# Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig


## Hierarchical matrices and the High-Frequency Fast Multipole Method for the Helmholtz Equation with Decay

(revised version: March 2014)

by


*Maryna Kachanovska*



Preprint no.: 13 2014


**Abstract**

The solution of boundary-value problems for the Helmholtz equation with decay is required by many physical applications, in particular viscoelastodynamics and electromagnetics. The boundary integral equation method allows to reduce the dimensionality of the problem by expressing the unknown quantity with the help of a boundary integral operator of a density given on the boundary of the domain. However, the BEM discretization of boundary integral formulations typically leads to densely populated matrices.

In the last three decades a new generation of data-sparse methods for the approximation of BEM matrices was designed. Among those are panel-clustering, hierarchical matrices ($\mathcal{H}$-matrices), $\mathcal{H}^2$-matrices and fast multipole methods (FMM).

In this work we review main concepts of data-sparse techniques. We present a description of the high-frequency fast multipole method (HF FMM) with some technical details, both for a real and complex wavenumber. A significant part of the report is dedicated to the error analysis of the HF FMM applied to the Helmholtz equation with a complex wavenumber. We compare the performance of the multilevel high-frequency fast multipole method and $\mathcal{H}$-matrices for the approximation of the single layer boundary operator for the Helmholtz equation with decay. Based on these results, a simple strategy to choose between these techniques is suggested.

# Contents

# Chapter 1

# Introduction

## 1.1 Overview

The solution of boundary-value problems for the Helmholtz equation with decay is required by many applications, particularly, in viscoelastodynamics (materials with damping) and electromagnetics (propagation of waves in the lossy media). The boundary integral equation method allows to reduce the dimensionality of the problem by expressing the unknown quantity via a boundary integral operator of a density given on the surface of the domain. This allows to reduce the dimensionality of the problem. Typically, however, the Galerkin discretization of such formulations leads to densely populated matrices.

In the last three decades a new generation of data-sparse methods for the approximation of BEM matrices was designed. Among those are panel clustering, hierarchical matrices ($\mathcal{H}$-matrices), $\mathcal{H}^2$-matrices and fast multipole methods (FMM). The main ideas of these techniques can be traced back to the classical Barnes-Hut algorithm for the fast $O(n \log n)$ evaluation of gravitational (Coulomb) interactions of $n$ bodies [11].

Hierarchical matrix techniques [57, 59, 63, 49, 14, 12] rely on a blockwise low-rank approximation of a BEM matrix. In the case when these ranks can be bounded by $C \log^{\alpha} M$, where $\alpha \geq 0$, $C > 0$ and $M$ is the size of the matrix, the corresponding approximation is of an almost linear complexity. This is true for many important applications, e.g. the Laplace equation in two and three dimensions, the biharmonic equation, the inverses of FEM matrices and BEM preconditioners. Unlike most of fast multipole methods, $\mathcal{H}$-matrix techniques do not require certain analytic expansions of an integral kernel to be known, but need a procedure to evaluate the integral kernel [12, 19, 48]. The applicability of $\mathcal{H}$-matrix techniques to the two-dimensional Helmholtz equation has been analyzed in [8] and to the three-dimensional Helmholtz equation in [14]. The questions of the application of $\mathcal{H}$-matrices to the Helmholtz equation with decay were studied in [7, 14].

Panel clustering techniques [64, 58] are closely related to one-level fast multipole methods, as well as to $\mathcal{H}^2$-matrices. The latter [20, 62, 60, 61] can be viewed as a generalization of fast multipole methods. The history of fast multipole methods starts with the seminal works [74, 52], where an algorithm for the fast evaluation of the sums

$$f_j = \sum_{n=1}^{N} q_n \frac{1}{\|x_n - x_j\|}, \; x_j \in \mathbb{R}^3, \; q_j \in \mathbb{R}, \; j = 1, \ldots, N,$$

$$f_j = \sum_{m=1}^{N} q_m \log \|x_m - x_j\|, \; x_j \in \mathbb{R}^2, \; q_j \in \mathbb{R}, \; j = 1, \ldots, N,$$

was developed. Similarly to $\mathcal{H}$-matrices, the fast multipole methods rely on certain matrix decompositions of blocks of a BEM matrix. These decompositions are chosen so that respective matrix-vector products can be efficiently evaluated. The complexity of the FMM depends on the complexity of such

transforms. Though recently several black-box fast multipole methods were developed [92, 36, 37], many FMMs heavily rely on the analytic expansions of integral kernels.

The one-level fast multipole method for the Helmholtz potential was introduced in [75]. An excellent algorithmic description of this method can be found in [30]. A wide range of works is dedicated to various improvements and the efficient implementation of the high-frequency fast multipole algorithm: see [80, 33, 26, 84, 28, 87, 56, 85] and references therein. In [31] the author developed the fast multipole algorithm coupled with the microlocal discretization; this method is particularly efficient for high frequencies. A stable for all frequencies fast multipole method for Maxwell equations was introduced in [34]. In [93, 55, 56] an alternative FMM was developed, based on the expansions of [38]. When applied to the high-frequency Helmholtz equation, this method is not of linear complexity, however, in practice it may perform better than the HF FMM for small discretizations. The applicability of this method to the Helmholtz equation with decay was numerically analyzed in [43]. In this work the authors numerically examined the effect of decay on the length of multipole expansions and suggested empirical formulas well suited for the error control. In particular, it was shown that the presence of decay decreases the complexity of multipole transforms, thus improving the efficiency of the algorithm.

In [27] the authors mentioned that the choice of the ranks of the multipole transforms can be performed ignoring the complex part of the wavenumber (decay), though for large decays more savings are possible. The fast multipole method for the Yukawa potential $\frac{\mathrm{e}^{-\lambda r}}{4\pi r}$, $\lambda > 0$, was developed in [53].

The applicability of the HF FMM of [27] to the Helmholtz equation with decay was analyzed in [51, 44]. There the authors suggested that if the relative error (instead of the scaled error defined in [43]) is of interest, the complexity of FMM transforms slightly increases with the increase of decay. In [51] the authors proposed a close to optimal empirical formula to determine the length of the multipole expansion.

The work [91] is dedicated to the numerical studies of the applicability of the high-frequency fast multipole method to the Helmholtz equation with decay; the authors demonstrated that if decay is sufficiently large, cancellation errors can occur, and proposed a strategy to avoid these errors (which resembles the use of the definition of the scaled error as in [43]). The cancellation and roundoff errors were studied as well in [6].

We refer to [43] for the review of other works on the fast multipole method for the Helmholtz equation with a complex wavenumber, as well as the list of possible applications.

The comparison of the $\mathcal{H}$-matrices and the FMM for the Helmholtz equation (without decay) is presented in [22], where both techniques were applied to solve the Burton-Miller integral formulation. The HF FMM used in this work was slightly modified for the efficient application with the half-space mirror techniques, see [23].

The primary goal of this technical report is two-fold. First, we present a description of the high-frequency fast multipole method with some important technical details. Second, we compare the performance of the HF FMM and $\mathcal{H}$-matrices for the approximation of boundary integral operators of the Helmholtz equation with decay. A secondary goal is to present an error analysis of the multilevel HF FMM that would incorporate the case of the complex wavenumber. This work extends and refines some of the results of [68].

This report is organized as follows. In the first section we present a sample problem that requires application of data-sparse techniques. Next, we review basic definitions from the theory of data-sparse techniques, as well as recall the notions of $\mathcal{H}$- and $\mathcal{H}^2$-matrices. We summarize the main results of the existing studies on the applicability of $\mathcal{H}$-matrices to the Helmholtz equation with decay. Chapter 3 is dedicated to the algorithmic description of the HF FMM. In Chapter 4 we provide error analysis and suggest a strategy to choose the parameters of the HF FMM based on theoretical and numerical considerations. For a reader interested solely in the implementation of the HF FMM, it is sufficient to read Sections 2.2, 2.4.1, 2.5, Chapter 3 and Section 4.5. Chapter 5 contains numerical experiments on the $\mathcal{H}$-matrix approximations and HF FMM. We discuss the efficiency of both strategies and suggest a heuristic that would allow to choose between the HF FMM and $\mathcal{H}$-matrices in a concrete case.

## 1.2 Motivation

Let $\Omega \subset \mathbb{R}^3$ be a bounded Lipschitz domain, $\Omega^c$ be its complement and $\Gamma$ be its boundary. The function $u$ solves the following Dirichlet boundary-value problem:

$$\begin{aligned} -\Delta u + s^2 u = 0, \qquad & \text{in } \Omega^c,\ s \in \mathbb{C} :\ \operatorname{Re} s > 0, \\ u(x) = g(x), \qquad & x \in \Gamma. \end{aligned} \tag{1.1}$$

If $\operatorname{Re} s = 0$, the unique solvability of this problem can be guaranteed only if the above problem is equipped with a radiation condition at infinity. For $\operatorname{Re} s > 0$ it is sufficient to require that $u(x) \in H^1(\Omega^c)$ [5]. The existence and uniqueness of solutions to this problem was proved in [5].

The unknown $u(x)$ can be written as the single-layer potential of an unknown density $\lambda(x),\ x \in \Gamma$, namely

$$u(x) = \int_\Gamma \frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}\lambda(y)d\Gamma_y, \qquad x \in \Omega^c.$$

The continuity of the single layer potential through $\Gamma$ allows to write the following boundary integral equation for $\lambda$:

$$g(x) = (V(s)\lambda)(x) = \int_\Gamma \frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}\lambda(y)d\Gamma_y, \qquad x \in \Gamma. \tag{1.2}$$

The well-posedness of this problem is due to [5], where the following proposition was deduced.

**Proposition 1.2.1.** *For $\operatorname{Re} s > 0$, the single layer boundary operator $V(s)$ is an isomorphism*

$$V(s) : H^{-\frac{1}{2}}(\Gamma) \to H^{\frac{1}{2}}(\Gamma).$$

*If $\operatorname{Re} s > \sigma_0$, for some $\sigma_0 > 0$, then*

$$\begin{aligned} \|V(s)\|_{H^{-\frac{1}{2}}(\Gamma)\to H^{\frac{1}{2}}(\Gamma)} &\le \frac{C_1}{\sigma_0}\max\left(\frac{1}{\sigma_0^2}, 1\right)|s|, \\ \|V^{-1}(s)\|_{H^{\frac{1}{2}}(\Gamma)\to H^{-\frac{1}{2}}(\Gamma)} &\le \frac{C_2}{\sigma_0}\max\left(\frac{1}{\sigma_0}, 1\right)|s|^2, \end{aligned}$$

*for some $C_1, C_2 > 0$ that depend on $\Gamma$ only. For all $\phi \in H^{-\frac{1}{2}}(\Gamma)$, the following coercivity estimate holds:*

$$\operatorname{Re}\langle \phi, sV(s)\phi\rangle \ge C_3 \min(\sigma_0, 1)|s|^{-1}\|\phi\|^2_{H^{-\frac{1}{2}}(\Gamma)},$$

*where $C_3 > 0$ does not depend on $s, \phi$.*

Here we concentrate on the efficient approximation of the Galerkin discretization of (1.2). In particular, we analyze $\mathcal{H}$- and $\mathcal{H}^2$-matrix techniques applied to the Galerkin discretization of the single-layer boundary integral operator for $-\Delta + s^2$, $s \in \mathbb{C}$, $\operatorname{Re} s > 0$. Let the boundary $\Gamma$ be subdivided into $N$ panels $(\pi_i)_{i=1}^N$. Let $(\phi_i)_{i=1}^M$ be Galerkin test and trial basis functions. Then the Galerkin discretization of $V(s)$ is defined as

$$\boldsymbol{V}_{ij} = \iint_{\Gamma\times\Gamma} \frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}\phi_i(x)\phi_j(y)d\Gamma_x d\Gamma_y, \qquad i,j = 1,\dots,M. \tag{1.3}$$

For simplicity, throughout this work we use piecewise constant basis functions, i.e.

$$\phi_i(x) = \begin{cases} 1, & x \in \pi_i, \\ 0, & \text{else}, \end{cases} \qquad i = 1,\dots,N = M.$$

All the arguments can be extended to a more general case under the assumption that the test and trial basis functions are compactly supported (more precisely, supported on a constant number of mesh elements). Additionally, we assume that given the meshwidth $\Delta x$, the number of boundary elements scales as $M = O\left(\frac{1}{(\Delta x)^2}\right)$.

The case $\operatorname{Re} s = 0$ is considered as well. However, the well-posedness of the problem (1.1) can be guaranteed only if it is equipped with a proper radiation condition. The equation (1.2) is no longer uniquely solvable and should be changed to another well-posed integral formulation (e.g. the combined field integral formulation)[1].

[1]For more information on the theory of boundary integral equations (BIE) see [71], [86] and [83]. Monographs [86] and [83] are dedicated to the numerical treatment of BIE.

# Chapter 2

# $\mathcal{H}$- and $\mathcal{H}^2$-matrices

The notion of $\mathcal{H}$-matrices was introduced in [57]. The questions of the efficient construction of $\mathcal{H}$-matrices and the complexity of algebraic operations are addressed in [49]. Recent monographs [59, 14] are dedicated to $\mathcal{H}$-matrix theory and provide both theoretical and numerical evidence of the efficiency of $\mathcal{H}$-matrix techniques. The non-exhaustive list of applications include the approximation of boundary integral operators [63], of the inverse of FEM matrices [16], and efficient $\mathcal{LU}$-preconditioners in BEM and FEM [13, 50].

In this section we review the main notions of the theory of $\mathcal{H}$-matrices, following [59].

## 2.1 Asymptotically Smooth Functions

**Definition 2.1.1.** *Given $X, Y \subset \mathbb{R}^d$, a function $k : X \times Y \to \mathbb{C}$ is called separable if it can be written in the following form:*

$$k(x,y) = \sum_{\nu=1}^{r} a_\nu(x) b_\nu(y),\ (x,y) \in X \times Y.$$

*The right hand side of*

$$k(x,y) = \sum_{\nu=1}^{r} a_\nu^{(r)}(x) b_\nu^{(r)}(y) + R_r(x,y),\ (x,y) \in X \times Y.$$

*is called an $r$-term separable expansion of $k(x,y)$ with the remainder $R_r$.*

One of the methods to obtain a separable expansion is polynomial interpolation. Namely, let $X$ be a box

$$X = [a_1,\ b_1] \times \dots [a_d,\ b_d]. \tag{2.1}$$

Let a set of interpolation points $(x_k^{(j)})_{j=1}^{N_k}$ be defined on intervals $[a_k,\ b_k]$, $k = 1, \dots, d$. Let

$$L_{k,\nu}(x) = \prod_{\mu=1,\dots,N_k,\ \mu \neq \nu} \frac{x - x_k^{(\mu)}}{x_k^{(\nu)} - x_k^{(\mu)}}$$

denote the $\nu^{th}$ Lagrange polynomial. Given $x = (x_1, \dots, x_d)$,

$$k^{(r)}(x,y) = \sum_{\nu_1=1}^{N_1} L_{1,\nu_1}(x_1) \sum_{\nu_2=1}^{N_2} L_{2,\nu_2}(x_2) \dots \sum_{\nu_d=1}^{N_d} L_{1,\nu_d}(x_d) k\left(\left(x_1^{(\nu_1)}, \dots x_d^{(\nu_d)}\right), y\right) \tag{2.2}$$

constitutes an $r = N_1 N_2 \dots N_d$-term separable expansion of $k(x, y)$. Let us for simplicity assume that $N_1 = N_2 = \dots N_d = m$. By $\Lambda_i$ we denote the Lebesgue constant for the set of interpolation points on $[a_i, b_i]$; if the interpolation points are chosen as Chebyshev nodes

$$\Lambda_i = \Lambda < 1 + \frac{2}{\pi} \log(m + 1).$$

The interpolation error, i.e. the remainder $R_r$, can be bounded by, see [59, Lemma B.3.4],

$$|k(x, y) - k^{(r)}(x, y)| \leq \frac{1}{m!} \Lambda^{d-1} \sum_{i=1}^{d} \|\omega_i\|_\infty \|\partial_{x_i}^m k(\cdot, y)\|_{\infty, X},$$

where $\omega_i = \omega_i(x_i) = \prod_{j=1}^{m} (x_i - x_j^{(i)})$. If all the partial derivatives of $k(x, y)$ do not grow too fast, his error can be easily controlled. This motivates the introduction of another important concept, namely asymptotic smoothness.

**Definition 2.1.2.** *Let $X, Y \subseteq \mathbb{R}^d$ and let $k : \{(x, y) \in X \times Y, x \neq y\} \to \mathbb{C}$ be smooth. Then $k$ is called asymptotically smooth if there exist $C, \kappa, \gamma, p$, s.t.*

$$|\partial_x^\alpha \partial_y^\beta k(x, y)| \leq C(\alpha + \beta)! |\alpha + \beta|^\kappa \gamma^{|\alpha+\beta|} \|x - y\|^{-|\alpha|-|\beta|-p}, \tag{2.3}$$

*for all $x \in X$, $y \in Y$, $x \neq y$, $\alpha, \beta \in \mathbb{N}_0^d$, $\alpha + \beta \neq 0$.*

The following lemma shows that under additional geometrical assumptions on domains $X$, $Y$ tensor-product interpolation (2.2) of the asymptotically smooth $k(x, y)$ converges to $k(x, y)$ exponentially.

**Lemma 2.1.3.** *[59, Proposition 4.2.13, p.69] Let $k(x, y)$ be asymptotically smooth in $X \times Y \subset \mathbb{R}^d \times \mathbb{R}^d$. Let additionally (2.1) hold. Let the number of interpolation points in each of the directions equal $m - 1$ and the Lebesgue constant $\Lambda_j = O(c^m)$, for some $c > 1$ and for all $j = 1, \dots, d$. Given $m + s \geq 0$ (where $s$ is defined by (2.3)), tensor-product interpolation (2.2) approximates $k(x, y)$ with the error*

$$\|k(., y) - k^{(r)}(., y)\|_{\infty, X} \leq c_1 \left( \frac{c_2 \operatorname{diam}_\infty(X)}{\operatorname{dist}(y, X)} \right)^m, \; y \in Y \setminus X, \tag{2.4}$$

*where $c_1, c_2$ do not depend on $m$ and*

$$\operatorname{diam}_\infty(X) = \max\{b_i - a_i : \; 1 \leq i \leq d\}.$$

Hence there exists a separable expansion for an asymptotically smooth $k(x, y)$ in $X, Y$ if

$$\eta \operatorname{diam}_\infty(X) < \operatorname{dist}(Y, X),$$

for some $\eta > 1$.

## 2.2 Cluster Trees and Block Cluster Trees

Let the boundary $\Gamma$ be subdivided into $M$ panels $\pi_i$, and let the corresponding index set be defined as $\mathcal{I} = \{1, \dots, M\}$. Note that when piecewise-constant basis functions are employed, $\operatorname{supp} \phi_i = \pi_i$, $i = 1, \dots, M$.

**Definition 2.2.1.** *Given a constant $C$, a tree $\mathcal{T}_\mathcal{I}$ is called a cluster tree corresponding to an index set $\mathcal{I}$ if $\mathcal{T}_\mathcal{I}$ is a binary labeled tree with the following properties:*

- *the label $\hat{\tau}$ of a vertex $\tau$ of $\mathcal{T}_\mathcal{I}$ is a subset of $\mathcal{I}$;*

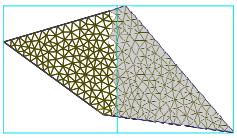

*Figure 2.1: Subdivision of a bounding box into two boxes of equal size, with two associated clusters filled with different colors (2D schematic representation).*

- *the label of the root of the tree is $\mathcal{I}$;*
- *the label of a vertex $\tau$ is a disjoint union of labels of its sons;*
- *for every leaf $\tau$, $\#\hat{\tau} \leq C$.*

The leaves of the cluster tree $\mathcal{T}_\mathcal{I}$ are denoted by $\mathcal{L}(\mathcal{T}_\mathcal{I})$. All the vertices located at the level $\ell$ of the cluster tree $\mathcal{T}_\mathcal{I}$ are denoted by $\mathcal{T}_\mathcal{I}^\ell$; the root is located at the level $\ell = 0$.

The structure of the cluster tree introduces a hierarchical subdivision of $\Gamma$ into sets of panels. A set of panels corresponding to a cluster $\tau$ is denoted by $\Omega_\tau$:

$$\Omega_\tau = \bigcup_{i \in \hat{\tau}} \pi_i.$$

In the $\mathcal{H}$-matrix theory, the bounding box of a cluster $\tau$ is the (axis-parallel) box containing the set $\Omega_\tau$; the center of the box we denote by $c_\tau$ and its diameter by $d_\tau$. The next definition can be found in [18, Def. 3.16].

**Definition 2.2.2.** *A predicate $\mathcal{A}$: $\mathcal{T}_\mathcal{I} \times \mathcal{T}_\mathcal{I} \to \{\text{true}, \text{false}\}$ is an admissibility condition for $\mathcal{T}_\mathcal{I} \times \mathcal{T}_{\mathcal{T}_I}$, if $\mathcal{A}(\tau, \upsilon) = \text{true}$ implies that for all $\tau' \in \text{sons}(\tau)$, $\mathcal{A}(\tau', \upsilon) = \text{true}$ and for all $\upsilon' \in \text{sons}(\upsilon)$, $\mathcal{A}(\tau, \upsilon') = \text{true}$.*

**Remark 2.2.3.** *When the high-frequency fast multipole method is employed, the actual definition of a bounding box is slightly different from the one used in the $\mathcal{H}$-matrix theory. Namely, as we show further, for the efficient implementation of the HF FMM it is crucial that the boxes on the same level of the cluster tree are of the same size. This can be achieved by a subsequent bisection of the bounding box of the domain. In case when some of the boundary elements do not belong to any of the boxes (they cross the border of two boxes), see Figure 2.1, they are assigned to either of the clusters.*

Now we have all the ingredients to introduce the concept of the admissible block-cluster tree. We adopt here a slightly modified definition, similar to the one used in the high-frequency fast multipole method [27]. In the $\mathcal{H}$-matrix theory it corresponds to the level-consistent admissible block-cluster tree.

**Definition 2.2.4.** *Let $\mathcal{T}_\mathcal{I}$ be a cluster tree. We will call an admissible block-cluster tree $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$ a subtree of a labeled tree $\mathcal{T}_\mathcal{I} \times \mathcal{T}_\mathcal{I}$ that satisfies the following conditions:*

1. *The root of the tree is $(\text{root}(\mathcal{T}_\mathcal{I}), \text{root}(\mathcal{T}_\mathcal{I}))$.*
2. *The son clusters of each block-cluster $b = (\tau, \sigma)$ are defined by*

$$\text{sons}(b) = \begin{cases} \{(\tau', \sigma'),\ \tau' \in \text{sons}(\tau),\ \sigma' \in \text{sons}(\sigma)\}, & \text{sons}(\tau) \neq \emptyset,\ \text{sons}(\sigma) \neq \emptyset, \\ \emptyset, & \text{sons}(\tau) = \emptyset \text{ or } \text{sons}(\sigma) = \emptyset; \end{cases}$$

3. *A block-cluster $(\tau, \sigma)$ is a leaf if and only if one of the following holds true:*
   (a) *$(\tau, \sigma)$ is admissible;*
   (b) *$(\tau, \sigma)$ is not admissible, and $\tau \in \mathcal{L}(\mathcal{T}_\mathcal{I})$ or $\sigma \in \mathcal{L}(\mathcal{T}_\mathcal{I})$;*

Let us note that the actual choice of the admissibility condition depends on the integration kernel. For asymptotically smooth kernels the natural choice is, see (2.4),

$$\eta \operatorname{dist}(\tau, \sigma) \geq \max\{d_\tau, d_\sigma\}, \tag{2.5}$$

for some $\eta > 0$.

In the literature on the fast multipole methods it is quite common to use an admissibility condition of the form (2.5), or a similar one: only the neighboring clusters are not admissible. We use a slightly different admissibility condition, see also [78].

**Definition 2.2.5.** *We will call a pair of clusters $(\tau, \sigma)$ admissible if for some fixed $\eta > 1$ the following holds true:*

$$\|c_\tau - c_\sigma\| \geq \frac{\eta}{2}(d_\tau + d_\sigma).$$

Thus, all the leaves of the admissible block-cluster tree can be split into two sets, namely $\mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ of admissible block-clusters and $\mathcal{L}_-(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ of non-admissible block-clusters. The first set is called the **far-field**, while the second one is referred to as the **near-field**.

**Remark 2.2.6.** *The definition of a cluster tree can be modified to include the case of an octree (instead of a binary tree), which is often used in the HF FMM theory. All the definitions of this section can be generalized to cover this case.*

*In the HF FMM a cluster octree is constructed using the following procedure. For each non-leaf cluster $\tau$, its sons are formed by splitting its bounding box into 8 identical boxes and associating each of the resulting bounding boxes with a set of indices. Since $\Gamma$ is a 2-dimensional surface, in average every non-leaf cluster will have only 4 non-empty sons.*

## 2.3 $\mathcal{H}$-matrices

We wish to approximate a Galerkin matrix of a (boundary) integral operator

$$(\mathcal{M})_{ij} = \iint\limits_{\Gamma\times\Gamma} k(x,y)\phi_i(x)\phi_j(y)d\Gamma_x d\Gamma_y,\ i,j = 1,\dots,M, \tag{2.6}$$

in the $\mathcal{H}$-matrix format. The main idea that lies behind $\mathcal{H}$-matrix techniques is the following.

