (35) as an upper bound for all quadrature nodes we finally get

$$\mathsf{L}_s = \begin{cases} L_s + \left\lfloor \log_4\left(\frac{8n_t}{d+1}\right)\right\rfloor, & d \geq 1 \\ L_s + \left\lfloor \log_4\left(\frac{8n_t}{\mu^m}\right)\right\rfloor, & d = 0\,. \end{cases} \tag{36}$$

5 Numerical Results

In this section we investigate initial Dirichlet boundary value problems for $\Omega \subset \mathbb{R}^3$ with boundary S over the time interval $I := [0, 0.5]$ by a direct Galerkin boundary integral approach. Thus seek the solution $q_h \in Q_h^{d_s,d_t}(S \times I)$ of

$$\langle \mathcal{V} q_h, w_h\rangle_{S\times I} = \langle \left(\tfrac{I}{2} + \mathcal{K}\right) g_D, w_h\rangle_{S\times I} \quad \forall w_h \in Q_h^{d_s,d_t}(S\times I) \tag{37}$$

for which the a priori estimate [15, eq. (7.25)]

$$\|q - q_h\|_{L^2(S\times I)} \leq C \max\left[\left(\frac{h_s^2}{h_t}\right)^{\frac{1}{4}}, \left(\frac{h_s^2}{h_t}\right)^{-\frac{1}{4}}\right](h_s^{d_s+1} + h_t^{d_t+1})\|q\|_{H^{d_s+1,d_t+1}(S\times I)} \tag{38}$$

holds. In all examples we impose an impulse point source located at $x_0 = (1.5, 1.5, 1.5)^\top \notin \Omega$ of the form $g_D(y,t) = G(x_0 - y, t)$ with $(y,t) \in (S \times I)$. The simple motivation for this choice lies in the fact that in this case we know the exact solution of (37), namely $q(y,t) = \partial/\partial_{n_y} G(x_0 - y, t)$ with $(y,t) \in (S \times I)$, which allows us to verify the estimate (38). Moreover, we also know the exact solution of the initial Dirichlet boundary value problem in the interior, i.e. $u(\tilde{y}, t) = G(x_0 - \tilde{y}, t)$ with $(\tilde{y}, t) \in (\Omega \times I)$, which enables us to compute the error at the interior, too.

In the following we divide this section into two types of examples regarding the space-time scaling condition

$$h_s^\alpha \sim h_t\,. \tag{39}$$

In Subsection 5.1 we show that our method maintains the optimal convergence rate at optimal complexity for $\alpha = 2$, in which case (38) simplifies to

$$\|q - q_h\|_{L^2(S\times I)} \leq C(h_s^{d_s+1} + h_t^{d_t+1})\|q\|_{H^{d_s+1,d_t+1}(S\times I)}\,. \tag{40}$$

We would like to emphasize that the Galerkin scheme does not require smooth domains, but can deal with piecewise Lipschitz domains, too. Therefore we choose our domain to be a cube, where we can still show (40).

In Subsection 5.2 we intend to demonstrate that our method maintains optimal complexity even for $\alpha \ll 2$. Such a setting occurs in applications where complicated geometries need to be resolved, while a rather coarse time discretization may still be sufficient to approximate the solution. We investigate the limit where $\alpha \to 0$ and thus (38) reduces to

$$\|q - q_h\|_{L^2(S\times I)} \leq C\|q\|_{H^{d_s+1,d_t+1}(S\times I)}\,. \tag{41}$$

This limit motivates the dyadic splitting and the application of the composite quadrature to approximate the time integrals of the Galerkin scheme with the FGT in space to solve the problem efficiently. Moreover, it shows that the Galerkin scheme does not suffer from the *geometrical stiffness* reported in [12].

Finally, we would like to remark that the solution of the linear systems (8) was obtained by using a CG solver with a block diagonal preconditioner.

5.1 Heat diffusion with uniform refinement in space and time

In this subsection we solve (37) with $\Omega = [-0.5, 0.5]^3$. Even though the estimate (40) for a uniform refinement in space and time with $\alpha = 2$ suggest to choose the polynomial ansatz orders to be $d_s = 1$ and $d_t = 0$ in order to achieve quadratic convergence in the spatial variable, we show results for the simpler case of $q_h \in Q_h^{0,0}$ yielding linear order of convergence. In order to show faster convergence, we evaluate the representation formula for $u(\tilde{x},t)$ with $(\tilde{x},t) \in (\Omega \times I)$ and compute the *rel.* $\ell_2(\mathfrak{I}\times\mathfrak{T})$ error $\|u-u_h\|_{\ell_2(\mathfrak{I}\times\mathfrak{T})}/\|u\|_{\ell_2(\mathfrak{I}\times\mathfrak{T})}$ with

$$\|u-u_h\|_{\ell_2(\mathfrak{I}\times\mathfrak{T})} := \left(\sum_{t_j\in\mathfrak{T}}\sum_{\tilde{x}_i\in\mathfrak{I}} \left(u(\tilde{x}_i,t_j) - u_h(\tilde{x}_i,t_j)\right)^2\right)^{\frac{1}{2}} \tag{42}$$

where we choose a fixed set $\mathfrak{I} := \{\tilde{x} : \text{vertices of } [-0.25, 0.25]^3\} \subset \Omega$ of spatial points and a fixed set $\mathfrak{T} := \{j/64\}_{j=1}^{32} \subset I$ of temporal points.

The coarsest discretization of $S \times I$ in our example consists of a uniform mesh with $N_s = 384$ triangles and $N_t = 32$ equidistant time steps. With $n_t = 1$ this leads to $L_t = 5$ while we start with the spatial root luster $L_s = 0$ leading to $\rho_L = 8$. It turns out that we can control the truncation error by neglecting all interactions from spatial cubes further apart than $n_f = 3$ and the interpolation error by choosing the spatial- and temporal expansion orders to be $q = 29$ and $p = 7$, respectively. These values had to be chosen this high, to resolve the error at the interior, while for the error on the boundary much lower values would have been sufficient.

lev.	N_sN_t	L_s/L_t	*rel.* $L_2(S\times I)$	*rel.* $\ell_2(\mathfrak{I}\times\mathfrak{T})$	it.	sol. [sec]	mem. [GB]
0	12,288	0/5	$1.23\ 10^{-1}$	$8.08\ 10^{-4}$	13	$1.92\ 10^{2}$	$1.48\ 10^{-1}$
1	196,608	1/7	$6.02\ 10^{-2}$	$2.57\ 10^{-5}$	13	$6.12\ 10^{2}$	$4.60\ 10^{-1}$
2	3,145,728	2/9	$2.99\ 10^{-2}$	$3.22\ 10^{-6}$	12	$1.08\ 10^{4}$	$1.99\ 10^{0}$
3	50,331,648	3/11	$1.49\ 10^{-2}$	$5.70\ 10^{-7}$	12	$1.85\ 10^{5}$	$1.07\ 10^{1}$

Table 3: pGFMM results for uniform refinement in space and time.

Table 3 shows the results for this uniform refinement in space and time. We compare the total number of unknowns $N_s N_t$ with the relative $L_2(S \times I)$ error, the relative $\ell_2(\mathfrak{I} \times \mathfrak{T})$ error, the solution time and the memory requirement, respectively. The relative $L_2(S \times I)$ error clearly reproduces the theoretical convergence rate $O(\sqrt{h_t})$ of (40), while the relative $\ell_2(\mathfrak{I} \times \mathfrak{T})$ shows $O(h_t)$ convergence of the interior solution. With $n_f = 3$ the spatial truncation comes into play at $\ell_s = 3$. Hence, we should be able to observe optimal computational complexity $O(N_s N_t)$ as is confirmed in Figure 2. Moreover, due to causality we only need to keep track of moment- and local expansions in two temporal clusters per level, i.e. neglecting logarithmic terms the memory requirement behaves like $O(N_s)$.