Let us assume that the kernel of the integral operator $k(x,y)$ is an asymptotically smooth function. The admissibility condition has to be chosen so that $k(x,y)$ has a separable expansion inside all the admissible clusters, see Section 2.1. Given $(\tau,\upsilon) \in \mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$, for all $i \in \hat{\tau}$, $j \in \hat{\upsilon}$, it should hold that

$$\iint\limits_{\Omega_\tau\times\Omega_\upsilon} k(x,y)\phi_i(x)\phi_j(y)d\Gamma_x d\Gamma_y \approx \sum_{k=1}^{r}\left(\int\limits_{\Omega_\tau} a_k^{(r)}(x)\phi_i(x)d\Gamma_x\right)\left(\int\limits_{\Omega_\upsilon} b_k^{(r)}(y)\phi_j(y)d\Gamma_y\right). \tag{2.7}$$

Let $\#\hat{\tau} = n$, $\#\hat{\upsilon} = m$. We denote by $\mathcal{M}|_{\hat{\tau}\times\hat{\upsilon}}$ the following matrix block:

$$(\mathcal{M}|_{\hat{\tau}\times\hat{\upsilon}})_{k_i\ell_j} = \mathcal{M}_{ij}, \qquad k_i \in \{1,\dots,n\},\ \ell_j \in \{1,\dots,m\},\ i \in \hat{\tau},\ j \in \hat{\upsilon}.$$

The expansion (2.7) shows that the matrix block $\mathcal{M}|_{\hat{\tau}\times\hat{\upsilon}}$ can be approximated by a rank $r$-matrix. Hence instead of storing all matrix entries it is possible to keep in the memory only $r$ $n$-dimensional vectors $A^{(\alpha)}$, $\alpha = 1,\dots r$,

$$A_{k_i}^{(\alpha)} = \int\limits_{\Omega_\tau} a_\alpha^{(r)}(x)\phi_i(x)d\Gamma_x, \qquad i \in \hat{\tau},$$

and $r$ $m$-dimensional vectors $B^{(\alpha)}$, $\alpha = 1, \dots r$,

$$B^{(\alpha)}_{\ell_j} = \int_{\Omega_v} b^{(r)}_\alpha(x)\phi_j(x)d\Gamma_x, \qquad j \in \hat{v}.$$

This allows to reduce storage costs and the time of the matrix-vector multiplication from $O(nm)$ to $O(r(n+m))$.

**Definition 2.3.1.** *Let $\mathcal{I}$ be an index set and $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$ be an admissible block-cluster tree. Let also $k : \mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}}) \to \mathbb{N}_+$. A matrix $M \in \mathbb{C}^{\mathcal{I}\times\mathcal{I}}$ is called an $\mathcal{H}$-matrix (or hierarchical matrix) if for each $b = (\tau, \sigma) \in \mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ the matrix $M|_b$ is a $k(b)$-rank matrix, i.e.*

$$\operatorname{rank} M|_b \le k(b)$$

*and is represented in the form*

$$M|_b = A_b B_b^T,$$

*where $A_b \in \mathbb{R}^{\tau\times\{1,\dots,k(b)\}}$, $B_b \in \mathbb{R}^{\sigma\times\{1,\dots,k(b)\}}$.*

An important notion for analyzing the complexity of $\mathcal{H}$-matrix arithmetic is the sparsity constant. We provide here a definition adapted to our needs, similar to [59] and [77]; for alternative definitions see the same reference, as well as [49].

**Definition 2.3.2.** *The sparsity constant for $\mathcal{L}_0 \subset \mathcal{L}(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ is defined as*

$$C_{sp}(\mathcal{L}_0) = \max\left\{\max_{\tau\in\mathcal{T}_\mathcal{I}}\{\#\sigma \in \mathcal{T}_\mathcal{I} : (\tau,\sigma) \in \mathcal{L}_0\},\ \max_{\sigma\in\mathcal{T}_\mathcal{I}}\{\#\tau \in \mathcal{T}_\mathcal{I} : (\tau,\sigma) \in \mathcal{L}_0\}\right\}.$$

In [64] it was demonstrated that under some mild assumptions on $\Gamma$, the sparsity constant can be bounded by a constant that depends on the admissibility condition and the space dimension.

The following lemma can be found in [49, Lemma 2.5] and [59, Lemma 6.3.6].

**Lemma 2.3.3.** *Let $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$ be a given admissible block-cluster tree with the sparsity constant $C_{sp}$ for $\mathcal{L}(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$. Let $M$ be an $\mathcal{H}$-matrix, and $k > 0$ be s.t. for all $b \in \mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$*

$$\operatorname{rank}(M|_b) \le k.$$

*Additionally, let $n_{\min} \in \mathbb{N}_+$ be s.t. for all $(\tau,\sigma) \in \mathcal{L}_-(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$:*

$$\#\tau \le n_{\min} \ \text{ and } \ \#\sigma \le n_{\min}.$$

*Then the following complexity estimates hold:*

1. *The storage costs scale as*

$$S_\mathcal{H} \le 2C_{sp}\max(n_{\min}, k)(\operatorname{depth}(\mathcal{T}_\mathcal{I}) + 1)\#\mathcal{I}.$$

2. *The complexity of the matrix vector product can be bounded by*

$$\mathcal{M} \le 2S_\mathcal{H}.$$

From now we assume that $C_{sp}$ does not depend on size of the discretization. The cluster tree is constructed so that $\operatorname{depth}(\mathcal{T}_\mathcal{I}) = O(\log \#\mathcal{I})$. Under these suppositions the storage costs, as well as the complexity of the matrix-vector product scale almost linearly, i.e. $O(M \log M)$, with respect to the size of the index set $\mathcal{I}$, and hence the number of Galerkin basis functions.

In practice the construction of $\mathcal{H}$-matrices is usually done using techniques based on the ideas from [47], e.g. ACA [12, 17], ACA+[48] or HCA [19] (although formula (2.4) is available). Such methods, besides being computationally efficient, possess major advantages over the polynomial expansion (2.4):

- no a priori information on ranks is needed, only evaluations of the integral kernel are used;
- low-rank approximations constructed with the help of such techniques can be close (and in practice are close) to optimal [47, 15].

The optimal low-rank approximation can be constructed with the help of the singular value decomposition. Due to high computation costs, the SVD is employed in the $\mathcal{H}$-matrix theory only rarely (e.g. for coarsening, see [48]).

## 2.4 $\mathcal{H}$-matrices for Helmholtz Boundary Integral Operators

Questions of the applicability of $\mathcal{H}$-matrices to the Helmholtz equation have been studied in various works [8, 7, 46], see also [14] and references therein. Let us address the simplest case, namely, the use of $\mathcal{H}$-matrices for the approximation of the Galerkin discretization of the boundary integral operator $V(s)$, namely

$$(\boldsymbol{V}(s))_{ij} = \iint\limits_{\Gamma\times\Gamma} \frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}\phi_i(x)\phi_j(y)d\Gamma_x d\Gamma_y,\ i,j=1,\ldots,M,$$

where $(\phi_i(x))_{i=1}^M$ are piecewise-constant test (and trial) basis functions.

In a nutshell, the results of these studies are the following.

1. If $s = i\kappa$, $\kappa \in \mathbb{R}$, the complexity (storage and matrix-vector multiplication) is bounded by $O(M|\kappa|\log M)$; the hidden constant depends on the accuracy. A typical assumption on the number of Galerkin basis functions is

   $$M = O(|\kappa|^2), \qquad \text{as } |\kappa| \to +\infty.$$

   Consequently, the complexity scales as

   $$O(M^{\frac{3}{2}}\log M).$$

   This result can be found in Section 3.4.5.2 of [14].

2. For complex $s : \operatorname{Re} s > 0$, in [7] it was shown that if $\frac{|\operatorname{Im} s|}{|\operatorname{Re} s|} < c$, for some $c > 0$, the Helmholtz kernel is asymptotically smooth and the complexity of the $\mathcal{H}$-matrix approximation is almost linear. Based on the results of [21], in [14, p.114, Theorem 3.18, p.157] it was demonstrated that the matrix-vector multiplication and storage costs depend on $M$ as

   $$O\left(M\left(C + \left|\frac{\operatorname{Im} s}{\operatorname{Re} s}\right|\right)\log M\right),$$

   for some $C > 0$. Under the assumption $M = O(|s|^2)$, the $\mathcal{H}$-matrix approximation is of almost linear complexity, namely $O(M\log M)$.

The complexity of $\mathcal{H}$-matrix assembly depends on the complexity of the procedure used for computing the low-rank approximations of matrix blocks. In particular, the complexity of the adaptive cross-approximation (ACA) [12, 17], see also [48, 19], applied to a matrix block of size $m\times n$, can be estimated by $O(r^2(m+n))$, where $r$ is the rank of the low-rank matrix approximation of this block computed in the course of the ACA [14]. When the ACA is applied to the Galerkin discretization of the Helmholtz single layer boundary operator, its complexity scales as

$$O(C_1 r(m+n) + C_2 r^2(m+n)), \tag{2.8}$$

where $C_1$ depends on the complexity of the evaluation of 4-dimensional BEM integrals, and $C_2$ is a (small) constant. The value $C_1 = O\left(\log^a M\right)$, where $a \le 4$, due to the use of tensor Gauss

quadratures with coordinate transformations, applied with the accuracy sufficient to preserve the stability of the Galerkin method [83, 39, 65, 81, 82].

It can be shown that the ranks $r$ can be estimated by $O(|\kappa|)$[14, p.157]), and thus the complexity of the ACA based assembly of an $\mathcal{H}$-matrix approximation of the Helmholtz single layer boundary operator in the case $s = i\kappa,\ \kappa \in \mathbb{R}$, is $O(|\kappa|^2 M) = O(M^2)$, i.e. the same as for the dense matrix. In practice, however, even for large problems the complexity of the matrix assembly scales close to $O(M^{\frac{3}{2}} \log^a M)$ (c.f. the numerical experiments in Section 5). There are two main reasons for this. First, in the complexity estimate of the ACA (2.8) usually $C_2 \ll C_1$, and the first term prevails even for large values of $r$ [14, Remark 3.31]. Second, the ranks produced by ACA-like procedures are typically close to optimal, and the estimate $r = O(|\kappa|)$ (which leads to the quadratic complexity) may appear too pessimistic in many practical cases.

In the case $s = |s|\mathrm{e}^{i\alpha},\ \alpha = \mathrm{const},\ \mathrm{Re}\, s > 0$, the asymptotic complexity of the matrix construction is larger than the storage and matrix-vector multiplication complexity by a factor up to $O\left(\log^a M\right)$, where $a \leq 4$ (this is due to the complexity of the evaluation of 4-dimensional boundary integrals), and hence scales not worse than $O\left(M \log^{a+1} M\right)$.

### 2.4.1 Efficient Construction of $\mathcal{H}$-Matrices for the Boundary Single Layer Operator of the Helmholtz Equation with Decay

In [43] the error of the fast multipole method for the Helmholtz equation with decay has been studied. It was suggested that the relative error in the case $\mathrm{Re}\, s > 0$ does not serve any more as a good error estimator. Clearly, the elements of the Galerkin matrix of the Helmholtz single layer boundary operator satisfy the inequality

$$\left| \int\limits_{\pi_i} \int\limits_{\pi_j} \frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|} \phi_i(x)\phi_j(y) d\Gamma_x d\Gamma_y \right| \leq \mathrm{e}^{-\mathrm{Re}\, s\, \mathrm{dist}(\pi_i,\pi_j)} \int\limits_{\pi_i} \int\limits_{\pi_j} \frac{1}{4\pi\|x-y\|} \phi_i(x)\phi_j(y) d\Gamma_x d\Gamma_y,$$

and hence become exponentially small when $\mathrm{dist}(\pi_i, \pi_j)$ gets larger. Therefore, the contribution of the blocks corresponding to the parts of the boundary $\Gamma$ that are distant from each other, can be neglected up to a certain tolerance when computing the matrix-vector product.

Accordingly, it is possible to skip constructing some blocks in the $\mathcal{H}$-matrix approximation. Let $(\tau, \sigma)$ denote an admissible block-cluster, and let the distance between the bounding boxes of the clusters $\tau$ and $\sigma$ equal $d > 0$. If

$$\frac{\exp(-d\,\mathrm{Re}\, s)}{4\pi d} < \epsilon, \tag{2.9}$$

for a fixed accuracy $\epsilon > 0$, the corresponding block can be approximated by a zero matrix. The actual choice of $\epsilon$ has to be made based on extensive numerical experiments.

The accuracy of the approximation of other blocks within the ACA+ algorithm may be reduced as well (see [76, pp.69-73] for the detailed description of the algorithm). When using the ACA/ACA+ algorithm, it is easier to control the relative accuracy of the approximation of matrix blocks, and hence we proceed as follows. If the distance between two admissible clusters is $d$, the relative accuracy of the approximation may be scaled by $\mathrm{e}^{\mathrm{Re}\, sd}$, what we also do. This is in correspondence with the definition of the scaled error in [43]. Alternatively, one could control the absolute accuracy of the approximation of each matrix block within the ACA/ACA+ algorithm.

## 2.5 $\mathcal{H}^2$-Matrices

The notion of $\mathcal{H}^2$-matrices was introduced in [62]. In [20] the authors developed a black-box algorithm that compresses a given matrix in the $\mathcal{H}^2$-matrix format. For the cases when the construction of a dense matrix is too expensive (e.g. discretizations of integral operators), an efficient method of the construction of $\mathcal{H}^2$-matrix based approximations was suggested in [61].

Another way to assemble an $\mathcal{H}^2$-matrix is based on the use of known explicitly separable expansions of an integral kernel, e.g. those coming from fast multipole methods. This was done for the discretization of the boundary single-layer operator for the Helmholtz equation in two dimensions in [8], as well as implicitly in [3].

In this section we review the main definitions of the $\mathcal{H}^2$-matrix theory based on recent monographs [18, 59] and lecture notes [76].

Let us fix a cluster tree $\mathcal{T}_{\mathcal{I}}$ and an admissible block-cluster tree $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$.

**Definition 2.5.1.** *A family of matrices $(V^t)_{t\in\mathcal{T}_{\mathcal{I}}}$, s.t. for all $t \in \mathcal{T}_{\mathcal{I}}$ the matrix $V^t \in \mathbb{C}^{\hat{t}\times K_t}$ for some finite index set $K_t$, is called a cluster basis.*

**Definition 2.5.2.** *(Uniform $\mathcal{H}$-matrix) Let $(V^t)_{t\in\mathcal{T}_{\mathcal{I}}}$, $(W^t)_{t\in\mathcal{T}_{\mathcal{I}}}$. A matrix $M \in \mathbb{C}^{\mathcal{I}\times\mathcal{I}}$ is called a uniform $\mathcal{H}$-matrix if for all admissible $(t,s) \in \mathcal{L}(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})$ there exists $S^{t,s} \in \mathbb{C}^{K^t\times K^s}$ s.t.*

$$M|_{\hat{t}\times\hat{s}} = V^t S^{t,s} \left(W^s\right)^T.$$

*Matrices $S^{t,s}$ are called coupling matrices.*

A uniform $\mathcal{H}$-matrix is an $\mathcal{H}$-matrix, since the ranks of all its subblocks corresponding to admissible clusters are bounded:

$$\operatorname{rank}\left(V^t S^{t,s}\left(W^s\right)^T\right) \leq \operatorname{rank} S^{t,s} \leq \min\left(\#K^t, \#K^s\right).$$

**Definition 2.5.3.** *A cluster basis $(V^t)_{t\in\mathcal{T}_{\mathcal{I}}}$ is called nested if for every non-leaf cluster $t$ and for all $t' \in \operatorname{sons}(t)$ there exists a matrix $T^{t'} \in \mathbb{C}^{K^{t'}\times K^t}$ ('transfer matrix'), such that*

$$V^t = V^{t'} T^{t'}.$$

**Definition 2.5.4.** *A uniform $\mathcal{H}$-matrix whose column and row cluster bases are nested is called an $\mathcal{H}^2$-matrix.*

The use of $\mathcal{H}^2$-matrices is motivated by a possible reduction of storage and computation costs when dealing with the nested cluster basis compared to the cluster basis. Namely, if $V^t$ are dense matrices, storing them for all $t \in \mathcal{T}_{\mathcal{I}}$ may be costly. In the case when the nested cluster basis is used, one only needs to store the cluster basis for leaves and (possibly) transfer matrices. If these are of a special structure (e.g. are sparse), storage costs and time for the computation of the matrix-vector product may be reduced significantly.

The algorithm for the efficient matrix-vector multiplication

$$y = Mx,$$

with $M$ being an $\mathcal{H}^2$-matrix, is performed in three stages.

1. Forward transformation. During the forward transformation vectors $x^s$, for all $s \in \mathcal{T}_{\mathcal{I}}$, are computed:

$$x^s = \left(W^s\right)^T x|_{\hat{s}}. \tag{2.10}$$

   If the cluster basis is nested, this computation can be performed recursively:

$$x^s = \begin{cases} \left(W^s\right)^T x|_{\hat{s}}, & \text{if } s \in \mathcal{L}_{\mathcal{T}_{\mathcal{I}}}, \\ \sum\limits_{t\in\operatorname{sons}(s)} \left(T_W^t\right)^T x^t, & \text{otherwise}, \end{cases} \tag{2.11}$$

   where $T_W^s$ are the transfer matrices of the cluster basis $(W^s)_{s\in\mathcal{T}_{\mathcal{I}}}$.

2. Multiplication. Let $R_t = \{s \in \mathcal{T}_{\mathcal{I}} : (t,s) \in \mathcal{L}_+(\mathcal{T}_{\mathcal{I}\times\mathcal{I}})\}$, $t \in \mathcal{T}_{\mathcal{I}}$. The result of the multiplication is

$$y^t = \sum_{s\in R_t} S^{t,s} x^s, \tag{2.12}$$

for all clusters $t \in \mathcal{T}_{\mathcal{I}}$.

3. Backward transformation. The result of the backward transformation is the vector $(y_j)_{j\in\mathcal{I}}$, given by

$$y_j = \sum_{t\in\mathcal{T}_{\mathcal{I}}:\ j\in\hat{t}} \left(V^t y^t\right)_j .$$

If the cluster basis is nested, this computation is performed recursively, similarly to the forward transformation. For $s \in \mathcal{T}_{\mathcal{I}}$ we first recursively compute:

$$y\,|_{\hat{s}} = y^s + T_V^s y^t, \qquad s \in \mathrm{sons}(t), \tag{2.13}$$

where $T_V^s$ are the transfer matrices of the cluster basis $(V^s)_{s\in\mathcal{T}_{\mathcal{I}}}$.

Next, for all $i \in \mathcal{I}$

$$y_i = \left(V^s y^{\hat{s}}\right)_i,\ s \in \mathcal{T}_{\mathcal{I}},\ i \in \hat{s}. \tag{2.14}$$

4. The non-admissible blocks are treated as in the case of $\mathcal{H}$-matrices.

**Remark 2.5.5.** *In the FMM the forward transformation and multiplication stages are split and done per level (rather than for the full cluster tree at once). For the level $\ell$ (starting from the lowest one), the vectors $x^s$ for this level are constructed (forward transformation), next the multiplication is done for all admissible block-clusters at the same level (multiplication). The same procedure is performed for the level $\ell-1$ (with the forward transformation that uses the vectors $x^s$ constructed at the level $\ell$) and so on. This procedure allows to avoid storing the result of the forward transformation for the full cluster tree, but rather keep it only per level.*

**Remark 2.5.6.** *If, for a given $\ell > 1$, there are no admissible clusters at the levels $k : 1 \le k < \ell$ of the block-cluster tree, there is no need to perform the forward and backward transformation for the levels $k < \ell$: for all such clusters $t$, $R_t = \emptyset$ (see Step 2 above), hence they do not contribute to the whole matrix-vector product.*

# Chapter 3

# High-Frequency Fast Multipole Method

In this section we describe the high-frequency fast multipole method (HF FMM) of [75, 27] in the framework of $\mathcal{H}^2$-matrices concentrating on technical and algorithmic questions. Before describing the method, we review the definitions of related special functions.

## 3.1 Special Functions

The Legendre polynomials are defined as

$$P_n(z) = \frac{1}{2^n n!} \frac{d^n}{dz^n}(z^2 - 1)^n.$$

They are orthogonal with respect to the $L_2$-product on $[-1,\ 1]$:

$$\int_{-1}^{1} P_n(x) P_m(x) dx = \frac{2}{2n+1} \delta_{nm},$$

where $\delta_{nm}$ is the Kronecker delta. The following classical theorem (see [35]) describes the convergence rate of the Legendre approximation to analytic functions.

**Theorem 3.1.1.** *Let $f(z)$ be analytic in the interior of a Bernstein ellipse*

$$E_\rho = \left\{ \frac{\rho \mathrm{e}^{i\phi} + \rho^{-1} \mathrm{e}^{-i\phi}}{2},\ \phi \in [0,\ 2\pi) \right\}$$

*for some $\rho > 1$, but not in the interior of any $E_{\rho'}$, with $\rho' > \rho$. Then*

$$f(z) = \sum_{n=0}^{\infty} a_n P_n(z)$$

*with*

$$a_n = \frac{2n+1}{2} \int_{-1}^{1} f(x) P_n(x) dx.$$

*The series converges absolutely and uniformly on any closed set in the interior of $E_\rho$ and diverges in the exterior of $E_\rho$. Moreover,*

$$\limsup_{n \to \infty} |a_n|^{\frac{1}{n}} = \frac{1}{\rho}.$$

The associated Legendre functions are defined with the help of the Legendre polynomials:

$$\begin{aligned} P_n^m(x) &= (-1)^m (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x), \qquad 0 \le m \le n, \\ P_n^{-m}(x) &\equiv (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x), \\ P_n^0(x) &\equiv P_n(x), \\ P_n^m(x) &\equiv 0, \qquad |m| > n. \end{aligned}$$

The normalized associated Legendre functions

$$\begin{aligned} \bar{P}_n^m &= \sqrt{\left(n+\frac{1}{2}\right)\frac{(n-m)!}{(n+m)!}} (1-x^2)^{\frac{m}{2}} \frac{d^m}{dx^m} P_n(x), \\ \bar{P}_n^m &\equiv \bar{P}_n^{-m}, \qquad m < 0. \end{aligned} \tag{3.1}$$

By $\hat{r}, \hat{s}, ...$ we denote unit vectors in $\mathbb{R}^3$, namely, given a vector $x \in \mathbb{R}^3$,

$$\hat{x} = \frac{x}{\|x\|}.$$

The spherical coordinates of a vector in $\mathbb{R}^3$ are given by $(\rho, \phi, \theta)$, with $\phi$ being the azimuth and $\theta$ the inclination. Then the Cartesian coordinates of a vector $\hat{s}$ on the unit sphere are read as

$$\hat{s} = (\cos\phi \sin\theta,\ \sin\phi \sin\theta,\ \cos\theta), \qquad \phi \in [0,\ 2\pi]\,,\ \theta \in [0,\ \pi]\,. \tag{3.2}$$

A spherical harmonic of degree $n$ and order $m$ is a function

$$\begin{aligned} Y_n^m &: \mathbb{S}^2 \to \mathbb{C}, \\ Y_n^m(\hat{s}) &\equiv Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} e^{im\phi} P_n^m(\cos\theta), \qquad |m| \le n, \\ Y_n^{m*}(\hat{s}) &\equiv (-1)^m Y_n^{-m}(\hat{s}). \end{aligned} \tag{3.3}$$

These functions constitute an orthonormal basis of $L_2(\mathbb{S}^2)$ and

$$\int_{\mathbb{S}^2} Y_n^m(\hat{s}) Y_p^{l*}(\hat{s}) d\hat{s} = \delta_{np}\delta_{ml}. \tag{3.4}$$

They are connected to the Legendre polynomials via the addition theorem. Given two unit vectors $\hat{x}$, $\hat{y}$,

$$P_n(\hat{x} \cdot \hat{y}) = \frac{4\pi}{2n+1} \sum_{m=-n}^{n} Y_n^m(\hat{x}) Y_n^{m*}(\hat{y}). \tag{3.5}$$

Spherical Bessel functions of the first kind $j_n(x)$ and spherical Bessel functions of the third kind $h_n^{(1)}(x)$, $h_n^{(2)}(x)$ are defined as in [2, (10.1.1)]. We denote

$$h_n(x) \equiv h_n^{(1)}(x).$$

The analytic expressions for these functions are given by Rayleigh's formulas:

$$j_n(z) = z^n \left(-\frac{1}{z}\frac{d}{dz}\right)^n \frac{\sin z}{z}, \tag{3.6}$$

$$h_n(z) = j_n(z) - iz^n \left(-\frac{1}{z}\frac{d}{dz}\right)^n \frac{\cos z}{z}. \tag{3.7}$$

Spherical Bessel functions are the coefficients of the expansion of the plane-wave function in the Legendre polynomial basis [2, (10.1.47)]:

$$e^{iz\cos\theta} = \sum_{n=0}^{\infty}(2n+1)i^n j_n(z)P_n(\cos\theta). \tag{3.8}$$

**Remark 3.1.2.** *Theorem 3.1.1 allows to conclude that the series (3.8) converges supergeometrically, since the function* $\mathrm{e}^{izt}$ *is entire in* $t$.

The following is a particular case of the Funk-Hecke theorem [54, Theorem 3.4.1].

**Theorem 3.1.3** (Funk-Hecke theorem)**.** *Let* $f$ *be a bounded integrable function on* $[-1,1]$. *Then* $f_{\hat{\alpha}}(\hat{s}) = f(\hat{\alpha}\cdot\hat{s})$, $\hat{\alpha}\in\mathbb{S}^2$, *is integrable on* $\mathbb{S}^2$ *and, for all* $n\in\mathbb{N}$,

$$\int\limits_{\mathbb{S}^2} f(\hat{\alpha}\cdot\hat{s})P_n(\hat{q}\cdot\hat{s})d\hat{s} = a_n(f)P_n\left(\hat{q}\cdot\hat{\alpha}\right),$$

*where*

$$a_n(f) = 2\pi\int\limits_{-1}^{1} f(t)P_n(t)dt.$$

The next identity can be immediately derived from the above theorem combined with (3.8) and Theorem 3.1.1:

$$\int\limits_{\mathbb{S}^2} \mathrm{e}^{i\lambda\hat{y}\cdot\hat{s}}P_k(\hat{s}\cdot\hat{x})d\hat{s} = 4\pi i^k j_k(\lambda)P_k(\hat{y}\cdot\hat{x}),\ k\in\mathbb{N},\ \lambda\in\mathbb{C}. \tag{3.9}$$

The following expression serves as the basis for the fast multipole method and is known under the name 'addition theorem' (or 'Gegenbauer's addition theorem'), see [2, (10.1.45), (10.1.46)]:

$$h_0(\kappa\|x-y\|) = \sum_{n=0}^{\infty}(2n+1)h_n(\kappa\|x\|)j_n(\kappa\|y\|)P_n(\hat{x}\cdot\hat{y}), \tag{3.10}$$
$$x,y\in\mathbb{R}^3 : \|x\| > \|y\|.$$

Another component of the fast multipole method is numerical integration over the unit sphere. In the fast multipole method literature it is often performed with the help of the quadrature rule introduced in the following lemma from [75].

**Lemma 3.1.4.** *Let* $f$ *be a spherical harmonic of degree* $n_1$, *and* $g$ *be a spherical harmonic of degree* $n_2$, $f = f(\hat{s})$, $g = g(\hat{s})$, *where* $\hat{s}$ *is given by (3.2). For any* $n_\theta \geq \left\lceil \frac{n_1+n_2+1}{2} \right\rceil$, $n_\phi \geq n_1+n_2+1$ *the quadrature rule on the unit sphere given by the nodes and weights*

$$\begin{aligned}(\phi_k,\theta_j) &= \left((k-1)\frac{2\pi}{n_\phi}, \arccos x_j\right),\\ w_{kj} &= \frac{2\pi}{n_\phi}\omega_j, \qquad k=1,..,n_\phi,\ j=1,..,n_\theta,\end{aligned} \tag{3.11}$$

*with* $(x_j)_{j=1}^{n_\theta}$, $(\omega_j)_{j=1}^{n_\phi}$ *being Gaussian quadrature nodes and weights on the interval* $[-1,1]$, *integrates the product of* $f$ *and* $g$ *exactly.*

*Proof.* Integration of the product of $f$ and $g$ requires the evaluation of the integrals of the type, see (3.3),

$$\int\limits_0^{2\pi} \mathrm{e}^{im\phi}\mathrm{e}^{im'\phi}d\phi \int\limits_{-1}^{1} P_{n_1}^m(x)P_{n_2}^{m'}(x)dx,$$

where $m, m' \in \mathbb{Z}$, $|m| \le n_1,\ |m'| \le n_2$. The first integral can be evaluated exactly with the help of the trapezoidal rule with at least $n = n_1 + n_2 + 2$ points, more precisely, the quadrature with nodes $\tilde{\phi}_k$ and weights $\tilde{\omega}_k$, $k = 1, \dots, n$:

$$\tilde{\phi}_k = \frac{2\pi}{n-1}(k-1),$$
$$\tilde{\omega}_1 = \tilde{\omega}_n = \frac{2\pi}{2(n-1)},\ \tilde{\omega}_j = \frac{2\pi}{n-1}, \qquad 2 \le j \le n-1.$$

The integrand of

$$\int_{-1}^{1} P_{n_1}^{m}(x) P_{n_2}^{-m}(x) dx$$

is a polynomial of the degree not larger than $n_1 + n_2$, hence this integral can be integrated exactly with any Gaussian quadrature rule with $\left\lceil \frac{n_1+n_2+1}{2} \right\rceil$ points. □

Here we employ the Gauss-Legendre quadrature. The abscissas of the quadrature of the order $n$ are given by the zeros of the Legendre polynomial $P_n$, and the weights by

$$w_\ell = \frac{2}{(1-x_\ell)^2 \left(P_n'(x_\ell)\right)^2},\ \ell = 1, \dots, n.$$

**Remark 3.1.5.** *In what follows we use the quadrature rule with* $n_\theta = \left\lceil \frac{n_1+n_2+1}{2} \right\rceil$ *and* $n_\phi = 2n_\theta$.

**Remark 3.1.6.** *We adopt a short notation for the quadrature rule defined in Lemma (3.1.4):*

$$(w_\ell, \hat{s}_\ell)_{\ell=1}^{L},\ (w_\ell, \hat{r}_\ell)_{\ell=1}^{L}, \dots \tag{3.12}$$

*stands for a quadrature rule with* $L = 2n_\theta^2$, *and* $\hat{s}_\ell$ $(\hat{r}_\ell, \dots)$ *is a vector (3.2) with* $\phi$, $\theta$ *given by (3.11).*

**Remark 3.1.7.** *We will denote the integral* $\int_{\mathbb{S}^2} f(\hat{s}) d\hat{s}$ *computed with the help of the quadrature* $(w_\ell, \hat{s}_\ell)_{\ell=1}^{L}$ *by* $Q_L\left[f(\hat{s})\right]$. *When necessary, the variable of the integration is stated explicitly in the upper index:*

$$\int_{\mathbb{S}^2} f(\hat{s} \cdot \hat{r}) d\hat{s} \approx Q_L^{\hat{s}}\left[f(\hat{s} \cdot \hat{r})\right]. \tag{3.13}$$

Additionally, we will use the following lemma which is a straightforward corollary of Lemma 3.1.4 and (3.5).

**Lemma 3.1.8.** *Given* $M \in \mathbb{N}_+$ *and* $m, n \in \mathbb{N}_0 :\ \left\lceil \frac{m+n+1}{2} \right\rceil \le M$,

$$Q_M^{\hat{s}}\left[P_m\left(\hat{q} \cdot \hat{s}\right) P_n\left(\hat{r} \cdot \hat{s}\right)\right] = \begin{cases} \frac{4\pi}{2n+1} P_n\left(\hat{q} \cdot \hat{r}\right), & \text{if } n = m, \\ 0, & \text{otherwise.} \end{cases}$$

*for all* $\hat{q}, \hat{r} \in \mathbb{S}_2$.

*Proof.* We use the addition theorem (3.5) for Legendre functions to rewrite

$$Q_M^{\hat{s}}\left[P_m\left(\hat{q} \cdot \hat{s}\right) P_n\left(\hat{r} \cdot \hat{s}\right)\right] = \frac{4\pi}{2m+1} \frac{4\pi}{2n+1} \sum_{\ell=-m}^{m} Y_m^{\ell}(\hat{q}) \sum_{k=-n}^{n} Y_n^{k*}(\hat{r}) Q_M^{\hat{s}}\left[Y_m^{\ell*}(\hat{s}) Y_n^{k}(\hat{s})\right].$$

The quadrature rule of order $M$ as in the statement of the lemma integrates the product of these spherical harmonics exactly. Hence for $m \neq n$ the result follows from orthogonality of spherical harmonics, see (3.4). For $m = n$, we again employ (3.4) to get

$$\begin{aligned} Q_M^{\hat{s}}\left[P_m\left(\hat{q} \cdot \hat{s}\right) P_n\left(\hat{r} \cdot \hat{s}\right)\right] &= \left(\frac{4\pi}{2n+1}\right)^2 \sum_{\ell=-m}^{m} Y_m^{\ell}(\hat{q}) Y_m^{\ell*}(\hat{r}) \\ &= \frac{4\pi}{2n+1} P_n\left(\hat{q} \cdot \hat{r}\right). \end{aligned}$$

□

## 3.2 High-Frequency Fast Multipole Algorithm

The high-frequency fast multipole method is based on the expansion (3.10). Namely, given $s \in \mathbb{C}$, $x, y, x_\beta, y_\alpha \in \mathbb{R}^3$, it holds:

$$\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|} = -\frac{s}{4\pi}\sum_{n=0}^{\infty}(2n+1)h_n(is\|c_{\alpha\beta}\|)j_n(is\|y-x+c_{\alpha\beta}\|)P_n(\hat{c}_{\alpha\beta}\cdot\hat{r}_{\alpha\beta}),$$
$$\text{for } \|c_{\alpha\beta}\| > \|r_{\alpha\beta}\|,$$

where $c_{\alpha\beta} = y_\alpha - x_\beta$ and $r_{\alpha\beta} = x - y + c_{\alpha\beta}$.

Truncating the above series at $N+1$ terms, employing (3.9) and interchanging the limits of integration, gives

$$\begin{aligned}\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|} = -\frac{s}{16\pi^2}\sum_{n=0}^{N}(2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|)\int_{\mathbb{S}^2}\mathrm{e}^{-s(r_{\alpha\beta},\hat{r})} \\ \times P_n(\hat{c}_{\alpha\beta}\cdot\hat{r})d\hat{r} + E_{tr}(N),\end{aligned} \tag{3.14}$$

where $E_{tr}(N)$ is the truncation error.

The next step is the discretization. The addition theorem (3.5) combined with (3.8) shows that the integrand of (3.14) is a sum of products of spherical harmonics, hence the quadrature rule of Lemma 3.1.4 can be employed. In [32] it was suggested that $L$ should be chosen so that $L \geq 2(N+1)^2$. This gives the following separable expansion of the Helmholtz kernel:

$$\begin{aligned}\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|} = \sum_{k=0}^{L} w_k \mathrm{e}^{s(y-y_\alpha,\hat{r}_k)}\left(-\frac{s}{(4\pi)^2}\sum_{n=0}^{N}(2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|)P_n(\hat{c}_{\alpha\beta}\cdot\hat{r}_k)\right) \\ \times \mathrm{e}^{-s(x-x_\beta,\hat{r}_k)} + E_{tr}(N) + E_I(L,N),\end{aligned}$$

where $E_I$ is the integration error.

Another way to discretize (3.14) based on the modification of the integrand and the use of the trapezoidal quadrature rule was recently suggested in [80, 26].

The fast multipole method can be cast into the framework of $\mathcal{H}^2$-matrices. Before, this was done in [8, 3]. Namely, the matrix-vector multiplication described in Section 2.5 can be viewed as a generalized description of the fast multipole algorithm, with properly defined cluster basis, transfer and coupling matrices.

Let us consider the Galerkin discretization of the Helmholtz single layer boundary operator

$$(\boldsymbol{V}(s))_{ij} = \iint_{\Gamma\times\Gamma}\phi_i(x)\phi_j(y)\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}d\Gamma_x d\Gamma_y, \qquad i,j = 1,\ldots,M.$$

In this section we show how this matrix can be approximated with the help of the fast multipole method. We will comment on the choice of the parameters and the error control in the further sections.

We fix the block-cluster tree $\mathcal{T}_{\mathcal{I}\times\mathcal{I}}$. We use the uniform partition of the domain, i.e. all the bounding boxes of the clusters located at the same level of the block-cluster tree are of the same size.

With each level $\ell$ of the cluster tree we associate a set of quadrature nodes on the unit sphere

$$(w_n, \hat{r}_n)_{n=1}^{L_\ell},\ L_\ell \in \mathbb{N}, \tag{3.15}$$

defined as in Lemma 3.1.4.

The version of the algorithm described here is the high-frequency FMM of [27] with minor modifications.

### 3.2.1 Cluster Basis

Given a cluster $\tau_\alpha$ located at the level $\ell$ of the cluster tree and $c_\alpha$ being the center of its bounding box, we define the column cluster basis as a matrix

$$\left(W^{\tau_\alpha}\right)_{k_j n}(s) = \int_{\tau_\alpha} \mathrm{e}^{-s(x-c_\alpha,\hat{r}_n)}\phi_j(x)d\Gamma_x, \tag{3.16}$$
$$k_j \in \{1,\dots,\#\hat{\tau}_\alpha\},\ j \in \hat{\tau}_\alpha,\ n = 1,\cdots,L_\ell.$$

The row cluster basis for the cluster $\tau_\alpha$ has a different form:

$$\left(V^{\tau_\alpha}\right)_{k_j n}(s) = w_n \int_{\tau_\alpha} \mathrm{e}^{s(x-c_\alpha,\hat{r}_n)}\phi_j(x)d\Gamma_x, \qquad k_j \in \{1,\dots,\#\hat{\tau}_\alpha\},\ j \in \hat{\tau}_\alpha,\ n = 1,\dots,L_\ell.$$

**Efficient Computation and Storage of the Cluster Basis**

It is sufficient to compute the column cluster basis only, whereas the row cluster basis can be constructed based on the symmetry of the quadrature points (3.11) on the unit sphere. Let $p := x - c_\alpha = (p_1,\ p_2,\ p_3) \in \mathbb{R}^3$. For quadrature nodes $\hat{r} = (\cos\phi_k \sin\theta_j,\ \sin\phi_k \sin\theta_j,\ \cos\theta_j)^T$, it holds that

$$\mathrm{e}^{s(x-c_\alpha,\hat{r})} = \mathrm{e}^{s(p_1\cos\phi_k\sin\theta_j + p_2\sin\phi_k\sin\theta_j + p_3\cos\theta_j)} = \mathrm{e}^{-s(p,\hat{q}_{kj})},$$
$$\hat{q}_{kj} = \left(\cos(\pi+\phi_k)\sin(\pi-\theta_j),\ \sin(\pi+\phi_k)\sin(\pi-\theta_j),\ \cos(\pi-\theta_j)\right).$$

Let $L_\ell = 2n_\theta^2$, $n_\theta \in \mathbb{N}$, see Remark 3.1.5. Since the nodes of the Gauss-Legendre quadrature are symmetric about 0,

$$\pi - \theta_j = \theta_{n_\theta - j + 1},$$

and also

$$\pi + \phi_k = \pi + \frac{\pi}{n_\theta}k = \frac{2\pi}{2n_\theta}\left((n_\theta+1)k \bmod 2n_\theta\right),$$

the vector $\hat{q}_{kj}$ indeed belongs to the set $(\hat{r}_n)_{n=1}^{L_\ell}$. Hence, only the column leaf cluster basis need to be computed and stored.

For some applications it is necessary to compute the cluster basis for many values of $s \in \mathbb{C}$. In this case storing all of the dense matrices (3.16) may be expensive. Alternatively, the function $f_\alpha(x,\hat{r}) = \mathrm{e}^{-s(x-c_\alpha,\hat{r})}$ can be interpolated in $x$ with the help of multivariate interpolation (as done when evaluating the boundary integrals with the help of a quadrature rule). Then the entries of the column leaf cluster basis are

$$\begin{aligned}\left(W^{\tau_\alpha}\right)_{jn}(s) &= \int_{\tau_\alpha} \mathrm{e}^{-s(x-c_\alpha,\hat{r}_n)}\phi_j(x)d\Gamma_x \\ &\approx \sum_{k=1}^{K} \mathrm{e}^{-s(x_{k,\alpha}-c_\alpha,\hat{r}_n)}\omega_{j,k}^{\alpha},\end{aligned} \tag{3.17}$$

where $x_{k,\alpha} \in \tau_\alpha$ are quadrature nodes and $\omega_{j,k}^{\alpha},\ k = 1,\dots,K$ are weights. Hence for all $s \in \mathbb{C}$ we can store only the weights $\omega_{j,k}^{\alpha}$, the interpolation points $x_{k,\alpha}$ and the centers of the clusters $c_\alpha$, and then compute $W^{\tau_\alpha}$ on the fly when reading the data from the memory to perform the matrix-vector multiplication.

### 3.2.2 Transfer Matrices

In the fast multipole method transfer matrices are represented by translation operators. Namely, for the column cluster basis transfer matrices correspond to the multipole-to-multipole (M2M) translations, and for the row cluster basis they are equivalent to local-to-local (L2L) translations.

Before defining transfer matrices, let us provide some information on one of the ingredients of these operators, namely the fast spherical harmonic transform as described in detail in [67].

**Fast Spherical Harmonic Transform**

Let

$$f : \mathbb{S}^2 \to \mathbb{C}.$$

Let us set $\hat{s}(\theta, \phi) = (\cos\phi \sin\theta,\ \sin\phi \sin\theta,\ \cos\theta)$. We assume that for some $K \in \mathbb{N}$,

$$f(\hat{s}(\theta, \phi)) = \sum_{k=0}^{K} \sum_{m=-k}^{k} f_k^m Y_k^m(\theta, \phi). \tag{3.18}$$

Given the values of the function $f$ on the grid $(\phi_\ell,\ \theta_n)_{\ell,n=1}^{n_\phi, n_\theta}$ defined by (3.11)

$$f_{\ell,n} = f(\hat{s}(\theta_n, \phi_\ell)),\ \ell = 1, \dots, n_\phi,\ n = 1, \dots, n_\theta,$$

we need to compute the values of the function

$$F(\hat{s}(\theta, \phi)) = \sum_{k=0}^{N} \sum_{m=-k}^{k} f_k^m Y_k^m(\theta, \phi), \tag{3.19}$$

on the grid of different size, namely $(\phi_\ell,\ \theta_n)_{\ell,n=1}^{n'_\phi, n'_\theta}$ defined by (3.11). We assume $n_\phi = 2n_\theta$ and $n'_\phi = 2n'_\theta$, set

$$L = 2n_\theta^2, \qquad L' = 2(n'_\theta)^2,$$

and define $R^{L',L}$ as an operator

$$F_{\ell,n} = \left(R^{L',L} f\right)_{\ell,n} = F\left((\cos\phi'_\ell \sin\theta'_n,\ \sin\phi'_\ell \sin\theta'_n,\ \cos\theta'_n)\right),$$
$$\ell = 1, \dots, n'_\phi,\ n = 1, \dots, n'_\theta.$$

To perform the truncation exactly, $n_\phi$ has to be chosen so that

$$n_\phi \geq K + N + 1. \tag{3.20}$$

A trivial algorithm for the spherical harmonic transform can be described in two steps.

1. Evaluate $f_m^n$ using the quadrature rule from Lemma 3.1.4:

$$f_n^m = \int_{\mathbb{S}^2} f(\hat{s}) Y_n^{-m}(\hat{s}) d\hat{s} = \sum_{\ell=1}^{n_\phi} \sum_{k=1}^{n_\theta} f_{\ell,k} w_{k,l}. \tag{3.21}$$

2. Define $F$ as in (3.19) and evaluate $F$ on the corresponding grid:

$$F_{\ell,n} = \sum_{k=0}^{N} \sum_{m=-k}^{k} f_k^m Y_m^k(\theta'_n, \phi'_\ell), \qquad \ell = 1, \dots, n'_\phi,\ n = 1, \dots, n'_\theta. \tag{3.22}$$

The fast spherical harmonic transform makes use of the structure of the sums (3.21,3.22) exploiting the fast Fourier and Legendre transforms. This algorithm proceeds as follows.

1. For every $n = 1, \dots, n_\theta$, compute

$$\hat{f}_n^m = \frac{2\pi}{n_\phi} \sum_{\ell=1}^{n_\phi} f_{\ell,n} \mathrm{e}^{i\frac{2\pi}{n_\phi}(\ell-1)m}, \qquad m = -n_\theta + 1, \dots, n_\theta - 1.$$

with the help of the inverse fast Fourier transform. Particularly, for $m < 0$

$$\hat{f}_n^m = \hat{f}_n^{(m+n_\phi) \bmod n_\phi}.$$

This operation is of the complexity $O(n_\theta n_\phi \log n_\phi)$.

2. For every $m = -N, \dots, N$, $k = 1, \dots, n'_\theta$, evaluate

$$\hat{F}_k^m = \epsilon_{N+1}^m \sum_{n=1}^{n_\theta} \hat{f}_n^m \tilde{w}_n \frac{\bar{P}_{N+1}^m(\cos\theta'_k)\bar{P}_N^m(\cos\theta_n) - \bar{P}_N^m(\cos\theta'_k)\bar{P}_{N+1}^m(\cos\theta_n)}{\cos\theta'_k - \cos\theta_n},$$

where $\bar{P}_n^m$ are the normalized associated Legendre functions, see (3.1), $\tilde{w}_n$ are the weights of the Gauss-Legendre quadrature of the order $n_\theta$ and

$$\epsilon_n^m = \sqrt{\frac{n^2 - m^2}{4n^2 - 1}}.$$

If $\cos\theta_n = \cos\theta'_n$, the quotient can be evaluated using l'Hôpital's rule. For $m = -N, \dots, N$ the matrices

$$(\mathcal{P}_1^m)_{kn} = \frac{\bar{P}_{N+1}^m(\cos\theta'_k)\bar{P}_N^m(\cos\theta_n)}{\cos\theta'_k - \cos\theta_n},$$
$$(\mathcal{P}_2^m)_{kn} = \frac{\bar{P}_N^m(\cos\theta'_k)\bar{P}_{N+1}^m(\cos\theta_n)}{\cos\theta'_k - \cos\theta_n}, \qquad k = 1, \dots, n'_\theta,\ n = 1, \dots, n_\theta,$$

can be efficiently represented in the $\mathcal{H}$-matrix format (as Nyström discretizations of the asymptotically smooth kernels, see Section 2.3) or with the help of the one-dimensional fast multipole method, see [90].

This operation can be performed with the asymptotic complexity $O(Nn'_\theta \log n'_\theta)$.

3. Compute the quantities

$$F_{m,n} = \sum_{\ell=-N}^{N} \hat{F}_n^\ell \mathrm{e}^{-i\frac{2\pi}{n'_\phi}\ell(m-1)}, \qquad m = 1, \dots n'_\phi,\ n = 1, \dots, n'_\theta,$$

with the help of the fast Fourier transform.

Note that from the description of the spherical harmonics transform it follows that, for all $N \in \mathbb{N}_+$ and bandlimited functions $f(\hat{s}) = \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} f_{\ell,m} Y_\ell^m(\hat{s})$

$$R^{N,N} f = f.$$

**Remark 3.2.1.** *In our implementation of the fast multipole algorithm, we use*

$$n_\theta = K + 1, \qquad n_\phi = 2n_\theta,$$
$$n'_\theta = N + 1, \qquad n'_\phi = 2n'_\theta.$$

*In the course of the fast multipole algorithm it is also necessary to evaluate the function given on the 'new' grid* $\left(\theta'_j, \phi'_k\right)$, $j = 1, \dots, n'_\theta, k = 1, \dots, n'_\phi$ *on the 'old' grid* $(\theta_j, \phi_k)$, $j = 1, \dots, n_\theta, k = 1, \dots, n_\phi$. *The algorithm proceeds as in 1-3, interchanging in the description* $n_\phi$ *and* $n'_\phi$ *and* $n_\theta$ *and* $n'_\theta$. *The matrix-vector multiplication in Step 2 has to be substituted by the matrix-vector multiplication with transposed matrices* $(\mathcal{P}_1^m)^T$, $(\mathcal{P}_2^m)^T$.

The transpose of the spherical harmonics transform is

$$\left(R^{N,M}\right)^T = R^{M,N}.$$

**Transfer Matrices (M2M and L2L Translation Operators)**

Let clusters $\tau_\alpha \notin \mathcal{L}_{\mathcal{T}_\mathcal{I}}$ and $\tau_\beta \in \text{sons}(\tau_\alpha)$ be located correspondingly at the levels $k$ and $k+1$ of the cluster tree. Let the centers of their bounding boxes be $c_\alpha,\ c_\beta$. Then the translation operators for the column cluster basis are defined as

$$T_c^{\tau_\beta}(s) = R^{L_{k+1},L_k} D^{\tau_\alpha,\tau_\beta}(-s),$$

where $R^{L_{k+1},L_k}$ is the fast spherical harmonics transform and $D^{\tau_\alpha,\tau_\beta}(s)$ is a diagonal translation operator. Its entries are explicitly given by

$$D_{\ell\ell}^{\tau_\alpha,\tau_\beta}(s) = \exp\left(s(c_\beta - c_\alpha, \hat{r}_\ell)\right), \qquad \ell = 1,\dots,L_k, \tag{3.23}$$

where $(\hat{r}_\ell)_{\ell=1}^{L_k}$ are as in (3.15).

The translation operators for the row cluster basis are defined similarly:

$$T_r^{\tau_\beta}(s) = R^{L_{k+1},L_k} D^{\tau_\alpha,\tau_\beta}(s). \tag{3.24}$$

**Efficient Computation and Storage of Translation Operators**

Let us consider a cluster $\tau_\beta$ with the bounding box centered at $c_\beta$ and its parent cluster $\tau_\alpha$ (whose bounding box is centered at $c_\alpha$). The cluster $\tau_\beta$ is located at the level $\ell$ of the cluster tree, and the cluster $\tau_\alpha$ at the level $\ell-1$. Then the multipole-to-multipole (local-to-local) translation operator $T_c^{\tau_\beta}$ depend only on the cluster basis rank $L_\ell$ on the level $\ell$, cluster basis rank $L_{\ell-1}$ on the level $\ell-1$ and on the $c_{\alpha\beta} = c_\beta - c_\alpha$. If the uniform partition of the domain is used, there exists only a fixed number of different $c_{\alpha\beta}$ per level, see Figure 3.1. Hence only a few translation operators need to be constructed and stored (and this is the reason to use the uniform partition of the domain).

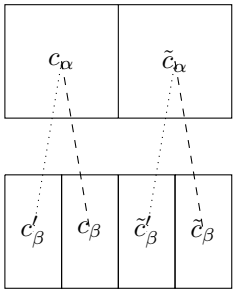


*Figure 3.1: Bounding boxes on two levels of a uniform binary cluster tree. In this case only two translation matrices per level are needed.*

### 3.2.3 Multipole-to-Local Operators, or Coupling Matrices

Given an admissible block-cluster $b = (\tau_\alpha, \tau_\beta)$ located at the level $\ell$ of the block-cluster tree, the corresponding multipole-to-local translation operator is defined as, see [33, 22],

$$S^b = D^b, \tag{3.25}$$

where $D^b$ is a diagonal matrix with elements

$$D_{kk}^b = -\frac{s}{16\pi^2} \sum_{n=0}^{n_b-1} (2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|) P_n(\hat{c}_{\alpha\beta}\cdot \hat{r}_k), \qquad k = 1,\dots,L_\ell, \tag{3.26}$$

where $c_{\alpha\beta} = c_\alpha - c_\beta$. Recall that $(\hat{r}_k)_{k=1}^{L_\ell}$ are the nodes of the quadrature on the unit sphere, see also (3.15).