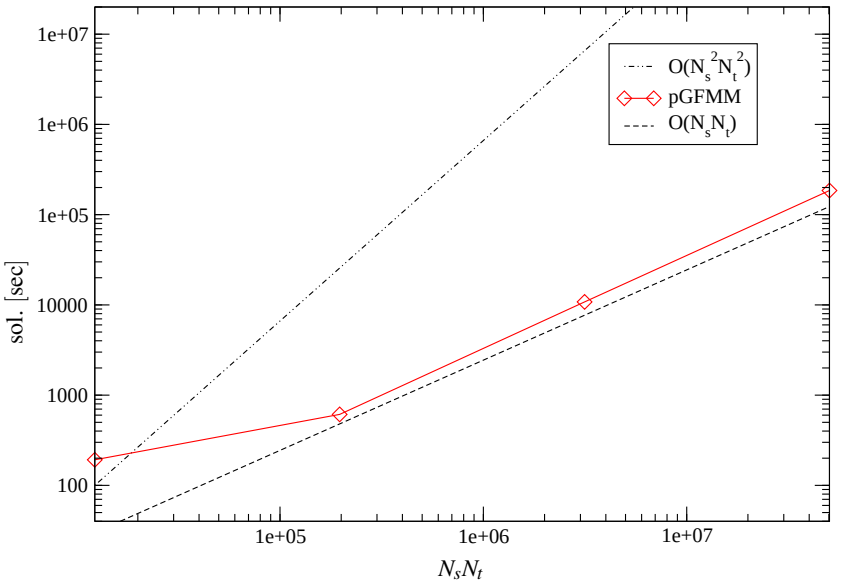


Figure 2: Solution time vs. number of unknowns for uniform refinement in space and time.

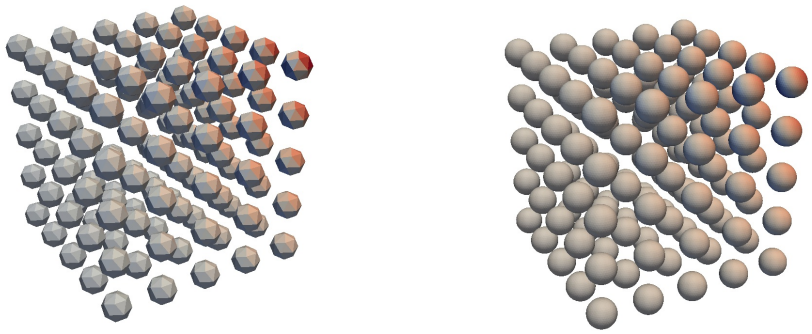
5.2 Heat diffusion with refinement in space and fixed time discretization

In this subsection we investigate different spatial refinements for a fixed temporal discretization. Violating the scaling in (39) with $\alpha \ll 2$ results in a smaller and smaller ρ_L. Hence, we conclude from (20) that we can not truncate the kernel in the computation of the temporal nearfield, which results in $O(N_s^2)$ complexity in terms of computation and storage. However, with the approach proposed in Section 4 we can keep n_f constant and derive from (34) with $\mu = 1/2$ that the number of FGTs in the temporal nearfield is $O(\log_2 N_s)$. Neglecting this logarithmic part and since we have localized the direct evaluations to $O(N_s)$ we indeed end up with an $O(N_s)$ algorithm. Furthermore, from Figure 1 and Table 2 we conclude that $g = 4$ should be sufficient for all quadratures in the temporal nearfield.

5.2.1 Uniform spatial spatial refinement of 125 spheres

We solve (37) for 125 spheres with a diameter of 0.125, placed on a uniform grid with spacings of 0.25, and centered at the origin. In refinement level zero each sphere is discretized by $N_{tr} = 36$ triangles. This is clearly too coarse of an approximation of the spheres, hence, in each further level they are uniformly refined (see Figure 3). We choose $L_s/L_t = 0/5$ with $n_t = 1$ to get $\rho_L = 8$ and find that $n_f = 2$ is sufficient to control the truncation error. Moreover, the expansion orders $p/q = 3/23$ lead to a bounded interpolation error.

In level zero we choose $M = 2$ to localize the direct evaluations of the nearfield. With $\mathsf{L}_s = 2$ each cluster for the direct evaluation of the nearfield contains approximately two spheres in refinement level zero, while in all subsequent levels M follows from (34) due to the refinement in space. Contrary to the example with uniform refinement in space and time in Subsection



(a) coarsest discretization (36 triangles/sphere) (b) finest discretization (2304 triangles/sphere)

Figure 3: Heat diffusion through different spatial refinements of 125 spheres on a uniform grid.

5.1 we observe that the number of iterations does not remain constant in this example. Thus in Figure 4 we plot the solution time/iterations vs. the number of spatial unknowns which shows the optimal complexity of the method. Moreover, we observe that the relative $L_2(S \times I)$ error is bounded, which is in agreement with (41).

lev.	N_{tr}	$N_s(N_t = 32)$	M	*rel.* $L_2(S \times I)$	it.	sol. [sec]	mem. [GB]
0	36	4,500(32)	2	$3.63\ 10^{-2}$	13	$1.17\ 10^{2}$	$1.09\ 10^{0}$
1	144	18,000(32)	4	$2.51\ 10^{-2}$	17	$1.41\ 10^{3}$	$4.84\ 10^{0}$
2	576	72,000(32)	6	$2.23\ 10^{-2}$	20	$6.60\ 10^{3}$	$2.09\ 10^{1}$
3	2304	288,000(32)	8	$2.11\ 10^{-2}$	26	$3.55\ 10^{4}$	$1.00\ 10^{2}$

Table 4: pGFMM results for 125 spheres with spatial refinement (*with* fast nearfield evaluation).

For comparison we show in Table 5 the results obtained *without* the fast nearfield evaluation presented in Section 4. In refinement level two we already observe the quadratic complexity, which is the reason why we could not solve level three efficiently anymore. Another observation is that

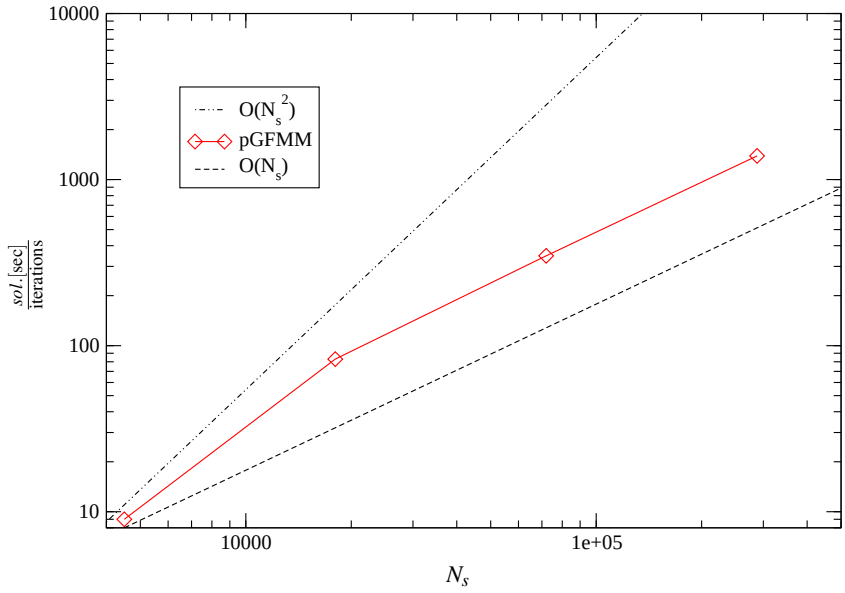


Figure 4: Solution time/iterations vs. number of spatial unknowns N_s for uniform spatial refinement of 125.

without the fast nearfield evaluation the number of iterations in the CG is bounded (compare columns 6 in Table 5 and 4). This is due to the fact that with the fast nearfield evaluation we also weaken the block diagonal preconditioner used within the CG solver.

lev.	N_{tr}	$N_s(N_t = 32)$	M	$rel.\ L_2(S \times I)$	it.	sol. [sec]	mem. [GB]
0	36	4,500(32)	0	$3.63\ 10^{-2}$	12	$3.98\ 10^{1}$	$1.06\ 10^{0}$
1	144	18,000(32)	0	$2.51\ 10^{-2}$	13	$2.77\ 10^{2}$	$9.88\ 10^{0}$
2	576	72,000(32)	0	$2.23\ 10^{-2}$	13	$1.47\ 10^{4}$	$1.34\ 10^{2}$

Table 5: pGFMM results for 125 spheres with spatial refinement (*without* fast nearfield evaluation).