**Remark 3.2.2.** *In the work [27], the multipole-to-local operator is defined slightly differently, namely,*

$$S^b = R^{L_\ell, N_\ell} \tilde{D}^b R^{N_\ell, L_\ell}, \tag{3.27}$$

*where $\tilde{D}^b$ is a diagonal matrix*

$$\tilde{D}^b_{kk} = -\frac{s}{16\pi^2} \sum_{n=0}^{n_b-1} (2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|) P_n(\hat{c}_{\alpha\beta} \cdot \hat{s}_k), \qquad k = 1, \ldots, N_\ell. \tag{3.28}$$

*Here $(\hat{s}_k)_{k=1}^{N_\ell}$ is the set of quadrature points on the unit sphere, see Lemma 3.1.4. Given $L_\ell = 2n_\ell^2$, an accurate choice of $N_\ell$ is*

$$N_\ell = 2n_{max}^2, \qquad n_{max} \geq \left\lceil \frac{2n_\ell + n_b + 1}{2} \right\rceil, \tag{3.29}$$

*for all admissible $b$ located at the level $\ell$, see also (3.20) and [27]. However, in practice, a slightly more efficient $N_\ell = 2 \max\limits_{b \in \mathcal{L}_+^\ell} n_b^2$ does not deteriorate the accuracy (this value also coincides with the heuristic suggested in [27] for the non-decay case).*

*In the present work we use the coupling matrices defined by (3.25), rather than (3.27): our numerical experiments did not encounter a significant deterioration of accuracy when a simpler and more efficient (3.25) is used.*

**Efficient Construction and Storage of Multipole-to-Local Operators**

A straightforward computation of the diagonal translation matrix (3.26) would require $O(n_b L_\ell)$ operations, which, for $s = -i\kappa$, $\kappa \in \mathbb{R}$, scales as $O(\kappa^3)$ (because of $L_\ell = O(\kappa^2)$ and $n_b = O(\kappa)$), see Section 4.2. Although this operation, as we show in the second part of this section, is repeated only a constant number of times per level, it can potentially destroy asymptotic complexity estimates of the fast multipole algorithm (see also [27]). There are several ways to deal with this problem, namely, the use of the Clenshaw summation algorithm [29] or the local interpolation approach, briefly described in [27]. We used the method that bears similarities with the latter one. More specifically, the function (see the expression (3.26))

$$f(t) = -\frac{s}{(4\pi)^2} \sum_{n=0}^{n_b-1} (2n+1) i^n h_n(is\|c_{\alpha\beta}\|) P_n(t),$$

is a polynomial in $t \in [-1,\ 1]$ of degree $n_b - 1$, hence can be represented by its values in $n_b$ Chebyshev points $\{t_j\}_{j=1}^{n_b}$ of the second kind. The evaluation at any other point $p \in [-1,\ 1]$ can be done with the help of the barycentric Lagrange interpolation [79]:

$$f(p) = \begin{cases} \dfrac{\sum\limits_{j=1}^{n_b}{}' (-1)^j \frac{f(t_j)}{p - t_j}}{\sum\limits_{j=1}^{n_b}{}' \frac{(-1)^j}{p - t_j}}, & p \neq t_j, \\ f(t_j), & p = t_j, \end{cases} \tag{3.30}$$

where the prime indicates that the terms $j = 1$ and $j = n_b$ are multiplied by $\frac{1}{2}$. Our task is to evaluate this fraction for $O(\kappa^2)$ points $p = p_1, \ldots, p_{L_\ell}$. Clearly, summations in the numerator and the denominator can be viewed as the multiplication of the matrix

$$M_{ij} = \begin{cases} \frac{1}{p_i - t_j}, & p_i \neq t_j, \\ 0, & \text{else}, \end{cases}$$
$$i = 1, \ldots, L_b,$$

by corresponding vectors. This matrix, in turn, for large $n_b,\ L_\ell$ can be efficiently approximated with the help of $\mathcal{H}$-matrix techniques (see Section 2), and the evaluation of (3.30) for $L_b$ points will require at most $O(L_\ell \log L_\ell)$ operations. The cases $p_j = t_i$ should be treated explicitly. The disadvantage of this method is that it needs the $\mathcal{H}$-matrix approximation to be very accurate and hence is efficient only for rather big values of $n_b,\ L_\ell$.

As before, the symmetry of the quadrature on the unit sphere, as well as the uniformity of the block-cluster tree allow us to construct and store per level only a small number of multipole-to-local translations (see also [22, 41]). This is due to the fact that the elements of the matrix $D^b$

$$D^b_{kk} = -\frac{s}{16\pi^2}\sum_{n=0}^{n_b-1}(2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|)P_n(\hat{c}_{\alpha\beta}\cdot\hat{r}_k), \qquad k=1,\ldots,N_\ell,$$

depend only on the direction $\hat{c}_{\alpha\beta}$ and on the distance $d_b = \|c_{\alpha\beta}\|$. The value $n_b$, as we show later, depends only on $d_b$ and the size of a cluster at the level $\ell$. Hence the elements of the matrix $D^b$ can be obtained by permuting the diagonal entries of the matrix

$$\tilde{D}^b_{kk} = -\frac{s}{16\pi^2}\sum_{n=0}^{n_b-1}(2n+1)(-i)^n h_n(is\|c_{\alpha\beta}\|)P_n(\hat{\lambda}\cdot\hat{r}_k), \qquad k=1,\ldots,N_\ell,$$

where $\hat{\lambda} = [\eta\hat{c}_{\alpha\beta,1},\ \mu\hat{c}_{\alpha\beta,2},\ \nu\hat{c}_{\alpha\beta,3}]$, with $\eta,\mu,\nu \in \{-1,+1\}$.

A more efficient realization of (3.27) in (2.12) reads as

$$y^t = R^{L_\ell,N_\ell}\sum_{v\in R_t} D^b R^{L_\ell,N_\ell} x^v.$$

**Remark 3.2.3.** *Although sections on efficient construction and storage of the cluster basis and translation matrices may seem to provide unnecessary technical details, they are* ***crucial*** *for the implementation of the fast multipole algorithm. An algorithm implemented without them appears to be unpractical even for quite large problems (with* $10^5$ *unknowns).*

# Chapter 4

# Error Analysis of the FMM for the Helmholtz Equation with Decay

The question of the proper choice of lengths of expansions in the fast multipole method had been intensively studied in [30, 73, 32, 33, 69, 25, 24]. Since Bessel functions are rather difficult to analyze, in most cases the error analysis is based on asymptotic expansions or explicit bounds on these functions; a very precise, near-optimal error analysis was recently made in [24, 25]. To our knowledge, these works use the fact that the wavenumber is purely real, and hence cannot be in a straightforward way adapted to the complex wavenumber case. In the recent works [43, 51] the authors numerically investigate the error of the truncation of the Gegenbauer's series in the case when the wavenumber has a decaying part, as well as provide empirical formulas for the choice of the length of the expansion. In this section we study analytically the error of the truncation of the Gegenbauer's series for the complex wavenumber, as well as analyze other sources of errors of the fast multipole method.

## 4.1 Behavior of Spherical Bessel and Hankel Functions

First we examine the behavior of spherical Bessel and Hankel functions of a complex argument. There exists a wide range of literature on these functions, see e.g. classical monographs [72, 89], and their asymptotic behavior is to a large extent known. We summarize these results here. First we consider spherical Bessel functions $j_n(z),\ \arg z \in (0,\ \pi)$ in different regimes.

1. **Fixed order $n$, small argument $z$.** As $z \to 0$ [2, (9.1.7)]:

$$j_n(z) \to \frac{\sqrt{\pi} z^n}{2^{n+1}} \frac{1}{\Gamma(n+\frac{3}{2})}.$$

.

2. **Order $n:\ n < |z|$, $|z| \to +\infty$.** We are interested in the range $\arg z \in (0,\ \pi)$. The expression [2, (10.1.14)]

$$j_n(z) = \frac{1}{2}(-i)^n \int_{-1}^{1} \mathrm{e}^{izt} P_n(t) dt \tag{4.1}$$

shows that

$$|j_n(-\bar{z})| = |j_n(z)|\,.$$

For $z \in \left(0, \frac{\pi}{2}\right)$, the following asymptotic expansion holds for $j_n(z)$, see expressions (10.17.13, 10.17.14, 10.17.15) combined with (10.4.4) in [1]:

$$j_n(z) = \frac{1}{2z}\left((-i)^{n+1}\mathrm{e}^{iz} + i^{n+1}\mathrm{e}^{-iz} - b_n\frac{(-i)^{n+2}\mathrm{e}^{iz}}{z} - b_n\frac{i^{n+2}\mathrm{e}^{-iz}}{z} + (-i)^{n+1}\mathrm{e}^{iz}R_n^+(z) + i^{n+1}\mathrm{e}^{-iz}R_n^-(z)\right),$$

where

$$b_n = \frac{(2n+1)^2 - 1}{8},$$

and

$$|R_n^+| \leq 2\frac{((2n+1)^2-1)((2n+1)^2-3^2)}{2\cdot 8^2}\frac{\mathrm{e}^{\frac{n^2+n}{|z|^2}}}{|z|^2},$$
$$|R_n^-| \leq 4\frac{((2n+1)^2-1)((2n+1)^2-3^2)}{2\cdot 8^2}\frac{\mathrm{e}^{2\frac{n^2+n}{|z|^2}}}{|z|^2}.$$

From the above we can see that

$$|j_n(z)| \sim \frac{\mathrm{e}^{\operatorname{Im} z}}{2|z|}\left|\left(1 - i\frac{b_n}{z}\right) + (-1)^{n+1}\mathrm{e}^{2iz}\left(1 + i\frac{b_n}{z}\right) + f(z,n)\right|,$$

where $f(z,n) = O\left(\frac{n^4}{|z|^2}\right)$. Alternatively,

$$|j_n(z)| \sim \frac{\mathrm{e}^{\operatorname{Im} z}}{2|z|}\left|\left(1 + (-1)^{n+1}\mathrm{e}^{2iz}\right) + \delta(z,n)\right|, \tag{4.2}$$

where $\delta(z,n) = O\left(\frac{n^2}{|z|} + \frac{n^4}{|z|^2}\right)$.

3. **Regime $n \approx |z|$, $n \to +\infty$.** Let $z = (n+\frac{1}{2})t$. We are interested in the case $\arg t \in (0,\ \pi)$. The asymptotic expansion for this regime can be found in [2, 9.3.35,10.4.59]:

$$j_n\left(\left(n+\frac{1}{2}\right)t\right) \sim \frac{1}{(2n+1)\sqrt{t}}\frac{\mathrm{e}^{(n+\frac{1}{2})\eta(t)}}{(1-t^2)^{\frac{1}{4}}} \tag{4.3}$$

where $\eta(t) = \sqrt{1-t^2} - \log\left(\frac{1+\sqrt{1-t^2}}{t}\right)$.

In the case $t \in \mathbb{R}$, spherical Bessel functions $j_n(t(n+\frac{1}{2}))$ oscillate, however, remain bounded as $n \to +\infty$.

4. **The order is much larger than the argument $n \gg |z|$, $n \to +\infty$.**
In this regime $j_n(z)$ decays super-exponentially, see [2, (9.3.1)]:

$$j_n(z) \sim \sqrt{\frac{\mathrm{e}}{2}}\frac{(\mathrm{e}z)^n}{(2n+1)^{n+1}}. \tag{4.4}$$

Another bound on spherical Bessel functions of complex argument valid for all $n \in \mathbb{N}$ is given by [2, 9.1.62]

$$|j_n(z)| \leq \frac{|z|^n}{(2n+1)!!}\mathrm{e}^{\operatorname{Im} z} = \frac{|z|^n(2n)!}{2^n n!}\mathrm{e}^{\operatorname{Im} z} \tag{4.5}$$

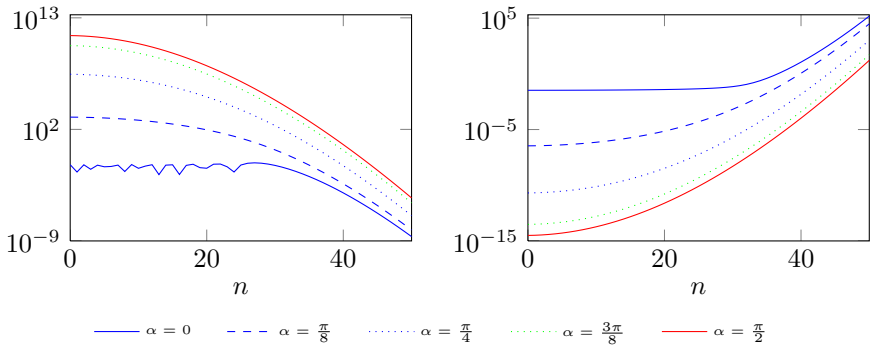


*Figure 4.1: In the left plot $|j_n(re^{i\alpha})|$ for different values of $\alpha$ and fixed $r = 30$ is depicted. The magnitude $|h_n(re^{i\alpha})|$ for the same values of $\alpha$ and $r$ is plotted on the right.*

Using Stirling's approximation

$$1 \leq \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} \leq \frac{\mathrm{e}}{\sqrt{2\pi}},$$

this can be rewritten as

$$|j_n(z)| \leq \mathrm{e}^{\operatorname{Im} z} \frac{\mathrm{e}}{2|z|\sqrt{\pi}} \left(\frac{|z|\mathrm{e}}{2(n+1)}\right)^{n+1}. \tag{4.6}$$

The behavior of spherical Hankel functions $h_n(z)$ is in some sense opposite to that of spherical Bessel functions: they decay exponentially with $\operatorname{Re} z$ in the regime $n < |z|$, see Figure 4.1.

1. **Fixed order $n$, small argument $z$.** The behavior of the function $h_n(z)$, $z \to 0$, is given by [2, 9.1.9]:

$$h_n(z) \to -i\frac{\Gamma(n+\frac{1}{2})}{\sqrt{\pi}} \frac{2^n}{z^{n+1}} \tag{4.7}$$

   as $z \to 0$.

2. **Order $n < |z|$.**
   According to [1, 10.17.13,10.17.14,10.17.15],

$$|h_n(z)| = \frac{\mathrm{e}^{-\operatorname{Im} z}}{|z|} \left|1 + i\frac{b_n}{z} + R_1(z)\right|,$$

   where $b_n = \frac{(2n+1)^2-1}{8}$ and, for $\arg z \in [0,\ \pi]$ (which is the case of interest for us),

$$|R_1(z)| \leq \frac{\left((2n+1)^2-1\right)\left((2n+1)^2-3^2\right)}{2 \cdot 8^2|z|^2} \mathrm{e}^{\frac{n^2+n}{|z|^2}}.$$

   This implies that in the regime $n < |z|$

$$|h_n(z)| \sim (1 + \gamma(|z|, n)) \frac{\mathrm{e}^{-\operatorname{Im} z}}{|z|}, \tag{4.8}$$

   where $\gamma(|z|, n) = O\left(\frac{n^2}{|z|} + \frac{n^4}{|z|^2}\right)$.

3. **Regime $n \approx |z|$, $n \to +\infty$.** Let $z = (n + \frac{1}{2})t$. We are interested in the case $\arg t \in (0,\ \pi)$.

   The behavior of the spherical Hankel function $h_n(z)$ is defined by the asymptotic expansion given in [2, 9.3.37, 10.4.59]:

$$h_n\left(\left(n+\frac{1}{2}\right)t\right) \sim -\frac{2i}{(2n+1)\sqrt{t}(1-t^2)^{\frac{1}{4}}}\mathrm{e}^{-(n+\frac{1}{2})\eta(t)}, \tag{4.9}$$

   where $\eta(t) = \sqrt{1-t^2} - \log\left(\frac{1+\sqrt{1-t^2}}{t}\right)$.

   If $t$ is purely real, the function oscillates but remains bounded with $n \to +\infty$.

4. **The order is much larger than the argument $n \gg |z|$, $n \to +\infty$.** In this regime $h_n(z) = j_n(z) + iy_n(z)$ experiences superexponential growth, see [2, 9.3.1] and (4.4):

$$|h_n(z)| \sim |y_n(z)| \sim \frac{1}{|z|}\sqrt{\frac{2}{\mathrm{e}}}\left(\frac{2n+1}{\mathrm{e}|z|}\right)^n. \tag{4.10}$$

Additionally, magnitudes of spherical Hankel functions are strictly monotonically increasing in their order, see [1, 10.37.1]:

$$|h_n(z)| < |h_m(z)|,\ m > n, \tag{4.11}$$

when $\operatorname{Re} z \geq 0$. The proof of this result can be found in [40].

## 4.2 Truncation of the Gegenbauer's Series

In this section we study the dependence of the truncation parameter $N$ in (3.14) on the complex wavenumber $s \in \mathbb{C}$, $\operatorname{Re} s > 0$.

Let $x, y \in \mathbb{R}^3$, $\|x\|$ and $\|y\|$ be fixed, and let also $\|x\| > \|y\|$. We are looking for $N$ s.t.

$$E_N^{\hat{x}\cdot\hat{y}} = \left|h_0(is\|x-y\|) - \sum_{\ell=0}^{N-1}(2\ell+1)h_\ell(is\|x\|)j_\ell(is\|y\|)P_\ell\left(\hat{x}\cdot\hat{y}\right)\right| < \epsilon, \tag{4.12}$$

for a fixed $\epsilon > 0$. Crucially, to truncate the Gegenbauer's series we use the criterion based on the absolute error rather than the relative one, which resembles similarities with [27] and [43].

**Remark 4.2.1.** *In particular in [43] instead of the relative error*

$$\mathcal{E}_{rel}^{\hat{x}\cdot\hat{y}}(N) = \frac{\left|h_0(is\|x-y\|) - \sum\limits_{\ell=0}^{N-1}(2\ell+1)h_\ell(is\|x\|)j_\ell(is\|y\|)P_\ell\left(\hat{x}\cdot\hat{y}\right)\right|}{|h_0(is\|x-y\|)|}$$

*the authors considered the weighted relative error, namely $\mathcal{E}_w^{\hat{x}\cdot\hat{y}}(N) = \mathcal{E}_{rel}^{\hat{x}\cdot\hat{y}}(N)\mathrm{e}^{-\operatorname{Re} s\|x-y\|}$, to account for rapid decay of the Helmholtz kernel with $\operatorname{Re} s$. This definition was adapted to take into consideration the geometry of a block-cluster tree. Namely, for a pair of admissible clusters whose bounding boxes are located at the distance $d$, the weighted error is defined as $\mathcal{E}_w^{\hat{x}\cdot\hat{y}}(N) = \mathcal{E}_{rel}^{\hat{x}\cdot\hat{y}}(N)\mathrm{e}^{-d\operatorname{Re} s}$, see Figure 4.2. In this work we do not use the definition in this form, and instead of the weighted relative error consider the absolute error. However, all the considerations that follow can be extended to include the definition of the relative error as per [43].*

Let $t = \hat{x}\cdot\hat{y}$, $t \in [-1,\ 1]$. Then the addition theorem for the spherical Bessel functions (3.10) is the Legendre polynomial expansion of the function

$$f(t) = h_0\left(is\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}\right) = -\frac{\mathrm{e}^{-s\left(\|x\|^2+\|y\|^2-2\|x\|\|y\|t\right)^{\frac{1}{2}}}}{s\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}}$$

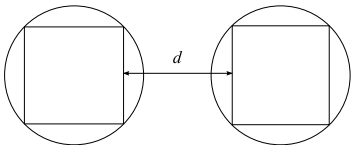


*Figure 4.2: Bounding boxes and their circumscribed spheres.*

in $t$. For $\operatorname{Re} t < t_{max} = \frac{\|x\|^2+\|y\|^2}{2\|x\|\|y\|}$ the function $f(t)$ is analytic, hence Theorem 3.1.1 can be applied. The parameter $\rho$ for the corresponding Bernstein ellipse is defined as

$$\rho = t_{max} + \sqrt{t_{max}^2 - 1} = \frac{\|x\|}{\|y\|}.$$

This rate of convergence coincides with the one deduced in [75, 73].

**Remark 4.2.2.** *The convergence rate of the Gegenbauer's series can be also deduced by analyzing the asymptotic behaviour of spherical Bessel and Hankel functions in the regime $n \gg |s|\|x\|$, see formulas (4.4) and (4.10). It is well known that in the no-decay case ($\operatorname{Re} s = 0$), for large $|s|$ the Gegenbauer's series starts converging earlier, namely when $j_n(is\|y\|)$ starts decaying superexponentially and $h_n(is\|x\|)$ remains bounded [32, 33], see (4.4). The length of the Gegenbauer's series (4.12) can be estimated by a semi-empirical formula, see [30],*

$$N = |s|\|y\| + C\log(\pi + |s|\|y\|),$$

*where $C$ is a constant that depends on the accuracy. For large $|s|$ the value of $N$ does not depend on $\|x\|$.*

Let us first summarize the results of this section. First, it is not difficult to see that under the condition

$$\left|\frac{\operatorname{Im} s}{\operatorname{Re} s}\right| < C, \qquad \operatorname{Re} s > \sigma > 0, \tag{4.13}$$

for fixed $C, \sigma > 0$, the length of the fast multipole expansion can be bounded by a constant independent of $\operatorname{Im} s$ (that depends on $C$, $\sigma$, $\epsilon$, $\rho$ though). This behavior is similar to that of $\mathcal{H}$-matrices, see Section 2.4. The result can be seen by noticing that there exists $r \in \mathbb{R}$, $r > \sigma$, s.t. for all $s \in \mathbb{C}$ with $|s| > r$:

$$|h_0(is\|x-y\|)| = \left|\frac{\mathrm{e}^{-\operatorname{Re} s\|x-y\|}}{|s|\|x-y\|}\right| < \epsilon. \tag{4.14}$$

Hence the length of the expansion (4.12) is bounded by the largest of the lengths of the expansions over all $s \in \mathbb{C}$ satisfying (4.13) and $|s| < r$. This justifies the use of the empirical formulas for the length of the expansion derived in [43]: it indeed can be bounded by a constant when decay is significantly large. Another implication of this is the complexity of the fast multipole approximation to the Galerkin discretization of the Helmholtz single layer boundary operator (under the assumption $\epsilon = \text{const}$): for multilevel fast multipole methods based on the expansion (4.12), for $s$ satisfying (4.13) and $M = O(|s|^2)$, it scales linearly, namely $O(M)$, even if in the high-frequency regime the method is of $O(M^{\frac{3}{2}}\log M)$ complexity (see also Appendix A).

**Remark 4.2.3.** *If $\epsilon \neq \text{const}$, but rather varies as $\epsilon = \epsilon_0|s|^{-\lambda}$, with $\epsilon_0$ being a constant accuracy parameter and $\lambda > 0$ (this may be required by the stability/convergence of the Galerkin method, see [83]), the length of the expansion still can be bounded by a constant. This is because $|s|^\lambda\,|h_0(is\|x-y\|)| \to 0$ as $|s| \to +\infty$ and (4.13) holds, c.f. (4.14).*

For $s = |s|\mathrm{e}^{i\alpha}$ with $|\alpha|$ close to $\frac{\pi}{2}$ this constant bound is far from optimal. This can be seen in Figure 4.3 ($\|y\| = 2$, $\|x\| = 4$, $\hat{x} \cdot \hat{y} = 1$): the length of the expansion needed to achieve a given accuracy increases with increasing $|s|$ on the whole interval under consideration. Notably, for $|\alpha| \leq \frac{\pi}{4}$, the length of the expansion does not seem to increase with $|s|$. Our goal is to provide some theoretical justification for this phenomenon. Here we present more refined bounds on the value of $N$ in (4.12), as well as motivate the error analysis in the subsequent sections.

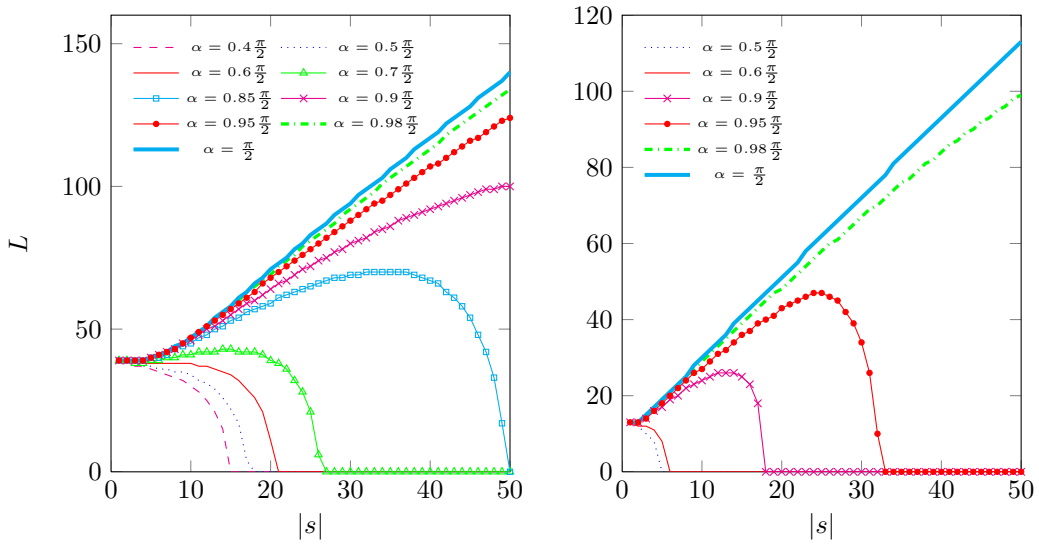


*Figure 4.3: Dependence of the length of the truncated expansion for accuracies $\epsilon = 10^{-12}$ (the left plot) and $\epsilon = 10^{-4}$ (the right plot) on $|s|$, for $s = |s|\mathrm{e}^{i\alpha}$ and different values of $\alpha$.*

The following result is due to [88].

**Theorem 4.2.4.** *Let the function $f$ be analytic inside and on a Bernstein ellipse $E_\rho$, $\rho > 1$. Let $\{a_n\}$ be the coefficients of the Legendre series expansion of $f$. Then the following bound holds true for all $n \geq 0$:*

$$|a_n| \leq (2n+1)\rho^{-n-1}\mathcal{M}\pi^{-1}l(E_\rho)(1-\rho^{-2})^{-1},$$

*where $\mathcal{M} = \max_{z \in E_\rho} |f(z)|$ and $l(E_\rho)$ is the circumference of the ellipse $E_\rho$.*

The next lemma bounds the values of the function $f(t)$ on the Bernstein ellipse.