5.2.2 Different number of spheres (N_{sp}) with const. number of triangles per sphere

In this example we want solve (37) for different agglomerations of spheres with a constant number of triangles per sphere, i.e. we place more and more spheres with a constant ratio of radius/distance on a uniform grid in $[-0.5, 0.5]^3$. Again we want to keep the temporal discretization fixed by splitting $I := [0, 0.5]$ into $N_t = 8$ equidistant time steps.

With such a refinement scheme we get $N_s = O(N_{sp})$ instead of $N_s = O(h_s^{-2})$. Hence we can not determine M from (34) but we have $M = O(\log_\mu N_{sp}^{-1})$ and with $\mu = 1/2$ we get $M = O(\log_2 N_{sp})$. Therefore, the spatial level where to perform the direct evaluations for the nearfield jumps with every other M which somehow dictates the refinement scheme.

We start by placing one sphere with $r = 0.4$ at the center of the unit cube and discretize it with 642 uniform triangles. In each subsequent refinement level we double the number of spheres in

linear direction scaling their radius by a factor of $1/2$ and keeping the number of elements per sphere constant. This way we end up with more and more spheres. We choose $L_s/L_t = 0/3$ to end up with $\rho_L = 2$ and compared to Example 5.2.1 we find that $n_f = 3$ and $q/p = 19/3$ are sufficient to control the truncation- and interpolation error, respectively.

Figure 5 and Table 6 illustrate that we can also maintain almost optimal complexity if we increase the number of geometric features rather than refine the mesh.

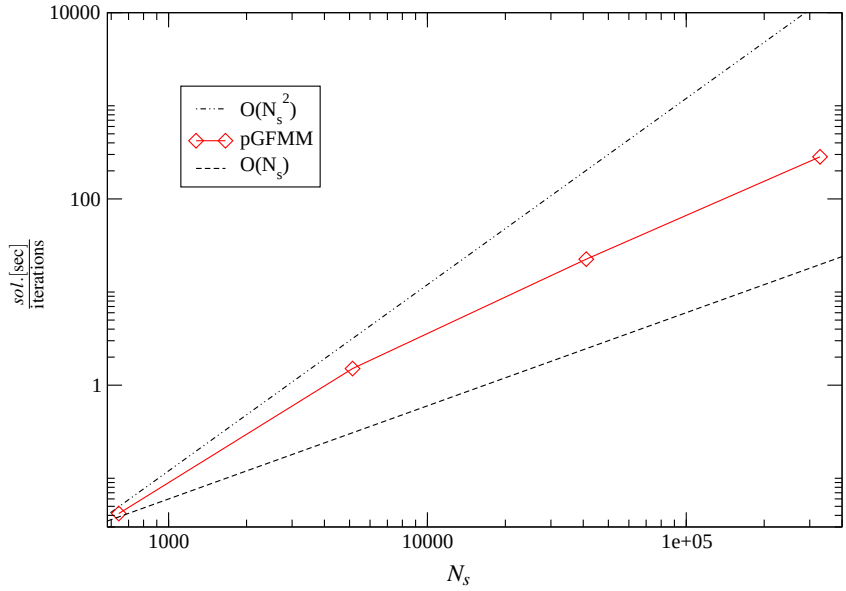


Figure 5: Solution time/iterations vs. number of spatial unknowns N_s for a different number N_{sp} of spheres.

lev.	N_{sp}	$N_s(N_t = 8)$	M	rel. $L_2(S \times I)$	it.	sol. [sec]	mem. [GB]
0	1	642(8)	0	$9.04\ 10^{-2}$	11	$1.17\ 10^{2}$	$9.67\ 10^{-2}$
1	8	5,136(8)	2	$8.52\ 10^{-2}$	15	$3.36\ 10^{2}$	$1.10\ 10^{0}$
2	64	41,088(8)	4	$8.33\ 10^{-2}$	25	$1.10\ 10^{3}$	$1.72\ 10^{1}$
3	512	328,704(8)	6	$8.31\ 10^{-2}$	41	$1.21\ 10^{4}$	$1.95\ 10^{2}$

Table 6: pGFMM results for a different number N_{sp} of spheres.

5.2.3 Ellipsoid with increasing curvature

In our last example we solve (37) where we choose our domain Ω to be the ellipsoid

$$\left(\frac{x_1}{\varepsilon}\right)^2 + \left(\frac{x_2}{\varepsilon}\right)^2 + x_3^2 < \frac{1}{4}$$

and $\varepsilon^{-1} \in \{1,2,4,8,16\}$. Since the curvature of the domain increases we have to refine the spatial discretization accordingly, i.e. we increase $N_s = O(\varepsilon^{-1})$, while we keep the temporal

discretization constant at $N_t = 32$. We choose the space time clustering parameters to be $L_s/L_t = 0/5$ leading to $\rho_L = 8$ and therefore $p/q = 3/23$ and $n_f = 2$ as in Subsection 5.2.1. With $\mu = 1/2$ we start with an initial $M = 0$, which we increase with every refinement level. This leads to an additional spatial cluster level in every other refinement. Thus the number of elements per FGT cluster is bounded, which leads to an optimal computational cost for the evaluation of the nearfield.

lev.	ε^{-1}	$N_s(N_t = 32)$	M	$rel.\ L_2(S \times I)$	it.	sol.[sec]	mem.[GB]
0	1	600(32)	0	$8.08\ 10^{-2}$	7	$5.75\ 10^{1}$	$7.61\ 10^{-2}$
1	2	1,210(32)	1	$6.09\ 10^{-2}$	10	$8.29\ 10^{1}$	$1.58\ 10^{-1}$
2	4	2,178(32)	2	$5.68\ 10^{-2}$	11	$1.67\ 10^{2}$	$3.01\ 10^{-1}$
3	8	4,312(32)	3	$5.52\ 10^{-2}$	13	$3.60\ 10^{2}$	$7.72\ 10^{-1}$
4	16	8,822(32)	4	$5.43\ 10^{-2}$	15	$1.02\ 10^{3}$	$2.02\ 10^{0}$
5	32	17,024(32)	5	$5.38\ 10^{-2}$	18	$2.20\ 10^{3}$	$4.62\ 10^{0}$

Table 7: pGFMM results for ellipsoid with increasing curvature.

From the data in Table 7 it can be concluded that the complexity is again almost optimal. Moreover, we see that while the spatial mesh must be refined with increasing curvature, it does not appear to be necessary to refine the temporal step size at the same time.

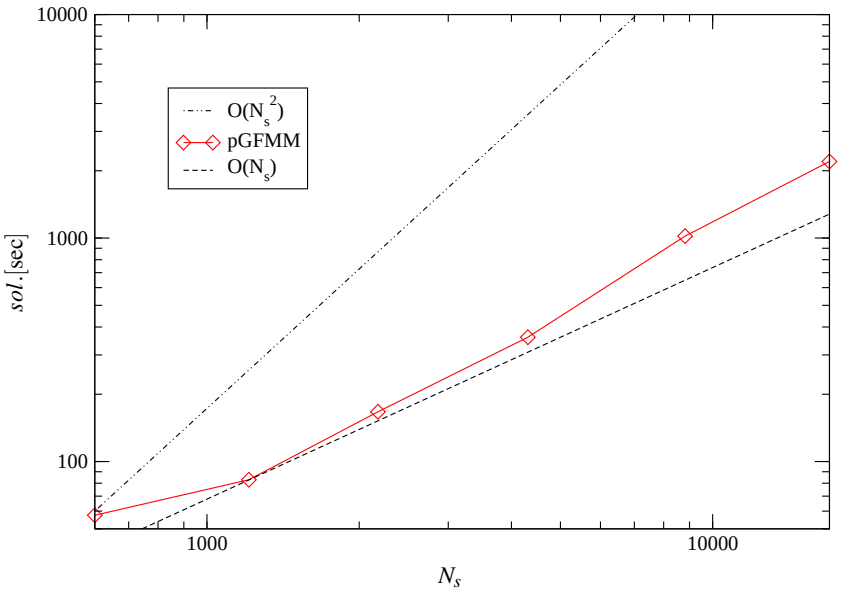


Figure 6: Solution time vs. number of unknowns N_s for an ellipsoid with increasing curvature ε.