**Lemma 4.2.5.** *Given $s = |s|e^{i\alpha}$, $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, the function*

$$f(t) = h_0\left(is\left(x^2 + y^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}\right),$$

*inside the Bernstein ellipse $E_\xi$, $\xi < \rho = \frac{\|x\|}{\|y\|}$, is bounded by*

$$|f(t)| \leq \max\left(1, \mathrm{e}^{\|y\||s|(|\sin\alpha| - \cos\alpha)\lambda(\rho)}\right)\left(|s|\|y\|\sqrt{\rho}\sqrt{\rho - \xi + \frac{1}{\rho} - \frac{1}{\xi}}\right)^{-1},$$

*where*

$$\lambda(\rho) = \frac{1}{2}\left(\rho - \frac{1}{\rho}\right). \tag{4.15}$$

*Proof.* Let us bound the numerator and the denominator of

$$f(t) = -\frac{\exp\left(-s\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}\right)}{s\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}}$$

on the boundary of the Bernstein ellipse

$$E_\xi = \left\{z : z = \frac{\xi \mathrm{e}^{i\phi} + \xi^{-1}\mathrm{e}^{-i\phi}}{2},\ \phi \in [0,\ 2\pi)\right\}, \qquad \xi < \rho = \frac{\|x\|}{\|y\|}.$$

The absolute value of the denominator

$$\begin{aligned} d(\phi) :&= \left|\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\frac{(\xi e^{i\phi} + \xi^{-1}e^{-i\phi})}{2}\right|^{\frac{1}{2}} \\ &= \sqrt{\|x\|\|y\|}\left|\rho + \rho^{-1} - \xi e^{i\phi} - \xi^{-1}e^{-i\phi}\right|^{\frac{1}{2}}. \end{aligned}$$

The minimum of this expression is achieved when $\phi = 0$:

$$d(\phi) \geq \sqrt{\rho}\|y\|\sqrt{\rho - \xi + \frac{1}{\rho} - \frac{1}{\xi}}.$$

Let, for a given $\xi < \rho$,

$$G(t) := \exp\left(-s\left(\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|t\right)^{\frac{1}{2}}\right),\ t \in E_\xi.$$

From the maximum principle it follows that for all $t$ in the interior of the Bernstein ellipse $E_\rho$ (and hence $E_\xi,\ 1 < \xi < \rho$),

$$|G(t)| \leq \max_{t' \in E_\rho} |G(t')|.$$

Hence, we are looking for $\max\limits_{t' \in E_\rho} |G(t')|$. Given $s = |s|\mathrm{e}^{i\alpha}$, for all $t \in E_\rho$,

$$\begin{aligned} |G(t)| &\leq \left|\exp\left(-s(2\|x\|\|y\|)^{1/2}\sqrt{\frac{\rho + \rho^{-1}}{2} - \frac{\rho e^{i\phi} + \rho^{-1}e^{-i\phi}}{2}}\right)\right| \\ &= \exp\left(-(2\|x\|\|y\|)^{1/2}|s|\sqrt{|z|}\cos\left(\alpha + \frac{\beta}{2}\right)\right), \end{aligned} \tag{4.16}$$

where

$$z = \frac{\rho + \rho^{-1}}{2} - \frac{\rho e^{i\phi} + \rho^{-1}e^{-i\phi}}{2},\ \beta = \arg z. \tag{4.17}$$

We find $|z| = Z(\beta)$ using the geometric meaning of (4.17).

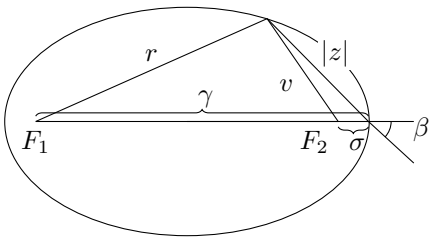


*Figure 4.4: Bernstein ellipse*

The foci of the Bernstein ellipse lie in the points

$$F_1 = (-1,\ 0),\ F_2 = (1,\ 0).$$

From the properties of the Bernstein ellipse, using Figure 4.4,

$$\gamma = \frac{\rho + \rho^{-1}}{2} + 1,\ \sigma = \frac{\rho + \rho^{-1}}{2} - 1, \tag{4.18}$$

$$r + v = \gamma + \sigma = \rho + \rho^{-1}, \tag{4.19}$$

$$r^2 = |z|^2 + \gamma^2 - 2|z|\gamma\cos\beta, \tag{4.20}$$

$$v^2 = |z|^2 + \sigma^2 - 2|z|\sigma\cos\beta, \tag{4.21}$$

where

$$\beta \in \left(-\frac{\pi}{2},\ \frac{\pi}{2}\right).$$

From equations (4.19) and (4.21), we obtain the following expression for $r,\ |z|$:

$$\begin{aligned} |z|^2 + \sigma^2 - 2|z|\sigma\cos\beta &= (\gamma + \sigma - r)^2 \\ &\overset{(4.20)}{=} \gamma^2 + \sigma^2 + (|z|^2 + \gamma^2 - 2|z|\gamma\cos\beta) + 2\gamma\sigma - 2(\gamma+\sigma)r. \end{aligned}$$

Then $r$ can be written as a function of $|z|$:

$$r = \gamma - |z|\frac{\gamma - \sigma}{\gamma + \sigma}\cos\beta \overset{(4.18)}{=} \gamma - \frac{2|z|\cos\beta}{\gamma + \sigma}.$$

Hence,

$$\begin{aligned} r^2 &= \left(\gamma - \frac{2|z|\cos\beta}{\gamma + \sigma}\right)^2 \\ &\overset{(4.20)}{=} |z|^2 + \gamma^2 - 2|z|\gamma\cos\beta. \end{aligned}$$

From this we obtain the following expression for $|z|$:

$$|z|^2\left(1 - \frac{4\cos^2\beta}{(\gamma+\sigma)^2}\right) = 2|z|\gamma\cos\beta\left(1 - \frac{2}{\gamma+\sigma}\right).$$

From this follows that $|z| = 0$ or

$$|z| = 2\gamma\cos\beta\left(1 - \frac{2}{\sigma+\gamma}\right)\left(1 - \frac{4\cos^2\beta}{(\sigma+\gamma)^2}\right)^{-1},\ \beta \in \left(-\frac{\pi}{2},\ \frac{\pi}{2}\right). \tag{4.22}$$

Note that for $|\beta| = \frac{\pi}{2}$ the above expression gives $|z| = 0$.

Hence, inserting (4.22) into (4.16), we obtain:

$$\max_{t\in E_\rho} G(t) = \max_{\beta\in\left[-\frac{\pi}{2},\ \frac{\pi}{2}\right]} g(\beta),$$

where

$$\begin{aligned} g(\beta) = \exp\Bigg(&-2\left(\|x\|\|y\|\gamma\left(1 - \frac{2}{\sigma+\gamma}\right)\right)^{1/2}|s| \\ &\times\sqrt{\cos\beta}\left(1 - \frac{4\cos^2\beta}{(\sigma+\gamma)^2}\right)^{-1/2}\cos\left(\alpha + \frac{\beta}{2}\right)\Bigg). \end{aligned} \tag{4.23}$$

This expression has to be maximized in $\beta$. Let $\mu := \frac{2}{\sigma+\gamma}$ and

$$R(\beta) := -\sqrt{\cos\beta}\cos\left(\alpha + \frac{\beta}{2}\right)\left(1 - \mu^2\cos^2\beta\right)^{-\frac{1}{2}}.$$

We consider two cases corresponding to $\alpha \geq 0$. The bounds for $\alpha \leq 0$ can be found from similar considerations.

1. $\alpha \in \left[0, \frac{\pi}{4}\right]$. In this case, for all $\beta \in \left(-\frac{\pi}{2},\ \frac{\pi}{2}\right)$,

$$-\cos\left(\alpha + \frac{\beta}{2}\right) \geq 0$$

and hence

$$R(\beta) \leq 0.$$

2. $\alpha \in \left(\frac{\pi}{4}, \frac{\pi}{2}\right]$. The maximum of $R(\beta)$ is achieved in some $\beta = \beta_* \in \left[-\frac{\pi}{2},\ \frac{\pi}{2}\right]$, s.t.

$$\cos\left(\alpha + \frac{\beta_*}{2}\right) \leq 0,$$

which shows that

$$\beta_* \in \left[\pi - 2\alpha,\ \frac{\pi}{2}\right].$$

Setting $p := \cos\beta$, with $\beta$ lying inside the above interval:

$$\begin{aligned} R(\beta) &= -\sqrt{p}(1-\mu^2p^2)^{-\frac{1}{2}}\left(\cos\alpha\sqrt{\frac{1+p}{2}} - \sin\alpha\sqrt{\frac{1-p}{2}}\right) \\ &\leq \sqrt{p}(1-\mu^2p^2)^{-\frac{1}{2}}\left(\sin\alpha\sqrt{\frac{1-p}{2}} - \cos\alpha\sqrt{\frac{1+p}{2}}\right) \\ &\leq \sqrt{\frac{p(1-p)}{2}}\left(1-\mu^2p^2\right)^{-\frac{1}{2}}(\sin\alpha - \cos\alpha). \end{aligned}$$

Next we find the maximum of the function $f(p) = \sqrt{\frac{p(1-p)}{2(1-\mu^2p^2)}}$ on $[0,\ 1]$. It is achieved at

$$p_* = \frac{1-\sqrt{1-\mu^2}}{\mu^2}$$

and equals

$$f(p_*) = \frac{\sqrt{1-\sqrt{1-\mu^2}}}{2\mu}.$$

Hence,

$$R(\beta) \leq \frac{\sqrt{1-\sqrt{1-\mu^2}}}{2\mu}(\sin\alpha - \cos\alpha).$$

The bound in the statement of the lemma can be deduced noting that

$$\rho^{-1} = \frac{1-\sqrt{1-\mu^2}}{\mu}, \tag{4.24}$$

$$\rho = \frac{1+\sqrt{1-\mu^2}}{\mu} \tag{4.25}$$

and $\gamma = \frac{1}{\mu} + 1$. More precisely, the coefficient in the exponent near $|s|\|y\|(\sin\alpha - \cos\alpha)$ in (4.23) is

$$\begin{aligned} \lambda(\rho) &= 2\sqrt{\rho\gamma(1-\mu)}\frac{\sqrt{1-\sqrt{1-\mu^2}}}{2\mu} \\ &= \sqrt{\rho}\sqrt{\gamma(1-\mu)}\frac{\sqrt{\rho^{-1}}}{\sqrt{\mu}}, \end{aligned}$$

where we applied (4.24). Inserting the explicit expression of $\gamma$ in terms of $\mu$, we obtain

$$\lambda(\rho) = \frac{\sqrt{1-\mu^2}}{\mu},$$

which, using (4.24) and (4.25), gives the explicit expression for $\lambda(\rho)$.

□

Theorem 4.2.4 and Lemma 4.2.5 allow us to formulate the following bound.

**Corollary 4.2.6.** *Let $\|x\| > \|y\| > 0$ and $1 < \xi < \rho = \frac{\|x\|}{\|y\|}$. Then for $s = |s|\mathrm{e}^{i\alpha}$ s.t. $\alpha \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ the following bound holds true:*

$$\begin{aligned} |h_n(is\|x\|)j_n(is\|y\|)| \leq \xi^{-n-1}\sqrt{\xi^2+\xi^{-2}}\max\left(1, \mathrm{e}^{\|y\||s|(|\sin\alpha|-\cos\alpha)\lambda(\rho)}\right) \\ \times \left(|s|\|y\|\sqrt{\rho}(1-\xi^{-2})\sqrt{\rho-\xi+\frac{1}{\rho}-\frac{1}{\xi}}\right)^{-1}, \end{aligned} \tag{4.26}$$

*where $\lambda(\rho)$ is given by (4.15).*

*Proof.* To apply Theorem 4.2.4, we need to bound the perimeter of the ellipse (analytically, it is expressed via the complete elliptic integral of the second kind):

$$l(E_\xi) \leq 2\pi\sqrt{\frac{a^2+b^2}{2}},$$

where $a$, $b$ are correspondingly the lengths of the semi-major and semi-minor axes of the ellipse $E_\xi$. This bound is due to Euler; other, more precise expressions can be found in e.g. [10]. Hence

$$l(E_\xi) \leq \pi\sqrt{\xi^2+\frac{1}{\xi^2}}.$$

□

For $s = |s|\mathrm{e}^{i\alpha}$, $|\alpha| \leq \frac{\pi}{4}$, the length of the truncated expansion (4.12) can be bounded by a constant that is independent of $|s|$, $\alpha$ for the range of $|s| > \sigma > 0$ for a fixed $\sigma > 0$. This is not the case for $\frac{\pi}{4} < |\alpha| \leq \frac{\pi}{2}$: Corollary 4.2.6 shows that

$$N = O\left(|s|\|y\|\left(|\sin\alpha| - \cos\alpha\right)\right),$$

which is tight for smaller values of $|s|$ and $\alpha$ close to $\frac{\pi}{2}$, however is pessimistic as $|s| \to +\infty$, as Figure 4.5 shows.

**Remark 4.2.7.** *Our numerical experiments show that the length of the truncated expansion for $\operatorname{Re} s > 0$, moderate values of $\rho$ ($\rho \geq 1.5$) and moderate values of $|s|$ satisfies:*

$$N \ll |s|\|x\|. \tag{4.27}$$

*The reason for this is that in the presence of decay, i.e. when $\operatorname{Re} s > 0$, the length of the truncated expansion is not larger the length of the truncated expansion in the no-decay case (keeping $\operatorname{Im} s$ fixed), see also Figure 4.6 and Remark 4.2.2. For the latter there exist tight formulas showing (4.27), see e.g. [25].*

Recall that the criterion based on (4.12) is often used to choose the length of the multipole expansion, see [33, 27]. Numerical experiments, see Figure 4.5, show that the length of the expansion for larger values of $|s|\|y\|$ can be smaller than for smaller $|s|\|y\|$. In the fast multipole algorithm, $\|y\|$ stands for the diameter of a cluster, and $\|x\|$ for the distance between the centers of the admissible clusters, see also Section 3.2. Assuming that $\rho = \frac{\|x\|}{\|y\|}$ is fixed, we can conclude that for larger clusters one may need the expansion of the smaller length than for smaller ones. It is not obvious if such choice of the length of the expansion leads to the deterioration of the accuracy when doing multipole-to-multipole and local-to-local transforms. This motivates the need for the analysis of the error of the multilevel fast multipole method for a complex wavenumber case.

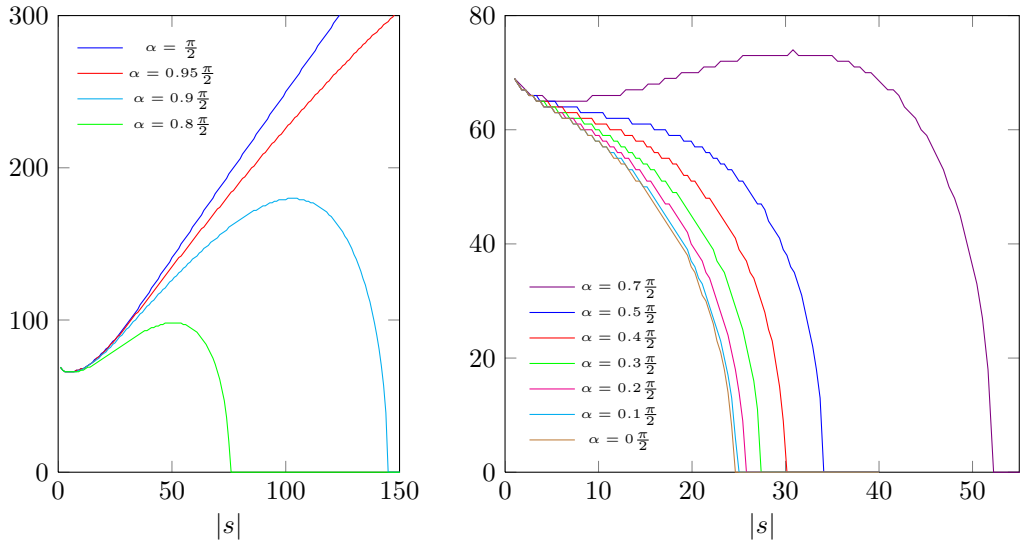


*Figure 4.5: In both plots the dependence of the length of the expansion (4.12) on $|s|$ ($s = |s|e^{i\alpha}$) for $\epsilon = 10^{-12}$ and various $\alpha$ is shown. Here $\|x\| = 3$, $\|y\| = 2$, $\hat{x} \cdot \hat{y} = 1$.*

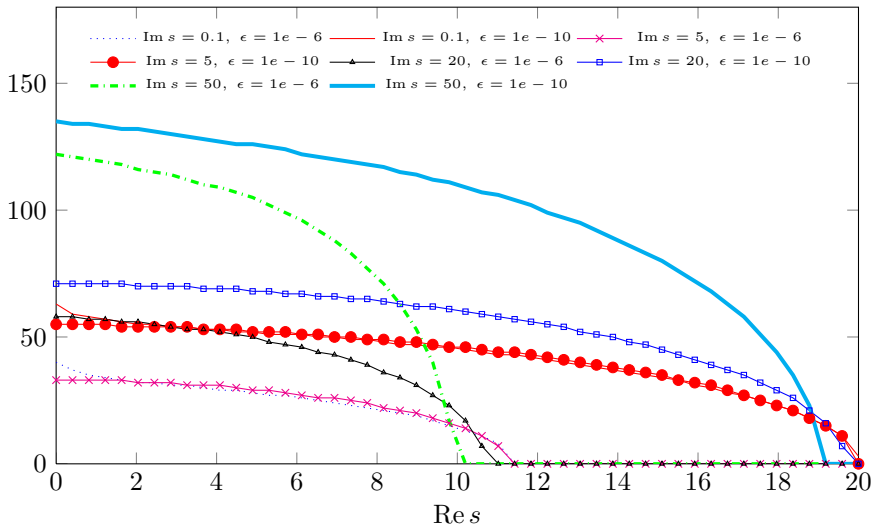


*Figure 4.6: The length of the expansion for $s = s_r + is_i$, as defined in (4.12), with varying $s_r$ and fixed $s_i$; $\|x\| = 3$, $\|y\| = 2$, $\hat{x} \cdot \hat{y} = 1$.*

## 4.3 Error Analysis

### 4.3.1 Error Control

Given a block-cluster tree $\mathcal{T}_{I\times I}$, let $\tau_\alpha$ and $\tau_\beta$ be two admissible clusters (the block cluster $(\tau_\alpha, \tau_\beta)$ belongs to the set of admissible leaves of the block-cluster tree). In this section we consider the error of the approximation of $h_0(is\|x-y\|)$ by $\tilde{h}_0(s,x,y)$ computed in the course of the fast multipole algorithm:

$$E = |h_0(is\|x-y\|) - \tilde{h}_0(s,x,y)|, \ s \in \mathbb{C}, \tag{4.28}$$

where $x \in \tau_\alpha$, $y \in \tau_\beta$.

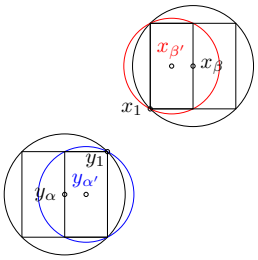


*Figure 4.7: A sample configuration of bounding boxes of the clusters with their circumscribed spheres. By $x_1, y_1$ we denote two closest points of these spheres.*

There exist several empirical formulas [30, 25, 43, 51] that provide tight estimates for the length of expansions in the fast multipole algorithm. In works [33, 27] authors suggest that it can be chosen analyzing the convergence of the Gegenbauer's series (4.12). We adopt this approach.

In [69] the authors analyzed the full error of the fast multipole algorithm in the case $is \in \mathbb{R}$. Their error analysis uses superexponential decay of spherical Bessel functions $j_n(is\|y\|)$ in the regime $n \gg |s|\|y\|$ and the geometric convergence of the quadrature on the unit sphere used for the interpolation of multipole expansions in the course of the fast multipole method. A straightforward application of such an error analysis to the case of the complex wavenumber may result in pessimistic error bounds, since it would not take into account the decay of spherical Hankel functions.

In this section we derive an explicit expression for the error of the multilevel fast multipole method for the case of general $s \in \mathbb{C}$ and comment on the choice of lengths of multipole expansions.

We study the following simple case.

**Assumption 4.3.1.** *Let admissible (in the sense of Definition 2.2.5) clusters $\tau_\alpha$ and $\tau_\beta$ be given. Let the points $x \in \tau_\alpha$, $y \in \tau_\beta$, and let the clusters $\tau_{\alpha'} \in \operatorname{sons}(\tau_\alpha)$ and $\tau_{\beta'} \in \operatorname{sons}(\tau_\beta)$ be s.t. $x \in \tau_{\alpha'}$ and $y \in \tau_{\beta'}$. We assume that $\tau_{\alpha'}$ is a leaf, and so is $\tau_{\beta'}$. By $y_\alpha$, $x_\beta$, $y_{\alpha'}$, $x_{\beta'}$ we denote the centers of the bounding boxes of the clusters $\tau_\alpha$, $\tau_\beta$, $\tau_{\alpha'}$, $\tau_{\beta'}$.*

**Remark 4.3.2.** *In the FMM theory instead of clusters one considers their bounding boxes. It is assumed that the error is the largest in either closest points of these boxes, see [43], or in the points of boxes that are most distant from their centers, see [26].*

*In this work we associate the clusters with circumscribed spheres of their bounding boxes. Then the largest error should be achieved in points $x, y$ lying on these spheres so that $\|x - x_\beta - y + y_\alpha\| \to \max$, due to a slower convergence of the Gegenbauer's series in this case, c.f. Remark 4.2.2 and Section 3.2. In [25] it was shown that the Gegenbauer's series truncation error is the largest in two most distant or two closest points of these spheres (while the local maximum in other points is typically smaller than these values) when $|s| \to +\infty$, $s = i\kappa$, $\kappa \in \mathbb{R}$. In the case $\operatorname{Re} s > 0$, as we show further with the help of extensive numerical experiments, considering the two closest points is enough. Such a configuration is shown in Figure 4.7.*

We assume that all spherical harmonic transforms are done exactly, i.e. no matrix approximation is used in the course of the transform, see Section 3.2.2 for more details. Recall that with every level of the cluster tree we associate a set of quadrature points on the unit sphere given by (3.11), i.e.

$$(\hat{s}_k)_{k=1}^{2n_\theta^2}. \tag{4.29}$$

We set $n_\theta = N$ at the level where the children clusters are located and $n_\theta = M$ at the level of the parent clusters. Let us also define

$$N_* = \min(N, M).$$

First, let us assume that $N \neq M$. The fast multipole algorithm proceeds in the following stages.

1. Evaluation of the multipole expansion for the cluster $\tau_{\beta'}$. The function $f(\hat{s}) = \mathrm{e}^{-s(\hat{s},x-x_{\beta'})}$ is sampled on the grid (4.29) of size $N \times 2N$, see Remark 3.1.5.

2. Evaluation of the multipole expansion for the cluster $\tau_\beta$. This is done in two stages. First, the multipole expansion for the cluster $\tau_{\beta'}$ is re-sampled on the grid of size $M \times 2M$ with the help of the spherical harmonic transform operator (and possibly the spherical harmonic expansion of $f(\hat{s})$ is truncated to $\min(N, M) = N_*$ summands, see Section 3.2.2 and Remark 3.2.1). The result of this operation is the vector of values of the function

$$b(\hat{s}) = \sum_{n=0}^{N_*-1} \sum_{m=-n}^{n} \beta_n^m Y_n^m(\hat{s}),$$
$$\beta_n^m = Q_N \left[ \mathrm{e}^{-s(\hat{q},x-x_{\beta'})} Y_n^{m*}(\hat{q}) \right]$$

in the points of the grid (4.29) of size $M \times 2M$. The expression for $\beta_n^m$ is obtained using Lemma 3.1.4. An alternative expression for $b(\hat{s})$ can be obtained using (3.5):

$$b(\hat{s}) = \sum_{n=0}^{N_*-1} \frac{2n+1}{4\pi} Q_N \left[ \mathrm{e}^{-s(\hat{q},x-x_{\beta'})} P_n(\hat{s} \cdot \hat{q}) \right].$$

Next,

$$B(\hat{s}) = \mathrm{e}^{-s(x_{\beta'}-x_\beta,\hat{s})} b(\hat{s})$$

is evaluated at the points of the grid (4.29) of size $M \times 2M$.

3. Multipole-to-local translation. At each point of the grid $\hat{s}_k, k = 1, \ldots, 2M^2$, $B(\hat{s}_k)$ is multiplied by

$$M_{\alpha,\beta}(\hat{s}_k) = \frac{1}{4\pi} \sum_{\ell=0}^{L-1} (2\ell+1)(-i)^\ell h_\ell(is\|y_\alpha - x_\beta\|) P_\ell \left( \frac{y_\alpha - x_\beta}{\|y_\alpha - x_\beta\|} \cdot \hat{s}_k \right), \tag{4.30}$$

where $L \in \mathbb{N}$. The result of this operation is the vector of values of the function

$$F(\hat{s}) = M_{\alpha,\beta}(\hat{s}) B(\hat{s}) \tag{4.31}$$

in the points of the grid (4.29) of size $M \times 2M$.