6 Conclusion

We presented an efficient implementation of a Galerkin BEM for the transient heat equation using the parabolic FMM algorithm. Due to the improved evaluation of the temporal nearfield

by a combination of numerical quadrature and the FGT our method allows a fast application of the resulting layer potentials for any relation between temporal- and spatial discretization parameters.

For simplicity we only presented results of the homogeneous initial Dirichlet boundary value problem obtained in the simplest possible conforming test- and trial space, which is the space-time tensor product space of piecewise constant functions in space and time.

Of course our method can be extended to higher orders by using higher order test- and trial functions. However, note that when going to higher orders in time, one has to use piecewise discontinuous polynomial test- and trial functions in order to preserve a block Toeplitz structure.

References

[1] D. Arnold and P. Noon. Coercivity of the single layer heat potential. *J. Comput. Math.*, 7: 100–104, 1989.

[2] Douglas N. Arnold and Wolfgang L. Wendland. On the asymptotic convergence of collocation methods. *Mathematics of Computation*, 41(164):pp. 349–381, 1983.

[3] M. M. Chawla and M. K. Jain. Error Estimates for Gauss Quadrature Formulas for Analytic Functions. *Mathematics of Computation*, 11(101):82–90, 1968.

[4] M. Costabel. Boundary integral operators for the heat equation. *Integral Equations Operator Theory*, 13(4):498–552, 1990.

[5] M. Costabel. Time-dependent problems with the boundary integral equation method. In E. Stein, R. de Borst, and T. Hughes, editors, *Encyclopedia of Computational Mathematics*. Wiley, 2004.

[6] P. J. Davis. *Interpolation and Approximation*. Blaisdell, 1963.

[7] L. Greengard and P. Lin. Spectral approximation of the free-space heat kernel. *Applied and Computational Harmonic Analysis*, 9:83–97, 1999.

[8] L. Greengard and J. Strain. A fast algorithm for the evaluation of heat potentials. *Comm. Pure Appl. Math.*, XLIII:949–963, 1990.

[9] L. Greengard and J. Strain. The fast Gauss transform. *SIAM J. Sci. Comput.*, 12:79–94, 1991.

[10] L. Greengard and X. Sun. A new version of the fast Gauss transform. *Doc. Math. J. DMV*, Extra Volume ICM 1998, III:575–584, 1998.

[11] G. Hsiao and J. Saranen. Boundary integral solution of the two-dimensional heat equation. *Math. Methods Appl. Sci.*, 16(2):87–114, 1993.

[12] J.R. Li and L. Greengard. High order accurate methods for the evaluation of layer heat potentials. *SIAM J. Sci. Comput.*, 31(5):3847–3860, 2009.

[13] Ch. Lubich and R. Schneider. Time discretization of parabolic boundary integral equations. *Numer. Math.*, 63(1), 1992.

[14] N. Nishimura. Fast multipole accelerated boundary integral equation methods. *Applied Mechanics Reviews*, 55(4):299–323, 2002.

[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] M. Schanz and H. Antes. Application of operational quadrature methods in time domain boundary element methods. *Meccanica*, 32(3):179–186, 1997.

[17] J. Tausch. A fast method for solving the heat equation by layer potentials. *J. Comput. Phys.*, 224:956–969, 2007.

[18] J. Tausch. Nyström discretization of parabolic boundary integral equations. *Appl. Numer. Math.*, 59(11):2843–2856, 2009.

[19] J. Tausch. Fast Nyström methods for parabolic boundary integral equations. In U. Langer, M. Schanz, O. Steinbach, and W.L. Wendland, editors, *Fast Boundary Element Methods in Engineering and Industrial Applications.*, volume 63 of *Lecture Notes in Applied and Computational Mechanics*, pages 185–219. Springer, 2011.

[20] J. Tausch and A. Weckiewicz. Multidimensional fast Gauss transforms by Chebyshev expansions. *SIAM J. Sci. Comput.*, 31(5):3547–3565, 2009.