4. Local-to-local translation. First, at each point of the grid $F(\hat{s})$ is multiplied by $\mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{s})}$ evaluated at this point. The result of this operation is the vector of values of

$$A(\hat{s}) = \mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{s})} F(\hat{s})$$

in the points of the grid (4.29) of size $M \times 2M$. Next, the (adjoint) spherical harmonic transform operator is applied to $A(\hat{s})$, possibly truncating its spherical harmonic expansion and re-sampling it on the grid (3.11) of size $N \times 2N$. The result of this operation is the vector of values of the function

$$a(\hat{s}) = \sum_{n=0}^{N_*-1} \sum_{m=-n}^{n} \alpha_n^m Y_n^m(\hat{s}),$$
$$\alpha_n^m = Q_M \left[ A(\hat{q}) Y_n^{m*}(\hat{q}) \right],$$

in the points of the grid (4.29) of size $N \times 2N$. An explicit expression for the coefficients $\alpha_n^m$ is:

$$\alpha_n^m = Q_M \left[ \mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{q})} M_{\alpha,\beta}(\hat{q}) \mathrm{e}^{-s(x_{\beta'}-x_\beta,\hat{q})} b(\hat{q}) Y_n^{m*}(\hat{q}) \right].$$

Using the addition theorem for Legendre functions (3.5),

$$a(\hat{s}) = \sum_{n=0}^{N_*-1} \frac{2n+1}{4\pi} Q_M \left[ A(\hat{q}) P_n(\hat{q}\cdot\hat{s}) \right].$$

Finally, the result is evaluated at the point $y$, giving the approximation

$$\tilde{h}_0(s,x,y) = Q_N \left[ \mathrm{e}^{-s(y_{\alpha'}-y,\hat{s})} a(\hat{s}) \right].$$

**Remark 4.3.3.** *If $N = M$, there is no need to perform a spherical harmonic transform when doing multipole-to-multipole and local-to-local translations. In this case the approximation error is equivalent to that of a one-level FMM, see Lemma 4.3.4.*

Making use of the linearity of the quadrature rule, we end up with the following approximation to $h_0(is\|x-y\|)$:

$$\begin{aligned}
\tilde{h}_0(s,x,y) &= Q_N^{\hat{q}} \left[ \mathrm{e}^{-s(y_{\alpha'}-y,\hat{q})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_M^{\hat{s}} \left[ \mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{s})} M_{\alpha,\beta}(\hat{s}) \mathrm{e}^{-s(x_{\beta'}-x_\beta,\hat{s})} \right.\right. \\
&\quad \left.\left. \times \sum_{n=0}^{N_*-1} \frac{2n+1}{4\pi} Q_N^{\hat{r}} \left[ \mathrm{e}^{-s(x-x_{\beta'},\hat{r})} P_n(\hat{s}\cdot\hat{r}) \right] P_k(\hat{q}\cdot\hat{s}) \right] \right] \\
&= Q_M^{\hat{s}} \left[ \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{q}} \left[ \mathrm{e}^{s(y-y_{\alpha'},\hat{q})} P_k(\hat{q}\cdot\hat{s}) \right] \right. \\
&\quad \left. \times \mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{s})} M_{\alpha,\beta}(\hat{s}) \mathrm{e}^{-s(x_{\beta'}-x_\beta,\hat{s})} \sum_{n=0}^{N_*-1} \frac{2n+1}{4\pi} Q_N^{\hat{r}} \left[ \mathrm{e}^{-s(x-x_{\beta'},\hat{r})} P_n(\hat{r}\cdot\hat{s}) \right] \right].
\end{aligned}$$

Our goal is to rewrite this expression in a more convenient form.
Let

$$\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s}) := \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{q}} \left[ \mathrm{e}^{s(y-y_{\alpha'},\hat{q})} P_k(\hat{q}\cdot\hat{s}) \right].$$

Then $\tilde{h}_0(s,x,y)$ can be rewritten as:

$$\tilde{h}_0(s,x,y) = Q_M^{\hat{s}} \left[ \mathcal{P}_N(-x_{\beta'}, -x_\beta, -x, \hat{s}) M_{\alpha,\beta}(\hat{s}) \mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s}) \right]. \tag{4.32}$$

From the expression (4.32) one can see that $\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s})$ approximates $\mathrm{e}^{s(y-y_\alpha,\hat{s})}$. Let us show this. First, let $u = y_{\alpha'} - y$. According to (3.8),

$$\begin{aligned}
\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s}) &:= \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{q}} \left[ \mathrm{e}^{-s(u,\hat{q})} P_k(\hat{q}\cdot\hat{s}) \right] \\
&= \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} \sum_{m=0}^{+\infty} (2m+1) i^m j_m(is\|u\|) Q_N^{\hat{q}} \left[ P_m(\hat{q}\cdot\hat{u}) P_k(\hat{q}\cdot\hat{s}) \right].
\end{aligned}$$

Making use of Lemma 3.1.8:

$$\begin{aligned}
\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s}) &= \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} \sum_{m=0}^{2N-N_*} (2m+1) i^m j_m(is\|u\|) \\
&\quad \times Q_N^{\hat{q}} \left[ P_m(\hat{q}\cdot\hat{u}) P_k(\hat{q}\cdot\hat{s}) \right] \\
&\quad + \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} \sum_{m=2N-N_*+1}^{+\infty} (2m+1) i^m j_m(is\|u\|) \\
&\quad \times Q_N^{\hat{q}} \left[ P_m(\hat{q}\cdot\hat{u}) P_k(\hat{q}\cdot\hat{s}) \right] \\
&= \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{m=0}^{\min(2N-N_*,N_*-1)} (2m+1) i^m j_m(is\|u\|) P_m(\hat{s}\cdot\hat{u}) \\
&\quad + \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{s})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} \sum_{m=2N-N_*+1}^{+\infty} (2m+1) i^m j_m(is\|u\|) \\
&\quad \times Q_N^{\hat{q}} \left[ P_m(\hat{q}\cdot\hat{u}) P_k(\hat{q}\cdot\hat{s}) \right].
\end{aligned}$$

Let us introduce

$$r_K(s, x, \hat{r}) = \sum_{n=K}^{+\infty} (2n+1) i^n j_n(is\|x\|) P_n(\hat{x}\cdot\hat{r}). \tag{4.33}$$

Then, using (3.8) and $2N - N_* > N_* - 1$, $\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{s})$ can be represented as a sum of three terms:

$$\begin{aligned}
\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{q}) &= \mathrm{e}^{s(y-y_\alpha,\hat{q})} - \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{q})} r_{N_*}(s, y_{\alpha'}-y, \hat{q}) \\
&\quad + \mathrm{e}^{s(y_{\alpha'}-y_\alpha,\hat{q})} \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{r}} \left[ r_{2N-N_*+1}(s, y_{\alpha'}-y, \hat{r}) P_k(\hat{r}\cdot\hat{q}) \right].
\end{aligned}$$

Hence, $\mathcal{P}_N(y_{\alpha'}, y_\alpha, y, \hat{q})$ approximates $\mathrm{e}^{s(y-y_\alpha,\hat{q})}$ with the error that is a sum of two errors, one coming from the truncation of series (3.8) to a finite number of terms (see (4.33)), and another induced by the imprecise quadrature.

The obtained explicit expressions for $\mathcal{P}_N(\cdot,\cdot,\cdot,\cdot)$ have to be inserted into (4.32). After computations one can see that

$$\tilde{h}_0(s, x, y) = h_0(is\|x-y\|) + \sum_{n=1}^{9} \mathcal{A}_n,$$

where

$$\mathcal{A}_1 = Q_M^{\hat{s}} \left[ \mathrm{e}^{s(y-y_\alpha,\hat{s})} M_{\alpha,\beta}(\hat{s}) \mathrm{e}^{s(x_\beta-x,\hat{s})} \right] - h_0(is\|x-y\|), \tag{4.34}$$

and

$$
\begin{aligned}
\mathcal{A}_2 &= -Q_M^{\hat{s}}\left[\mathrm{e}^{s(y-y_\alpha+x_\beta-x_{\beta'},\hat{s})}M_{\alpha,\beta}(\hat{s})r_{N_*}(s,x-x_{\beta'},\hat{s})\right],\\
\mathcal{A}_3 &= -Q_M^{\hat{s}}\left[\mathrm{e}^{s(x_\beta-x-y_\alpha+y_{\alpha'},\hat{s})}M_{\alpha,\beta}(\hat{s})r_{N_*}(s,y_{\alpha'}-y,\hat{s})\right],\\
\mathcal{A}_4 &= Q_M^{\hat{s}}\left[\mathrm{e}^{s(y-y_\alpha+x_\beta-x_{\beta'},\hat{s})}M_{\alpha,\beta}(\hat{s})\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{r}}\left[r_{2N-N_*+1}(s,x-x_{\beta'},\hat{r})P_k(\hat{r}\cdot\hat{s})\right]\right],\\
\mathcal{A}_5 &= Q_M^{\hat{s}}\left[\mathrm{e}^{s(x_\beta-x-y_\alpha+y_{\alpha'},\hat{s})}M_{\alpha,\beta}(\hat{s})\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{q}}\left[r_{2N-N_*+1}(s,y_{\alpha'}-y,\hat{q})P_k(\hat{q}\cdot\hat{s})\right]\right],\\
\mathcal{A}_6 &= Q_M^{\hat{s}}\left[\mathrm{e}^{s(y_{\alpha'}-y_\alpha+x_\beta-x_{\beta'},\hat{s})}r_{N_*}(s,y_{\alpha'}-y,\hat{s})M_{\alpha,\beta}(\hat{s})r_{N_*}(s,x-x_{\beta'},\hat{s})\right],\\
\mathcal{A}_7 &= -Q_M^{\hat{s}}\left[\mathrm{e}^{s(y_{\alpha'}-y_\alpha+x_\beta-x_{\beta'},\hat{s})}r_{N_*}(s,x-x_{\beta'},\hat{s})M_{\alpha,\beta}(\hat{s})\right.\\
&\quad\left.\times\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{q}}\left[r_{2N-N_*+1}(s,y_{\alpha'}-y,\hat{q})P_k(\hat{q}\cdot\hat{s})\right]\right],\\
\mathcal{A}_8 &= -Q_M^{\hat{s}}\left[\mathrm{e}^{s(y_{\alpha'}-y_\alpha+x_\beta-x_{\beta'},\hat{s})}r_{N_*}(s,y_{\alpha'}-y,\hat{s})M_{\alpha,\beta}(\hat{s})\right.\\
&\quad\left.\times\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{r}}\left[r_{2N-N_*+1}(s,x-x_{\beta'},\hat{r})P_k(\hat{r}\cdot\hat{s})\right]\right],\\
\mathcal{A}_9 &= Q_M^{\hat{s}}\left[\mathrm{e}^{s(y_{\alpha'}-y_\alpha+x_\beta-x_{\beta'},\hat{s})}M_{\alpha,\beta}(\hat{s})\right.\\
&\quad\times\sum_{m=0}^{N_*-1}\frac{2m+1}{4\pi}Q_N^{\hat{r}}\left[r_{2N-N_*+1}(s,x-x_{\beta'},\hat{r})P_m(\hat{r}\cdot\hat{s})\right]\\
&\quad\left.\times\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{q}}\left[r_{2N-N_*+1}(s,y_{\alpha'}-y,\hat{q})P_k(\hat{q}\cdot\hat{s})\right]\right].
\end{aligned}
\tag{4.35}
$$

Notice that $\mathcal{A}_2$ has the same structure as $\mathcal{A}_3$, $\mathcal{A}_4$ as $\mathcal{A}_5$ and $\mathcal{A}_7$ as $\mathcal{A}_8$. More precisely, for a fixed $M_{\alpha,\beta}(\hat{s})$, $\mathcal{A}_2=\mathcal{A}_2(x,y,x_\beta,x_{\beta'},y_\alpha,y_{\alpha'},M_{\alpha,\beta})$ and $\mathcal{A}_3=\mathcal{A}_2(-x,-y,-y_\alpha,-y_{\alpha'},-x_\beta,-x_{\beta'},M_{\alpha\beta})$.

Our goal is to show how $L$ in (4.30), $N$ and $M$ have to be chosen to control each of the terms $\mathcal{A}_k,\ k=1,\dots 9$. We will use the trivial bound, see also Lemma 3.1.4:

$$
|Q_M\left[f(\hat{q})\right]| = \left|\sum_{k=1}^{2M^2} w_k f(\hat{q}_k)\right| \le \sup_{\hat{q}\in\mathbb{S}^2}|f(\hat{q})|\sum_{k=1}^{2M^2} w_k = 4\pi\sup_{\hat{q}\in\mathbb{S}^2}|f(\hat{q})|. \tag{4.36}
$$

The sum of the quadrature weights $w_k=4\pi$, since $\sum\limits_{k=1}^{2M^2} w_k$ equals the quadrature rule of Lemma 3.1.4 applied to approximate $\int\limits_{\mathbb{S}^2} d\hat{s}=4\pi$, which in turn is computed by this quadrature rule exactly (as an integral over the product of two spherical harmonics $Y_0^0(\hat{s})\equiv 1$), see Lemma 3.1.4.

Recall that

$$
|P_n(t)|\le 1, \qquad t\in[-1,\ 1]\,. \tag{4.37}
$$

We will make use of (3.8):

$$
e^{iz\cos\theta} = \sum_{n=0}^{\infty}(2n+1)i^n j_n(z)P_n(\cos\theta). \tag{4.38}
$$

For the sake of brevity, from now on

$$
c_{\alpha\beta} := y_\alpha - x_\beta. \tag{4.39}
$$

First, let us assume that $M \geq L$, where $L$ is as in (4.30). Due to the monotonicity result (4.11) and the bound (4.37),

$$|M_{\alpha,\beta}(\hat{s})| = \frac{1}{4\pi}\left|\sum_{\ell=0}^{L-1}(2\ell+1)(-i)^{\ell}h_{\ell}(is\|c_{\alpha\beta}\|)P_{\ell}\left(\hat{c}_{\alpha\beta}\cdot\hat{s}\right)\right| \leq \frac{1}{4\pi}\sum_{\ell=0}^{L-1}(2\ell+1)|h_{\ell}(is\|c_{\alpha\beta}\|)|$$
$$\leq \frac{1}{4\pi}|h_{L-1}(is\|c_{\alpha\beta}\|)|\sum_{\ell=0}^{L-1}(2\ell+1) \leq \frac{L^2}{4\pi}|h_{L-1}(is\|c_{\alpha\beta}\|)|. \tag{4.40}$$

The following lemmas show how the errors $\mathcal{A}_j,\ j=1,\ldots,9$, can be controlled. In particular, the error $\mathcal{A}_1$, see (4.34), does not depend on $N$, and hence is not related to multipole-to-multipole (local-to-local) translations. More precisely, it equals the error of the one-level fast multipole method. Similar statements (though formulated slightly differently) have been proved in [73, 69]. For simplicity, we denote

$$v := y_{\alpha} - y - x_{\beta} + x. \tag{4.41}$$

**Lemma 4.3.4.** *Let $s \in \mathbb{C}$ be s.t. $\operatorname{Re} s \geq 0$, $x, y, x_{\beta}, y_{\alpha} \in \mathbb{R}^3$ satisfy Assumption 4.3.1, $v$ be given by (4.41) and $c_{\alpha\beta}$ by (4.39). Let $L$ be as in (4.30) and $M \geq L$. Let us denote*

$$E_{tr}(L, v, c_{\alpha\beta}, s) = \left|\sum_{m=L}^{\infty}(2m+1)j_m(is\|v\|)h_m(is\|c_{\alpha\beta}\|)P_m\left(\hat{v}\cdot\hat{c}_{\alpha\beta}\right)\right|, \tag{4.42}$$

$$E_I(M, L, v, c_{\alpha\beta}, s) = \left|\sum_{k=2M-L+1}^{+\infty}(2k+1)i^k j_k(is\|v\|)Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right)P_k\left(\hat{s}\cdot\hat{v}\right)\right]\right|. \tag{4.43}$$

*Then $\mathcal{A}_1 = \mathcal{A}_1(M, L, x, x_{\beta}, y, y_{\alpha}, s)$ defined by (4.34) satisfies*

$$|\mathcal{A}_1(M, L, x, x_{\beta}, y, y_{\alpha}, s)| \leq E_{tr}(L, v, c_{\alpha\beta}, s) + E_I(M, L, v, c_{\alpha\beta}, s).$$

*Proof.* Let

$$H := Q_M^{\hat{s}}\left[\mathrm{e}^{s(y-y_{\alpha},\hat{s})}M_{\alpha,\beta}(\hat{s})\mathrm{e}^{s(x_{\beta}-x,\hat{s})}\right].$$

Then $\mathcal{A}_1$ defined by (4.34) equals $\mathcal{A}_1 = H - h_0(is\|x-y\|)$. Let us separately consider $H$:

$$H = Q_M^{\hat{s}}\left[\mathrm{e}^{-s(v,\hat{s})}M_{\alpha,\beta}(\hat{s})\right]$$
$$\overset{(4.38,4.30)}{=} \frac{1}{4\pi}\sum_{k=0}^{\infty}(2k+1)i^k j_k(is\|v\|)\sum_{\ell=0}^{L-1}(2\ell+1)(-i)^{\ell}h_{\ell}(is\|c_{\alpha\beta}\|)Q_M^{\hat{s}}\left[P_{\ell}\left(\hat{c}_{\alpha\beta}\cdot\hat{s}\right)P_k(\hat{v}\cdot\hat{s})\right].$$

We split

$$H = \frac{1}{4\pi}\sum_{k=0}^{2M-L}(2k+1)i^k j_k(is\|v\|)\sum_{\ell=0}^{L-1}(2\ell+1)(-i)^{\ell}h_{\ell}(is\|c_{\alpha\beta}\|)Q_M^{\hat{s}}\left[P_{\ell}\left(\hat{c}_{\alpha\beta}\cdot\hat{s}\right)P_k(\hat{v}\cdot\hat{s})\right]$$
$$+ \frac{1}{4\pi}\sum_{k=2M-L+1}^{+\infty}(2k+1)i^k j_k(is\|v\|)\sum_{\ell=0}^{L-1}(2\ell+1)(-i)^{\ell}h_{\ell}(is\|c_{\alpha\beta}\|)Q_M^{\hat{s}}\left[P_{\ell}\left(\hat{c}_{\alpha\beta}\cdot\hat{s}\right)P_k(\hat{v}\cdot\hat{s})\right].$$

Applying Lemma 3.1.8 to the above and using $M \geq L$, we obtain:

$$\begin{aligned} H &= \sum_{k=0}^{L-1}(2k+1)j_k(is\|v\|)h_k(is\|c_{\alpha\beta}\|)P_k\left(\hat{c}_{\alpha\beta}\cdot\hat{v}\right) \\ &+ \sum_{k=2M-L+1}^{+\infty}(2k+1)i^k j_k(is\|v\|)Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right)P_k\left(\hat{s}\cdot\hat{v}\right)\right] \\ &\overset{(3.10)}{=} h_0(is\|x-y\|) - \sum_{k=L}^{+\infty}(2k+1)j_k(is\|v\|)h_k(is\|c_{\alpha\beta}\|)P_k\left(\hat{c}_{\alpha\beta}\cdot\hat{v}\right) \\ &+ \sum_{k=2M-L+1}^{+\infty}(2k+1)i^k j_k(is\|v\|)Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right)P_k\left(\hat{s}\cdot\hat{v}\right)\right]. \end{aligned}$$

The definition of $\mathcal{A}_1 = H - h_0(is\|x-y\|)$ combined with (4.42) and (4.43) gives the desired bound. □

The following is a trivial corollary of the above lemma.

**Corollary 4.3.5.** *Let $s \in \mathbb{C}$ be s.t. $\operatorname{Re} s \geq 0$, $x, y, x_\beta, y_\alpha \in \mathbb{R}^3$ satisfy Assumption 4.3.1, $v$ be given by (4.41) and $c_{\alpha\beta}$ by (4.39). Let $\epsilon > 0$, $L$ as in (4.30), $M \geq L$ and*

$$E_{tr}(L, v, c_{\alpha\beta}, s) + E_I(M, L, v, c_{\alpha\beta}, s) < \epsilon,$$

*where $E_{tr}$ and $E_I$ are defined by (4.42) and (4.43). Then $\mathcal{A}_1 = \mathcal{A}_1(M, L, x, x_\beta, y, y_\alpha, s)$ defined by (4.34) satisfies*

$$|\mathcal{A}_1(M, L, x, x_\beta, y, y_\alpha, s)| < \epsilon.$$

*Proof.* The corollary trivially follows from Lemma 4.3.4. Such $M$ and $L$ exist, since the Gegenbauer's series converges geometrically, see Section 4.2, $\left|Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right)P_k\left(\hat{s}\cdot\hat{v}\right)\right]\right|$ can be uniformly bounded for all $k$ (see also Lemma 4.3.6) and $\sum\limits_{m=0}^{+\infty}(2m+1)|j_m(z)|$ converges supergeometrically for all $z \in \mathbb{C}$, see Remark 3.1.2. □

The following lemma shows how the term $E_I(M, L, v, c_{\alpha\beta}, s)$ can be bounded.

**Lemma 4.3.6.** *Let $s \in \mathbb{C}$ be s.t. $\operatorname{Re} s \geq 0$, $x, y, x_\beta, y_\alpha \in \mathbb{R}^3$ satisfy Assumption 4.3.1, $v$ be given by (4.41) and $c_{\alpha\beta}$ by (4.39). Let $L$ be as in (4.30) and $M \geq L$. Then the following bounds hold for $E_I$ defined by (4.43):*

$$E_I(M, L, v, c_{\alpha\beta}, s) < L^2\left|h_{L-1}(is\|c_{\alpha\beta}\|)\right|\sum_{m=L}^{\infty}(2m+1)|j_m(is\|v\|)|. \tag{4.44}$$

*and*

$$E_I(M, L, v, c_{\alpha\beta}, s) < L^2\sum_{m=L}^{\infty}(2m+1)|j_m(is\|v\|)h_m(is\|c_{\alpha\beta}\|)|. \tag{4.45}$$

*Proof.* Inequalities (4.36,4.37,4.40) let us bound

$$\begin{aligned} E_I(M, L, v, c_{\alpha\beta}, s) &= \left|\sum_{k=2M-L+1}^{+\infty}(2k+1)i^k j_k(is\|v\|)Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right)P_k\left(\hat{s}\cdot\hat{v}\right)\right]\right| \\ &\leq L^2\sum_{k=L}^{\infty}(2k+1)\left|j_k(is\|v\|)h_{L-1}(is\|c_{\alpha\beta}\|)\right|. \end{aligned}$$

The monotonicity of spherical Hankel functions (4.11) and the above bound imply that

$$E_I(M,L,v,c_{\alpha\beta},s) \le L^2 \sum_{k=L}^{\infty} (2k+1)\,|j_k(is\|v\|)h_{L-1}(is\|c_{\alpha\beta}\|)| \\ \le L^2 \sum_{k=L}^{\infty} (2k+1)|j_k(is\|v\|)h_k(is\|c_{\alpha\beta}\|)|.$$

□

The error terms $\mathcal{A}_j,\ j=2,\dots,9$, defined by (4.35), occur due to the multipole-to-multipole and local-to-local transforms (each of these expressions contains a term that depends on $N$). To show how these errors can be controlled, we will need the following auxiliary quantities. Let us set

$$R_K(d) = \sum_{m=K}^{+\infty} (2m+1)|j_m(d)|.$$

Clearly $|r_K(s,x-x_{\beta'},\hat{s})| < R_K(is\|x-x_{\beta'}\|)$ for arbitrary $\hat{s}\in\mathbb{S}_2$. Moreover, since the series (4.38) converges supergeometrically (Remark 3.1.2), $R_K(d)$ decays rapidly as $K\to+\infty$.

Given $d_\alpha, d_\beta, d_{\alpha'}, d_{\beta'}$ being the diameters of bounding boxes of clusters $\tau_\alpha,\tau_\beta,\tau_{\alpha'},\tau_{\beta'}$, let us introduce auxiliary quantities:

$$\begin{aligned} r_p &= \frac{1}{2}\max(d_\alpha,d_\beta),\\ r_c(x) &= \|x-x_{\beta'}\|,\\ r_c(y) &= \|y-y_{\alpha'}\|,\\ r_d &= \max\left(\|y_\alpha-y_{\alpha'}\|,\ \|x_\beta-x_{\beta'}\|\right). \end{aligned} \tag{4.46}$$

Clearly, $r_c(x)\le\frac{1}{2}d_{\beta'}$ and $r_c(y)\le\frac{1}{2}d_{\alpha'}$.

The following simple lemma demonstrates how the errors $\mathcal{A}_j$, $j=2,\dots,9$, defined by (4.35) can be made small.

**Lemma 4.3.7.** *Let $s\in\mathbb{C}$ be s.t. $\operatorname{Re} s\ge 0$, $x,y,x_\beta,x_{\beta'},y_\alpha,y_{\alpha'}\in\mathbb{R}^3$ satisfy Assumption 4.3.1, $r_p,r_d,r_c(x),r_c(y)$ be defined by (4.46) and $c_{\alpha\beta}$ by (4.39). Given $\epsilon>0$, $L$ as in (4.30), let $N$, $M$ be chosen so that $N_*=\min(N,M)$ satisfies:*

$$E^{(1)} = L^2N_*^2\mathrm{e}^{\operatorname{Re} s(r_p+r_d)}\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\max\left(R_{N_*}(isr_c(x)),R_{N_*}(isr_c(y))\right)\le\epsilon, \tag{4.47}$$

$$E^{(2)} = L^2N_*^4\mathrm{e}^{2\operatorname{Re} sr_d}\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\max\left(R^2_{N_*}(isr_c(x)),R^2_{N_*}(isr_c(y))\right)\le\epsilon. \tag{4.48}$$

*Then the following bound holds for $\mathcal{A}_j=\mathcal{A}_j(L,M,N,x,y,x_\beta,y_\alpha,x_{\beta'},y_{\alpha'},s),\ j=2,\dots,9$, defined by (4.35):*

$$|\mathcal{A}_j|<\epsilon,\qquad j=2,\dots,6.$$

*Proof.* We bound each of the errors $\mathcal{A}_j$, $j=2,\dots,9$:

$$|\mathcal{A}_2| = \left|Q_M^{\hat{s}}\left[\mathrm{e}^{s(y-y_\alpha+x_\beta-x_{\beta'},\hat{s})}M_{\alpha,\beta}(\hat{s})r_{N_*}(s,x-x_{\beta'},\hat{s})\right]\right| \\ \overset{(4.36,4.40)}{\le} \mathrm{e}^{\operatorname{Re} s(r_p+r_d)}L^2\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\,R_{N_*}(isr_c(x)) \le \frac{\epsilon}{N_*^2},$$

where the last bound follows from (4.47). Similarly the bound for $|\mathcal{A}_3|$ can be derived.

$$|\mathcal{A}_4| = \left|Q_M^{\hat{s}}\left[\mathrm{e}^{s(y-y_\alpha+x_\beta-x_{\beta'},\hat{s})}M_{\alpha,\beta}(\hat{s})\sum_{k=0}^{N_*-1}\frac{2k+1}{4\pi}Q_N^{\hat{q}}\left[r_{2N-N_*+1}(s,x-x_{\beta'},\hat{q})P_k(\hat{q}\cdot\hat{s})\right]\right]\right| \\ \overset{(4.36,4.37,4.40)}{\le} L^2\mathrm{e}^{\operatorname{Re} s(r_p+r_d)}\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\,R_{2N-N_*+1}(isr_c(x))\sum_{k=0}^{N_*-1}(2k+1) \\ \le L^2N_*^2\mathrm{e}^{\operatorname{Re} s(r_p+r_d)}\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\,R_{N_*}(isr_c(x)),$$

where we used $N \geq N_*$. The bound (4.47) gives $|\mathcal{A}_4| < \epsilon$. Again, the bound $|\mathcal{A}_5| < \epsilon$ can be derived using the same arguments. Similarly, we bound

$$|\mathcal{A}_6| = \left| Q_M^{\hat{s}} \left[ \mathrm{e}^{s(y_{\alpha'} - y_\alpha + x_\beta - x_{\beta'}, \hat{s})} r_{N_*}(s, y_{\alpha'} - y, \hat{s}) M_{\alpha,\beta}(\hat{s}) r_{N_*}(s, x - x_{\beta'}, \hat{s}) \right] \right|$$
$$\overset{(4.36,4.40)}{\leq} L^2 \mathrm{e}^{2\,\mathrm{Re}\, s r_d} \max\left( R_{N_*}^2(isr_c(x)), R_{N_*}^2(isr_c(y)) \right) |h_{L-1}(is\|c_{\alpha\beta}\|)| \leq \frac{\epsilon}{N_*^4},$$

where the last bound was obtained using (4.48).

$$|\mathcal{A}_7| = \left| Q_M^{\hat{s}} \left[ \mathrm{e}^{s(y_{\alpha'} - y_\alpha + x_\beta - x_{\beta'}, \hat{s})} r_{N_*}(s, x - x_{\beta'}, \hat{s}) M_{\alpha,\beta}(\hat{s}) \right.\right.$$
$$\left.\left. \times \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{q}} \left[ r_{2N-N_*+1}(s, y_{\alpha'} - y, \hat{q}) P_k(\hat{q} \cdot \hat{s}) \right] \right] \right|$$
$$\overset{(4.36,4.40)}{\leq} \mathrm{e}^{2\,\mathrm{Re}\, s r_d} R_{N_*}(isr_c(x)) L^2 |h_{L-1}(is\|c_{\alpha\beta}\|)| R_{2N-N_*+1}(isr_c(y)) \sum_{k=0}^{N_*-1} (2k+1)$$
$$\leq L^2 N_*^2 \max\left( R_{N_*}^2(isr_c(x)), R_{N_*}^2(isr_c(y)) \right) |h_{L-1}(is\|c_{\alpha\beta}\|)| \leq \frac{\epsilon}{N_*^2},$$

where $N_* \leq N$ and (4.48) were used. Analogically the same bound for $|\mathcal{A}_8|$ can be derived. Finally, the error

$$|\mathcal{A}_9| = \left| Q_M^{\hat{s}} \left[ \mathrm{e}^{s(y_{\alpha'} - y_\alpha + x_\beta - x_{\beta'}, \hat{s})} M_{\alpha,\beta}(\hat{s}) \right.\right.$$
$$\times \sum_{m=0}^{N_*-1} \frac{2m+1}{4\pi} Q_N^{\hat{r}} \left[ r_{2N-N_*+1}(s, x - x_{\beta'}, \hat{r}) P_m(\hat{r} \cdot \hat{s}) \right]$$
$$\left.\left. \times \sum_{k=0}^{N_*-1} \frac{2k+1}{4\pi} Q_N^{\hat{q}} \left[ r_{2N-N_*+1}(y_{\alpha'} - y, \hat{q}) P_k(\hat{q} \cdot \hat{s}) \right] \right] \right|$$
$$\overset{(4.36),(4.40)}{\leq} \mathrm{e}^{2\,\mathrm{Re}\, s r_d} N_*^4 L^2 |h_{L-1}(is\|c_{\alpha\beta}\|)| \max\left( R_{2N-N_*+1}^2(isr_c(x)), R_{2N-N_*+1}^2(isr_c(y)) \right) \leq \epsilon.$$

□

The results of numerical experiments in Table 4.1 show that a good estimate for

$$\max_{x \in \tau_{\beta'}, y \in \tau_{\alpha'}} \max\left( R_m(isr_c(x)), R_m(isr_c(y)) \right)$$

is $R_m(isr_c)$, where $r_c = \max\left( \frac{d_{\alpha'}}{2}, \frac{d_{\beta'}}{2} \right)$, $m \geq 1$. Hence a criterion based on the use of this value can serve for choosing the length of multipole expansions.

**Remark 4.3.8.** *The values $E^{(1)}, E^{(2)}$ decay with $\mathrm{Re}\, s > 0$, as $\mathrm{Im}\, s \to +\infty$, independently of the choice $N$, $M$ (however, with $L$ satisfying Lemma 4.3.6). To show this, let us fix the admissibility condition (assuming that the diameters $d_\alpha$, $d_\beta$ of the bounding boxes of clusters $\tau_\alpha$ and $\tau_\beta$ are equal):*

$$\|c_{\alpha\beta}\| \geq \frac{3}{4}(d_\alpha + d_\beta) = 3r_p. \tag{4.49}$$

*We will need two ingredients.*

1. *Exponentially fast decay of spherical Hankel functions as $\mathrm{Re}\, s \to +\infty$. As $|s|\,\mathrm{diam}\,\tau_\alpha \gg 1$, see Remark 4.2.7,*

$$L \ll |s|\|c_{\alpha\beta}\|.$$

*Hence, the spherical Hankel function $h_{L-1}(is\|c_{\alpha\beta}\|)$ is in the asymptotic regime (4.8), namely*

$$|h_{L-1}(is\|y_\alpha - x_\beta\|)| \sim (1 + \gamma(|s|\|c_{\alpha\beta}\|)) \frac{\mathrm{e}^{-\mathrm{Re}\, s\|c_{\alpha\beta}\|}}{|s|\|c_{\alpha\beta}\|}, \tag{4.50}$$

*where $\gamma(|s|\|c_{\alpha\beta}\|, L) = O\left( \frac{L^2}{|s|\|c_{\alpha\beta}\|} + \frac{L^4}{|s|^2\|c_{\alpha\beta}\|^2} \right)$.*

2. *Exponentially fast (but with a relatively small rate) growth of $|R_{N_*}(r_c)|$. Let us show that, given $r_c = \frac{1}{2}\max(d_{\alpha'}, d_{\beta'})$,*

$$R_{N_*}(isr_c(x)) \le C_1|s|^2 r_c^2 \mathrm{e}^{\mathrm{Re}\, sr_c}, \qquad C_1 > 0. \tag{4.51}$$

*The same holds for $R_{N_*}(isr_c(y))$. The expression (4.1) allows to derive a trivial bound:*

$$|j_n(isr_c(x))| = \left| \frac{1}{2}(-i)^n \int_{-1}^{1} \mathrm{e}^{-sr_c(x)} P_n(t)dt \right| \overset{(4.37)}{\le} \frac{1}{2}\mathrm{e}^{\mathrm{Re}\, sr_c}. \tag{4.52}$$

*Recall that spherical Bessel functions satisfy the bound (4.6):*

$$|j_n(z)| \le \mathrm{e}^{\mathrm{Im}\, z} \frac{\mathrm{e}}{2|z|\sqrt{\pi}} \left( \frac{|z|\mathrm{e}}{2(n+1)} \right)^{n+1}.$$

*Let $N' \in \mathbb{N}$, $N' \ge |z|\mathrm{e}$. Then for all $n \ge N'$, it holds that*

$$|j_n(z)| \le \mathrm{e}^{\mathrm{Im}\, z} \frac{\mathrm{e}}{2|z|\sqrt{\pi}} \left( \frac{|z|\mathrm{e}}{2(n+1)} \right)^{n+1} \le \mathrm{e}^{\mathrm{Im}\, z} \frac{\mathrm{e}}{2|z|\sqrt{\pi}} \left( \frac{1}{2} \right)^{n+1}. \tag{4.53}$$

*Then, given $N' \ge |s|r_c\mathrm{e} \ge |s|r_c(x)\mathrm{e}$,*

$$\begin{aligned} R_{N_*}(isr_c(x)) \le R_0(isr_c(x)) &= \sum_{n=0}^{\infty} (2n+1)|j_n(isr_c(x))| \\ &= \sum_{n=0}^{N'-1} (2n+1)|j_n(isr_c(x))| + \sum_{n=N'}^{+\infty} (2n+1)|j_n(isr_c(x))|. \end{aligned}$$

*The first term can be bounded using (4.52) by*

$$\frac{1}{2}N'^2 \mathrm{e}^{\mathrm{Re}\, sr_c(x)} < C|s|^2 r_c^2 \mathrm{e}^{\mathrm{Re}\, sr_c},$$

*for a constant $C$ that does not depend on $s$ and $r_c$. The second term can be bounded with the help of (4.53)*

$$\sum_{n=N'}^{+\infty} (2n+1)|j_n(isr_c(x))| \le C' \mathrm{e}^{\mathrm{Re}\, sr_c(x)} \le C' \mathrm{e}^{\mathrm{Re}\, sr_c},$$

*for some $C' > 0$. This lets obtaining (4.51). The above considerations can be repeated to bound $R_{N_*}(r_c)(y)$.*

*Now let us consider the errors $E^{(1)}$ and $E^{(2)}$. We insert (4.50,4.51) into the expression for $E^{(1)}$ to obtain*

$$E^{(1)} \sim C_1 L^2 N_*^2 \left(1 + \gamma\left(|s|\|c_{\alpha\beta}\|, L\right)\right) |s|^2 r_c^2 \frac{\mathrm{e}^{\mathrm{Re}\, s(r_p + r_d + r_c - \|c_{\alpha\beta}\|)}}{|s|\|c_{\alpha\beta}\|}.$$

*Clearly, $r_d < r_p$, $r_c < r_p$, hence, using (4.49),*

$$r_p + r_d + r_c - \|c_{\alpha\beta}\| < 0,$$

*which shows the exponential decay of $E^{(1)}$ with* $\mathrm{Re}\, s$.

*Similarly, we can substitute $R_{N_*}(r_c)$ and $h_{L-1}(is\|c_{\alpha\beta}\|)$ in (4.48) by their estimates (4.51, 4.50) to obtain:*

$$\begin{aligned} E^{(2)} &= L^2 N_*^4 \mathrm{e}^{2\,\mathrm{Re}\, sr_d} \left|h_{L-1}(is\|c_{\alpha\beta}\|)\right| \max\left(R_{N_*}^2(isr_c(x)), R_{N_*}^2(isr_c(y))\right) \\ &\sim L^2 N_*^4 C_1^2 |s|^4 r_c^4 \left(1 + \gamma(|s|\|c_{\alpha\beta}\|)\right) \frac{\mathrm{e}^{\mathrm{Re}\, s(2r_d + 2r_c - \|c_{\alpha\beta}\|)}}{|s|\|c_{\alpha\beta}\|}. \end{aligned}$$

*We consider two cases, the first one when an octree based partitioning of the domain is used, and another one when a binary partitioning is employed (see [22]). We assume that bounding boxes of clusters $\tau_\alpha, \tau_\beta$ are cuboids with sides $a, b, c$. Then $r_p = \frac{1}{2}\sqrt{a^2 + b^2 + c^2}$.*

1. *if an octree partitioning is used,* $r_c = \frac{1}{2} r_p = r_d$*, hence we obtain:*

$$E^{(2)} \sim L^2 N_*^4 C_1^2 |s|^3 r_c^4 \left(1 + \gamma(|s| \|c_{\alpha\beta}\|)\right) \frac{\mathrm{e}^{-\operatorname{Re} s(\|c_{\alpha\beta}\| - 2r_p)}}{\|c_{\alpha\beta}\|}.$$

*Due to the admissibility condition (4.49),* $\|c_{\alpha\beta}\| > 2r_p$*, hence* $E^{(2)}$ *decays with* $\operatorname{Re} s$*.*

2. *if a binary cluster tree is used, a parent cluster is split into two children clusters. More precisely, one of the sides of the bounding box is halved, and the bounding boxes of children clusters are defined as cuboids obtained as a result of this subdivision. We assume w.l.o.g.* $r_d = \frac{a}{4}$ *and* $r_c = \frac{1}{2}\sqrt{\frac{a^2}{4} + b^2 + c^2}$*. Hence, using (4.49), we obtain:*

$$E^{(2)} \sim L^2 N_*^4 C_1^2 |s|^3 r_c^4 \left(1 + \gamma(|s| \|c_{\alpha\beta}\|)\right) \frac{\mathrm{e}^{\operatorname{Re} s(\frac{a}{2} + \sqrt{\frac{a^2}{4} + b^2 + c^2} - \frac{3}{2}\sqrt{a^2 + b^2 + c^2})}}{\|c_{\alpha\beta}\|}.$$

*Since* $\frac{a}{2} + \sqrt{\frac{a^2}{4} + b^2 + c^2} - \frac{3}{2}\sqrt{a^2 + b^2 + c^2} = \frac{a}{2} - \frac{1}{2}\sqrt{a^2 + b^2 + c^2} + \sqrt{\frac{a^2}{4} + b^2 + c^2} - \sqrt{a^2 + b^2 + c^2} < 0$*, the above expression decays exponentially with* $\operatorname{Re} s \to +\infty$*.*

*Note that these considerations are also true when applied to the admissibility condition typically used in the HF FMM (*$\|c_{\alpha\beta}\| \geq \frac{4}{\sqrt{3}} r_p$*) combined with the uniform octree partitioning.*

In the next section we discuss practical implications of the above bounds. We proceed as follows. First, we consider the expressions of Lemma 4.3.4 and with the help of extensive numerical experiments demonstrate that in practice only (4.42) for $\hat{v} \cdot \hat{c}_{\alpha\beta} = 1$ needs to be controlled. Unfortunately, we were not able to provide a formal theoretical justification to this fact. Second, we numerically study the expressions of Lemma 4.3.7. We demonstrate that usually there is no need to check the conditions of Lemma 4.3.7, however, in some cases the use of these bounds can improve the complexity of the algorithm.

### 4.3.2 One-Level FMM Error

Let us consider expressions (4.42) and (4.43), namely

$$E_{tr}^{\hat{v}\cdot\hat{c}_{\alpha\beta},L} := \left| \sum_{m=L}^{\infty} (2m+1) j_m(is\|v\|) h_m(is\|c_{\alpha\beta}\|) P_m\left(\hat{v}\cdot\hat{c}_{\alpha\beta}\right) \right|, \tag{4.54}$$

$$E_I^{\hat{v},\hat{c}_{\alpha\beta},L,M} := \left| \sum_{k=2M-L+1}^{+\infty} (2k+1) i^k j_k(is\|v\|) Q_M^{\hat{s}}\left[M_{\alpha,\beta}\left(\hat{s}\right) P_k\left(\hat{s}\cdot\hat{v}\right)\right] \right|. \tag{4.55}$$

Under the assumption $M \geq L$, the error of the one-level FMM can be bounded by the sum of these expressions, see Lemma 4.3.4. Recall that given $c_{\alpha\beta} \in \mathbb{R}^3$ both errors have to be controlled for a range of $v$, more precisely, for $\hat{v} \in \mathbb{S}^2$ and $\|v\| \in \left[0,\ \rho^{-1}\|c_{\alpha\beta}\|\right]$ (here $\rho$ is a fixed number defined by the admissibility condition, see Section 4.2). We will numerically study only the case $\|v\| = \rho^{-1}\|c_{\alpha\beta}\|$, as it is often done in the HF FMM theory. In the following two sections we conduct numerical experiments that demonstrate that to control the error of the one level FMM it is sufficient to consider only $E_{tr}^{1,L}$. Unfortunately we were not able to provide a theoretical justification for this fact, though a related result has been proved in [25].

For further considerations we fix $\epsilon > 0$ being the desired accuracy, $\|c_{\alpha\beta}\|$ and $\|v\| = \rho^{-1}\|c_{\alpha\beta}\|$.

#### Truncation Error

There exists several strategies to truncate the series (4.54):

1. use the bound

$$\left|h_0(is\|c_{\alpha\beta}-v\|)-\sum_{\ell=0}^{L-1}(2\ell+1)h_\ell(is\|c_{\alpha\beta}\|)j_\ell(is\|v\|)P_\ell\left(\hat{c}_{\alpha\beta}\cdot\hat{v}\right)\right| \\ \leq\sum_{\ell=L}^{\infty}(2\ell+1)\left|h_\ell(is\|c_{\alpha\beta}\|)j_\ell(is\|v\|)\right|=:E_{abs}^{L}, \tag{4.56}$$

   and then estimate the sum of this series by the first term (as geometrically convergent series), see e.g. [33] and [27]. Remarkably precise formulas for $L$ valid for large purely imaginary $s$ were analytically derived in [25]. Alternatively, one can use empirically derived bounds for $L$. We will denote the smallest $L$ stemming from the condition $E_{abs}^{L}<\epsilon$ by $L_{abs}$.

2. make use of (4.12) explicitly; more precisely, let $L_{unif}$ be the smallest integer satisfying

$$E_{tr}^{\cos\psi,L_{unif}}<\epsilon$$

   for all $\psi\in[0,2\pi)$. This is an approach suggested in [25]; the results of [25] were used in [26].

For the real wavenumber case, as numerical experiments in [25] show, the difference between $L_{abs}$ and $L_{unif}$ is significant only for very large values of $|s|$, close to $10^3$ (e.g. for $is=200$, for accuracies $10^{-2}$ and $10^{-5}$ the difference in $L_{abs}$ and $L_{unif}$ does not exceed $L_{abs}-L_{unif}\leq 3$). In the decay

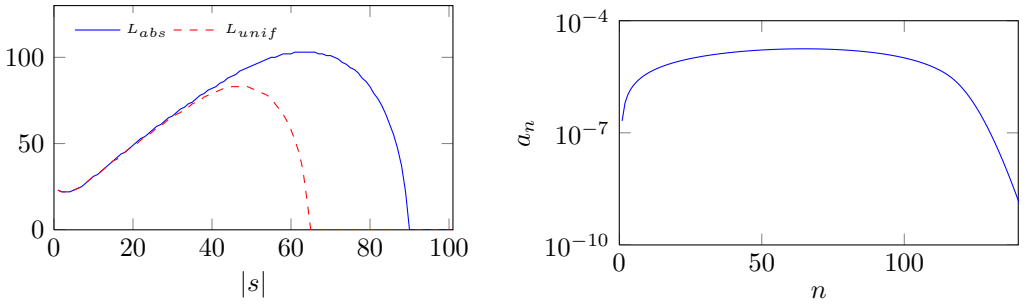


*Figure 4.8: In the left plot we show $L_{unif}$ and $L_{abs}$ determined numerically, $\|c_{\alpha\beta}\|=3$, $\|v\|=2$, $\epsilon=10^{-4}$, $s=|s|\mathrm{e}^{i0.95\frac{\pi}{2}}$, for different values of $|s|$. The right plot depicts absolute values of first 140 terms of the Gegenbauer's series (namely $a_n=(2n+1)\left|h_n(is\|c_{\alpha\beta}\|)j_n(is\|v\|)\right|$, $0\leq n\leq 139$), with $\|c_{\alpha\beta}\|=3$, $\|v\|=2$ and $s=60\mathrm{e}^{i0.95\frac{\pi}{2}}$. In this case $|h_0(is\|c_{\alpha\beta}-v\|)|<1.6e-4$, however, $\left|\sum_{\ell=L}^{+\infty}(2\ell+1)|h_\ell(is\|c_{\alpha\beta}\|)j_l(is\|v\|)\right|<1e-4$ only starting with $L\geq 102$.*

case the situation is different, as Figure 4.8: even for $|s|<100$ $L_{abs}$ and $L_{unif}$ differ significantly. Indeed, in points where $h_0(is\|x-y\|)$ is close to (or smaller than) $\epsilon$ the sum

$$\sum_{\ell=0}^{+\infty}(2\ell+1)|j_\ell(is\|v\|)h_\ell(is\|c_{\alpha\beta}\|)|>\epsilon,$$

though each of the terms of the above series is smaller than $\epsilon$. This is also shown in Figure 4.8.

In further sections we show that the use of $L_{unif}$ instead of $L_{abs}$ provides a good error control for the truncation error. First, however, let us address a question of an actual evaluation of $L_{unif}$. For this, let us study the dependence of $L_\psi$, which we define as the smallest integer for which $E_{tr}^{\cos\psi,L}<\epsilon$, on $\cos\psi$. In [25] it was proved that in the no-decay case for large $|s|$ and $L\geq|s|\|v\|-\frac{1}{2}$ the error

$$\mathcal{E}^{\hat{c}_{\alpha\beta}\cdot\hat{v},L}=\left|e^{-s\|c_{\alpha\beta}-\|v\|\hat{v}\|}-\|c_{\alpha\beta}-v\|\sum_{\ell=0}^{N-1}(2\ell+1)h_\ell(is\|c_{\alpha\beta}\|)j_\ell(is\|v\|)P_\ell(\hat{c}_{\alpha\beta}\cdot\hat{v})\right| \tag{4.57}$$

achieves its maximum when $\hat{c}_{\alpha\beta} \cdot \hat{v} = \pm 1$. The formulas provided in [25] show that the errors in $\hat{c}_{\alpha\beta} \cdot \hat{v} = 1$ and $\hat{c}_{\alpha\beta} \cdot \hat{v} = -1$ differ insignificantly.

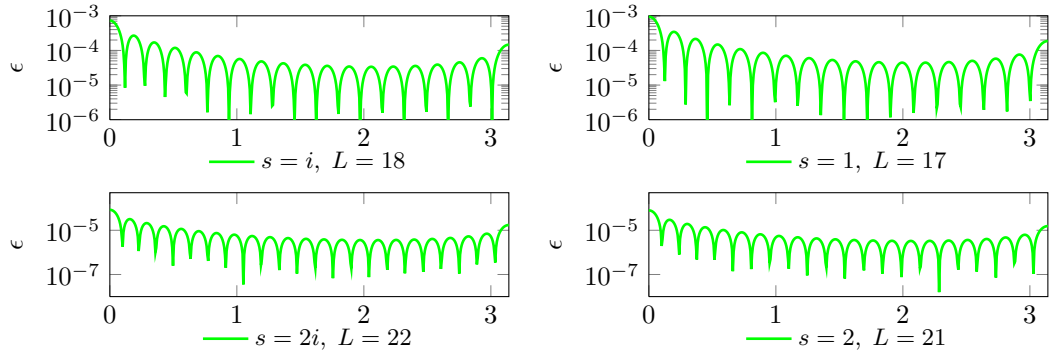


*Figure 4.9: Given the length of the expansion $L$, $\|c_{\alpha\beta}\| = 3$, $\|v\| = 2$, we study the dependence of the error of the truncation of the series $E_{tr}^{\cos\psi,L}$ defined by (4.57) on $\psi$. In all the experiments the error has a global maximum at $\psi = 0$.*

The results of our numerical experiments in Figures 4.9, 4.10 and 4.11 confirm that the maximum of the absolute error $E_{tr}^{\cos\psi,L}$ is almost always achieved when $\cos\psi = 1$, even for small values of $L$ (to check this we computed the smallest $L$, s.t. $E_{tr}^{1,L} < \epsilon$, and evaluated $E_{tr}^{\cos\psi,L}$ for a range of $\psi$). In case when this does not hold, the maximum of the error never exceeded the value of the error in $\psi = 0$ more than twice. In $\psi = \pi$ there is a local maximum. Hence, in practice it is reasonable to use the value $\psi = 0$ for controlling the truncation error (4.54). Unfortunately we were not able to prove this formally for $s \in \mathbb{C}$, as it was done in [25] for the no-decay case in the asymptotic regime $|s| \to +\infty$.

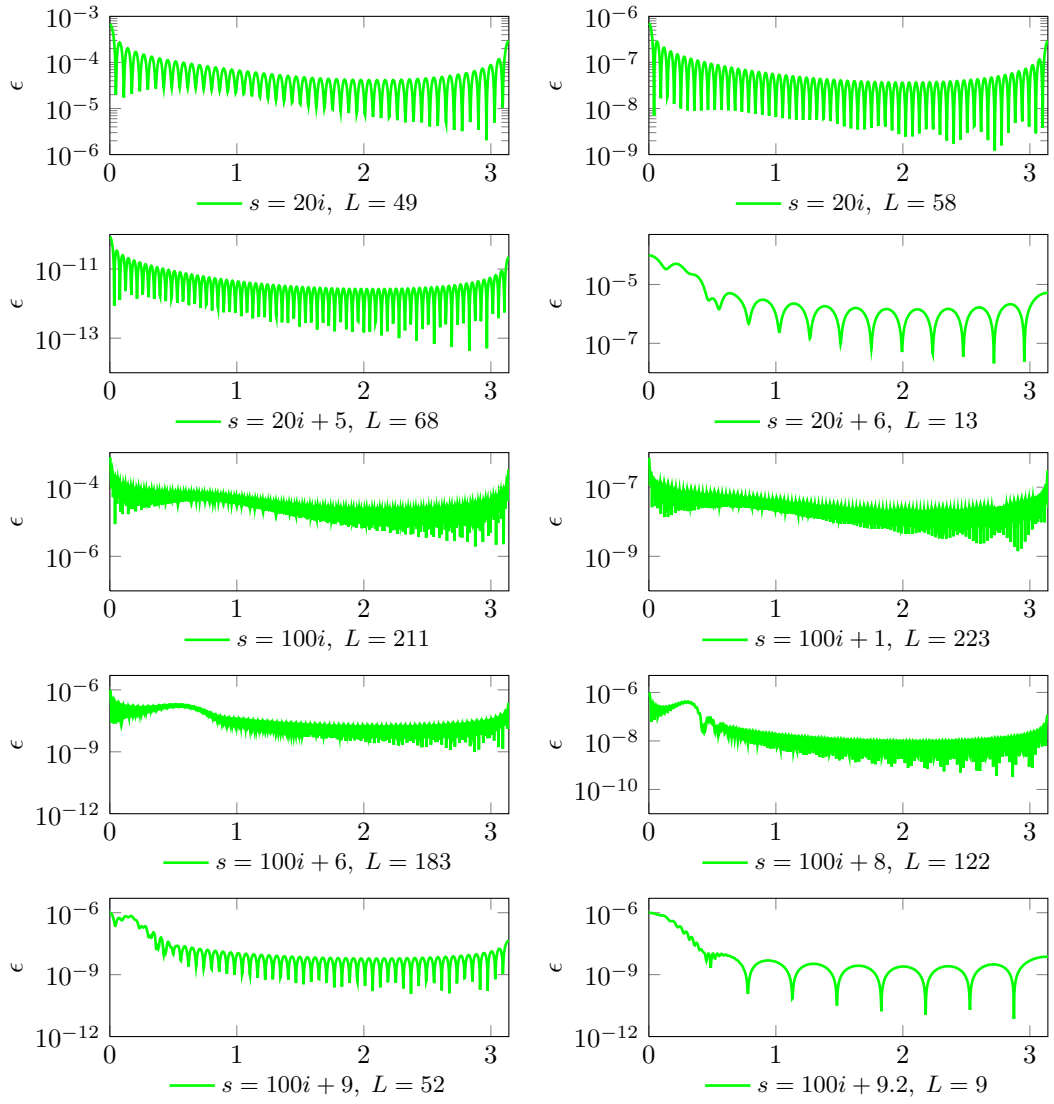


*Figure 4.10: Given the length of the expansion $L$, $\|c_{\alpha\beta}\| = 3$, $\|v\| = 2$, we study the dependence of the error of the truncation of the series $E_{tr}^{\cos\psi,L}$ defined by (4.57) on $\psi$. In all the experiments the error has a global maximum at $\psi = 0$.*

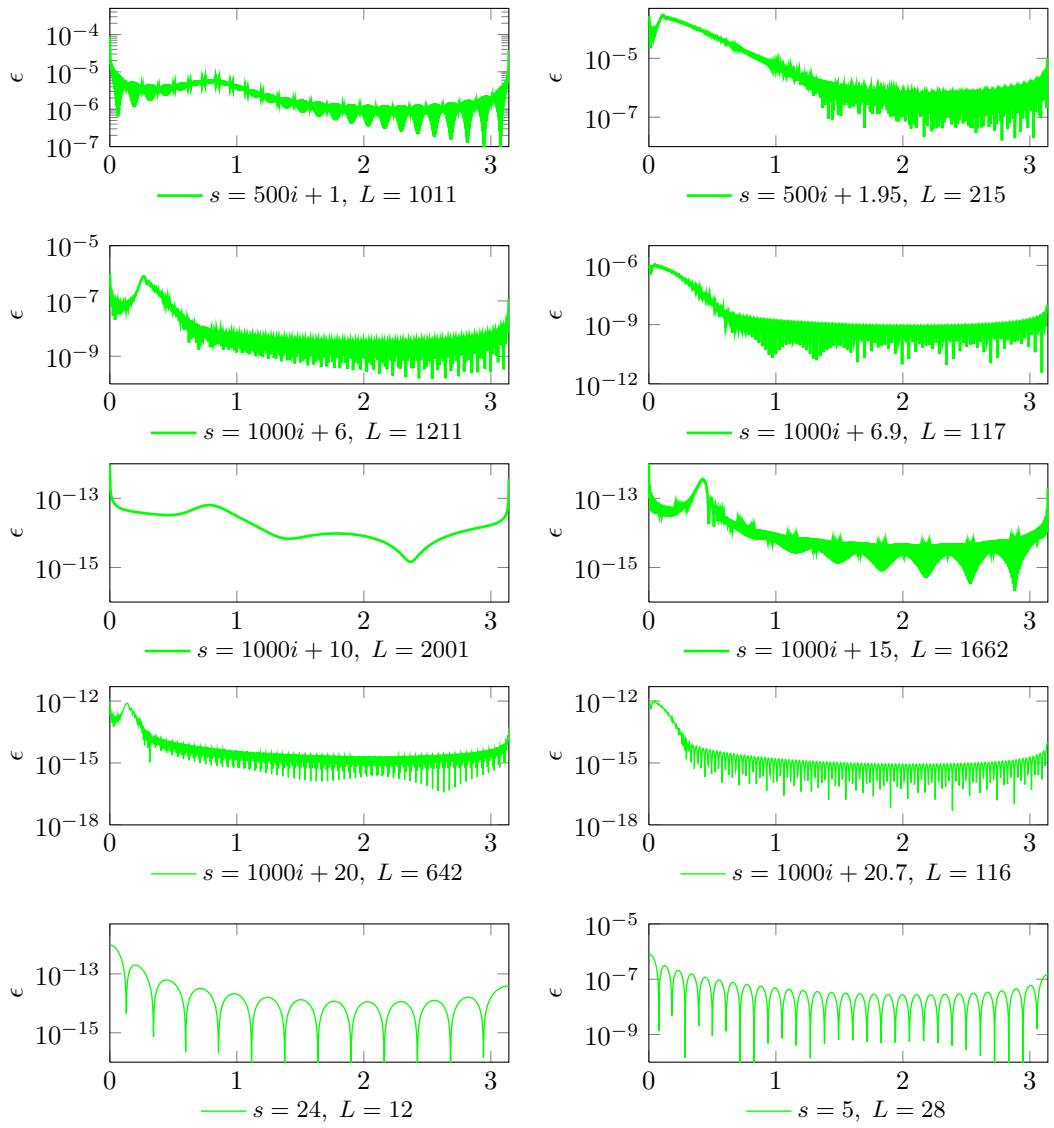


*Figure 4.11: Given the length of the expansion $L$, $\|c_{\alpha\beta}\| = 3$, $\|v\| = 2$, we study the dependence of the error of the truncation of the series $E_{tr}^{\cos\psi, L}$ defined by (4.57) on $\psi$. In almost all the experiments the error has a global maximum at $\psi = 0$; if this is not the case the value of the error at $\psi = 0$ differs insignificantly from the global maximum.*

**The Total Error of the One-Level FMM**

Recall that according to Lemma 4.3.4, under assumption $M \geq L$, the error of the one-level fast multipole method can be bounded by a sum of truncation (4.42) and integration errors (4.43). More precisely, the integration error (4.43) is defined as

$$E_I^{\hat{v},\hat{c}_{\alpha\beta},L,M} = \left| \sum_{k=2M-L+1}^{+\infty} (2k+1) i^k j_k(is\|v\|) Q_M^{\hat{s}} \left[ M_{\alpha,\beta}(\hat{s}) P_k(\hat{s}\cdot\hat{v}) \right] \right|,$$

where $M_{\alpha,\beta}(\hat{s}) = \frac{1}{4\pi} \sum_{\ell=0}^{L-1} (2\ell+1)(-i)^\ell h_\ell(is\|c_{\alpha\beta}\|) P_\ell(\hat{c}_{\alpha\beta}\cdot\hat{s})$ and $M \geq L$. In this section we would like to demonstrate that there is no need to account for the integration error separately when determining the length of the multipole expansion, and the criterion based solely on the truncation error analysis can be used.

In numerical experiments of this section we compute the total error and demonstrate that it can be estimated by the truncation error $E_{tr}^{1,L}$, and hence the integration error is significantly smaller than the truncation error. In the no-decay case this was discussed in [69], for the case $\operatorname{Re} s = 0$. This fact was also confirmed by various experiments of [27]. In the latter work it was shown that choosing the length of the multipole expansion by examining the last term of the Gegenbauer's series provides a good total error control. Numerical experiments of this section confirm that the choice of the length of multipole expansion based solely on the analysis of the convergence of the Gegebauer's series allows to control the total error of the one-level FMM in the decay regime, i.e. $\operatorname{Re} s > 0$.

We fix $\|c_{\alpha\beta}\| = 3$, $\|v\| = 2$, $\epsilon = 10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}$, find the smallest $L$, s.t. $E_{tr}^{1,N} < \epsilon$ for all $N \geq L$, and compute the total error, setting the parameter $M$ that defines the length of the multipole expansion to the smallest possible value, namely $M = L$. The results of this comparison can be found in Figure 4.12. In general, the integration error (and hence the total error, see also Lemma 4.3.4) $E_I^{\hat{v},\hat{c}_{\alpha\beta},L,L}$ depends on directions $\hat{c}_{\alpha\beta}$ and $\hat{v}$. From the definition of $E_I^{\hat{v},\hat{c}_{\alpha\beta},L,L}$, see (4.55), it follows that w.l.o.g. we may fix $\hat{c}_{\alpha\beta} = (1,0,0)$. Given $\hat{c}_{\alpha\beta}$, $\|c_{\alpha\beta}\|$, $\|v\|$, we study numerically the dependence of the total error on $\hat{v} \in \mathbb{S}^2$.

In Figure 4.12 we show the dependence of the computed value (4.34)

$$\mathcal{E}_t = \max_{\hat{v}\in\mathbb{S}^2} |\mathcal{A}_1(L, L, x(\hat{v}), x_\beta, y(\hat{v}), y_\alpha, s)| \tag{4.58}$$

on $s$ (here $x(\hat{v}), y(\hat{v}) \in \mathbb{R}^3$ are chosen so that $v = x - x_\beta + y_\alpha - y$, $\|v\|$ fixed). The value $L$ is chosen as suggested in Section 4.3.2, i.e. as the smallest $N$, s.t. $E_{tr}^{1,N} < \epsilon$. To exclude the contribution of the roundoff errors, see Section 4.4, we compute $\mathcal{E}_t$ only for values of $s, L$ s.t.

$$L|h_{L-1}(is\|c_{\alpha\beta}\|)| \mathrm{e}^{\operatorname{Re} s\|v\|} \epsilon_m < \epsilon,$$

where $\epsilon_m$ is the machine precision, see (4.65). We can see that in all the numerical experiments the total error $\mathcal{E}_t$ can be bounded by $CE_{tr}^{1,L}$, where $C$ is a constant close to 1. We are not aware of any works dedicated to derivation of nearly optimal error bounds for the FMM integration error, even for the case of the real wavenumber. However, the results of numerical studies suggest that the criterion solely based on the analysis of $E_{tr}^{1,L}$ provides a good error control of the HF FMM.

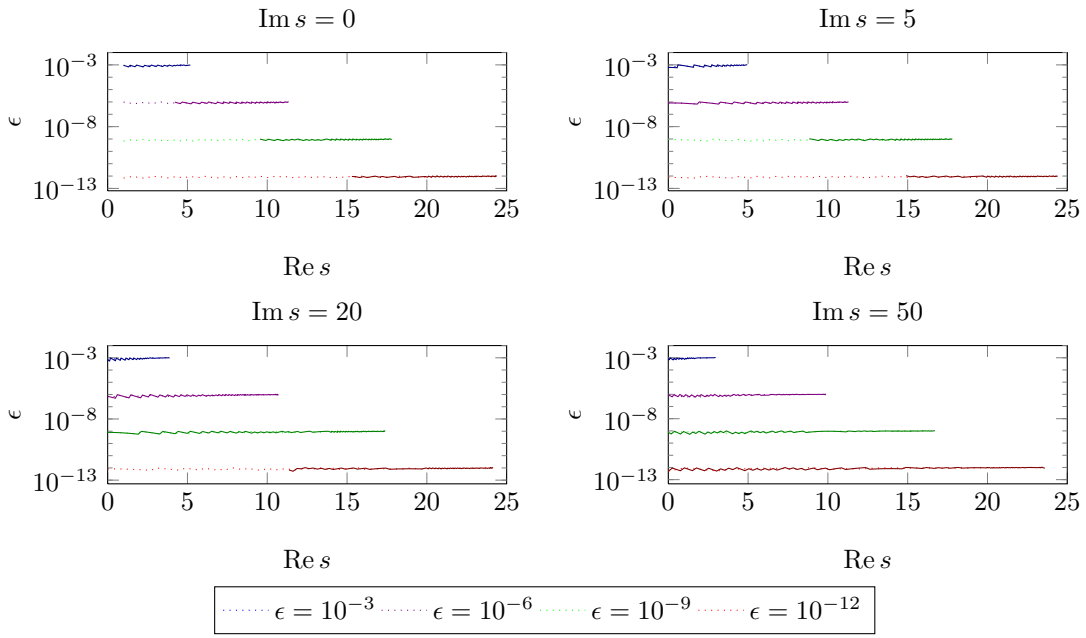


*Figure 4.12: Given the smallest $L$ satisfying $E_{tr}^{1,L} < \epsilon$, we show the numerically computed truncation error $E_{tr}^{1,L}$ and $\mathcal{E}_t$ defined by (4.58) for different values of $s$ and $\epsilon$. For a fixed $\epsilon$, the dark line corresponds to the total error (not measured when the roundoff error exceeds $\epsilon$) and the light dotted one to $E_{tr}^{1,L}$. In all the experiments $\|c_{\alpha\beta}\| = 3$ and $\|v\| = 2$.*

### 4.3.3 Errors due to M2M and L2L Translations

In order to control the error during multipole-to-local and local-to-local translations we make use of Lemma 4.3.7. We adopt the same notation as in Section 4.3.1, see Remark 4.3.1. In particular, given $\epsilon > 0$, it is suggested that $N_* = \min(N, M)$ (where $N$ defines the length of the multipole expansion at the children level and $M$ at the parent level, see Section 4.3.1) satisfies:

$$\begin{aligned} E^{(1)} &= L^2 N_*^2 \mathrm{e}^{\operatorname{Re} s(r_p+r_d)} \, |h_{L-1}(is\|c_{\alpha\beta}\|)| \max\left(R_{N_*}(isr_c(x)), R_{N_*}(isr_c(y))\right) \le \epsilon, \\ E^{(2)} &= L^2 N_*^4 \mathrm{e}^{2\operatorname{Re} sr_d} \, |h_{L-1}(is\|c_{\alpha\beta}\|)| \max\left(R_{N_*}^2(isr_c(x)), R_{N_*}^2(isr_c(y))\right) \le \epsilon, \end{aligned} \tag{4.59}$$

for all $y \in \tau_\alpha$, $x \in \tau_\beta$. Recall that $M, L$ satisfy Lemma 4.3.7 and $R_m(isr) = \sum_{n \ge m} (2n+1)|j_n(isr)|$.

First, let us numerically demonstrate that for all $x \in \mathbb{R}^3$, s.t. $\|x\| \le r_c$, the value $R_M(isr_c)$ serves as a good bound for $R_M(is\|x\|)$. This is shown in Table 4.1.

| $s$ | $R(is)$ | $s$ | $R(is)$ | $s$ | $R(is)$ | $s$ | $R(is)$ |
|---|---|---|---|---|---|---|---|
| $10i$ | 0 | $50i$ | 0.0056 | $400i$ | 0.0055 | $100i$ | 0.0033 |
| $10i + 0.25$ | 0 | $50i + 0.25$ | 0.00062 | $400i + 0.25$ | 0.0013 | $200i$ | 0.0064 |
| $10i + 1$ | 0 | $50i + 1$ | 0 | $400i + 1$ | 0 | $500i$ | 0.0061 |
| $10i + 5$ | 0 | $50i + 5$ | 0 | $400i + 5$ | 0 | $1000i$ | 0.0044 |
| $10i + 10$ | 0 | $50i + 10$ | 0 | $400i + 10$ | 0 | $2000i$ | 0.0019 |
| $10i + 20$ | 0 | $50i + 20$ | 0 | $400i + 20$ | 0 | $5000i$ | 0.0009 |

*Table 4.1: The value $R(is) = \max\limits_{1 \le n \le n_{\max}} \frac{\max\limits_{\alpha \in (0,\,1]} R_n(is\alpha) - R_n(is)}{R_n(is)}$ determined numerically ($n_{\max}$ is chosen so that $|R_{n_{\max}}(is)| < 1e-16$) for different values of $s$. We can see that the maximum of $R_n(is\alpha)$, with $\alpha \in (0, 1]$, is close to $R_n(is)$ (or equals $R_n(is)$ when $\operatorname{Re} s \ge 1$).*

First we would like to demonstrate that conditions of Lemma 4.3.7 and Lemma 4.3.4 indeed

allow to control the error of the multilevel HF FMM.

Next, based on these numerical experiments we discuss if it is always necessary to check the conditions of Lemma 4.3.7, or the use of Lemma 4.3.4 can solely provide a good error control.

We start with numerical experiments for several different configurations of the clusters shown in Figure 4.13.

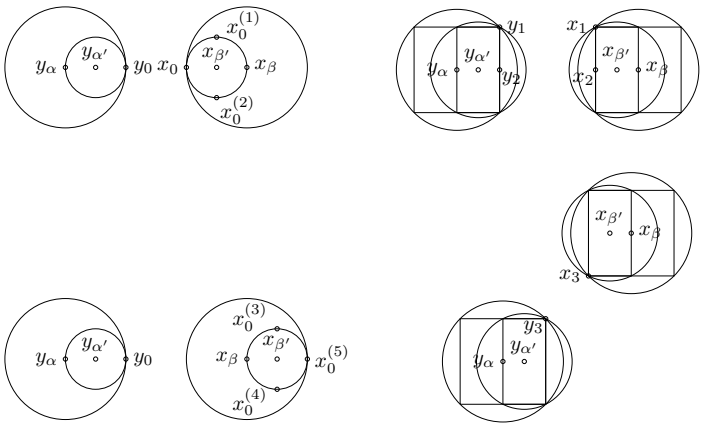


*Figure 4.13: Schematic 2D presentation of configurations that we used in experiments. Cluster parameters are given in the table below.*

| *Value* | *Upper left plot* | *Upper right plot* | *Lower left plot* | *Lower right plot* |
|---|---|---|---|---|
| $y_\alpha$ | $(0,0,0)$ | $(0,0,0)$ | $(0,0,0)$ | $(0,0,0)$ |
| $x_\beta$ | $(3,0,0)$ | $(3,0,0)$ | $(3,0,0)$ | $3\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ |
| $y_{\alpha'}$ | $(0.5,0,0)$ | $(\frac{1}{2\sqrt{3}},0,0)$ | $(0.5,0,0)$ | $(\frac{1}{2\sqrt{3}},0,0)$ |
| $x_{\beta'}$ | $(2.5,0,0)$ | $\left(3-\frac{1}{2\sqrt{3}},0,0\right)$ | $(3.5,0,0)$ | $(\sqrt{3}-\frac{1}{2\sqrt{3}},\sqrt{3},\sqrt{3})$ |
| $r_p$ | 1 | 1 | 1 | 1 |
| $r_d$ | 0.5 | $\frac{1}{2\sqrt{3}}$ | 0.5 | $\frac{1}{2\sqrt{3}}$ |
| $r_c$ | 0.5 | $\frac{\sqrt{3}}{2}$ | 0.5 | $\frac{\sqrt{3}}{2}$ |

*We compute a multilevel FMM approximation to $h_0(is\|x_j-y_j\|)$, $j=0,\dots,3$, and $h_0(is\|x_0^{(k)}-y_0\|)$, $k=1,\dots,5$. The coordinates of these points are provided below.*

| | |
|---|---|
| $x_0=(2,0,0)$ | $y_0=(1,0,0)$ |
| $x_0^{(1)}=(2.5,0.5,0)$ | $x_0^{(2)}=(2.5,-0.5,0)$ |
| $x_1=\left(3-\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ | $y_1=\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ |
| $x_2=\left(3-\frac{1}{\sqrt{3}},0,0\right)$ | $y_2=\left(\frac{1}{\sqrt{3}},0,0\right)$ |
| $x_0^{(3)}=(3.5,0.5,0)$ | $x_0^{(4)}=(3.5,-0.5,0),\ x_0^{(5)}=(4,0,0)$ |
| $x_3=3\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)-\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ | $y_3=\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ |

First, given $\epsilon>0$, we define $L$ as the smallest integer satisfying

$$\left|h_0(is(\|c_{\alpha\beta}\|-2r_p))-\sum_{k=0}^{L-1}(2k+1)h_k(is\|c_{\alpha\beta}\|)j_k(2isr_p)\right|<\epsilon.$$

Next, to control the errors $E^{(1)},E^{(2)}$, we look for $N>0$ s.t.

$$\begin{aligned}|(L-1)h_{L-1}(is\|c_{\alpha\beta}\|)\exp(\operatorname{Re}s(r_p+r_d))j_N(isr_c)N|<\epsilon,\\ \left|(L-1)h_{L-1}(is\|c_{\alpha\beta}\|)\exp(2\operatorname{Re}sr_d)j_N^2(isr_c)N\right|<\epsilon.\end{aligned}\tag{4.60}$$

Compared to the explicit expressions for $E^{(1)}$, $E^{(2)}$, we dropped some powers of $N,L$, since our estimates in Lemma 4.3.7 are non-optimal. Additionally, we estimated $R_N(isd)$ by $|j_N(isd)|$, since the corresponding series converges supergeometrically. Then the length of the expansion on the

level where the children clusters are located is chosen as $2N^2$. To choose the length of the multipole expansion $2M^2$ on the level where the parent clusters are located, we set $M = \max(L, N)$. We compute a multilevel HF FMM approximation to $h_0(is\|x_j - y_j\|)$, $j = 0, \ldots, 3$, and $h_0(is\|x_0^{(k)} - y_0\|)$, $k = 1, \ldots, 5$, for configurations depicted in Figure 4.13. For all $s$ the largest values of the error was achieved for pairs $(x_0, y_0)$ and $(x_3, y_3)$. The measured error never exceeded $2\epsilon$.

We plot the dependence of the measured error of the approximation of $h_0(is\|x_0 - y_0\|)$ on $\operatorname{Re} s$ in the right plot of Figure 4.14.

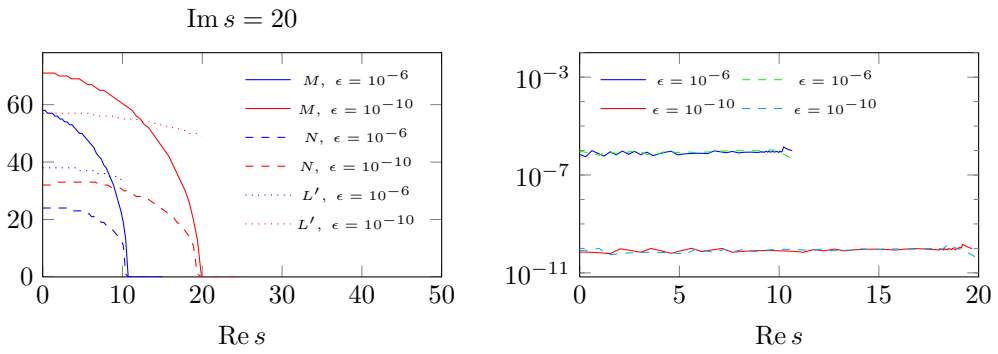


*Figure 4.14: In the left plot we depict $M = L, N$ computed as in Lemma 4.3.7 and Lemma 4.3.4 and $L'$ computed per Lemma 4.3.4 (under the assumption that a child cluster has an admissible neighbor, see (4.61) with $\tilde{c} = 3r_c$) for $\operatorname{Im} s = 20$ and two different accuracy settings. In the right plot the dependence of the error of the approximation of $h_0(is\|x_0 - y_0\|)$ on $\operatorname{Re} s$ for a fixed imaginary $\operatorname{Im} s$ is shown, for two given accuracy settings, $\epsilon = 10^{-6}$ and $\epsilon = 10^{-10}$. With dashed lines we demonstrate the computed error for $\operatorname{Im} s = 50$ and with solid lines the computed error for $\operatorname{Im} s = 20$ is shown.*

The results of these numerical experiments show that the error control provided by the use of such length of expansions is fairly robust (see Figure 4.14). Let us now discuss if the conditions of Lemma 4.3.7 require to be checked at all. First, let the parent clusters be located at the level $\ell - 1$ and the children clusters at the level $\ell$.

Let us assume that at the level $\ell$ all the clusters are leaves and there exists at least one pair of admissible clusters. Let also the centers of the bounding boxes of these clusters are located at the distance $\tilde{c}$. In this case the length of the multipole expansion at the level $\ell$ is not smaller than the one provided by Lemma 4.3.4. More precisely, as discussed before, we require that $N \geq L'$, where $L'$ is s.t.

$$\left| \sum_{n=L'}^{\infty} (2n+1) h_n(is\tilde{c}) j_n(2isr_c(x)) P_n(\cos\alpha) \right| < \epsilon, \quad \text{for all } 0 < r_c(x) \leq r_c \text{ and } \alpha \in [0, 2\pi), \tag{4.61}$$

where $r_c$ is the half-diameter of a bounding box on the level $\ell$. Such inequality should hold for all admissible pairs at the level $\ell$.

In the case of no-decay and $|s| \to +\infty$, the truncation parameter $L'$, and hence $N$, has to be chosen larger than $|2isr_c|$ (this is shown [25]), namely, when $j_{L'}(2isr_c)$ starts decaying superexponentially. Similarly, at the parent level $M \geq L$ and $L > |2isr_p|$. In this case it is guaranteed that errors $E^{(1)}$ and $E^{(2)}$ are small, since $N_* \geq \min(L, L') \gg |2isr_c|$ (recall that $|s|$ is large), and hence $|j_{N_*}(isr_c)|$ in (4.59) is in the regime of the superexponential decay (4.4). Therefore the expressions of (4.59) are negligibly small.

In the decay case, the same arguments as above can also apply. However, if decay is large enough, it may occur that $\min(L, L') < |2isr_c|$. Nevertheless, as results of the numerical experiments in Figure 4.14 show, the computed value of $L'$ is always larger than $N$ computed in accordance with (4.60). We were not able to detect a case when this does not hold.

Nevertheless, in practice we recommend performing the checks (4.60). Let us assume that an FMM accelerated matrix-vector product has to be computed for a range of complex frequencies

and the same block-cluster tree should be used for all of them. In this case it may appear that at the levels of the block-cluster tree where leaves are located, due to the low-frequency breakdown of the FMM there may be no admissible clusters (this will be explained in Section 4.4). For some geometries this is also possible at other levels of the block-cluster tree, when a binary cluster tree is employed. In these cases to determine the length of the expansion we may *assume* that there is at least one pair of admissible clusters at this level and compute the length of the expansion as discussed above. However, as the results of numerical experiments show, it may appear advantageous for the efficiency of the algorithm (and in some cases easier to implement) to make use of bounds (4.60) instead: they will provide a smaller length of the expansion sufficient to achieve a desired accuracy.

## 4.4 Numerical Stability and Control of Roundoff Errors

There are two sources of round-off errors when the high-frequency fast multipole method is applied to Helmholtz problems with complex wavenumbers. The first one is connected to the exponential growth of spherical Hankel functions $h_n(z)$ when $n \gg |z|$ and is also inherent to the HF FMM applied to the problems with a purely real wavenumber, see e.g. [73]. The second one is intrinsic to the HF FMM applied to the Helmholtz equation with large decay and was studied in [91]. Importantly, these errors occur in different situations: the first one appears only when small clusters are considered, while the second one is likely to appear when applying the high-frequency fast multipole method to distant (and hence, due to the definition of the admissibility condition, large) clusters. In the following section we study an effect of these errors on the high-frequency FMM.

The low-frequency breakdown of the fast multipole method occurs when performing the multipole-to-local transform between small admissible clusters. One of the ways to control this error was suggested in [27]: there numerically determined bounds on size of clusters were used (e.g. to achieve an accuracy $10^{-3}$, the authors recommend to use HF FMM only for clusters whose size exceeds $\frac{1}{4}$ of a wavelength). Indeed, such strategy has to be adapted to different admissibility conditions, as well as to the presence of decay, which can (though not always) decrease the magnitude of rounding errors.

Our strategy of the roundoff error control is based on the following observation. In the simplest case of the one-level fast multipole method $h_0(is\|x-y\|)$ is approximated by the scalar product

$$
\begin{aligned}
h_0(is\|x-y\|) &\approx \sum_{\ell=1}^{2M^2} w_\ell \mathrm{e}^{-s(x-x_\beta,\hat{s}_\ell)} M_{\alpha,\beta}(\hat{s}_\ell) \mathrm{e}^{s(y-y_\alpha,\hat{s}_\ell)} = A^T B, \qquad (4.62)\\
A &= \left[ w_1 \mathrm{e}^{-s(x-x_\beta,\hat{s}_1)}, \ldots, w_{2M^2} \mathrm{e}^{-s(x-x_\beta,\hat{s}_{2M^2})} \right]^T,\\
B &= \left[ M_{\alpha,\beta}(\hat{s}_1) \mathrm{e}^{s(y-y_\alpha,\hat{s}_1)} \ldots, M_{\alpha,\beta}(\hat{s}_{2M^2}) \mathrm{e}^{s(y-y_\alpha,\hat{s}_{2M^2})} \right]^T,
\end{aligned}
$$

where $(w_k, \hat{s}_k)_{k=1}^{2M^2}$ are the quadrature nodes and weights and $M_{\alpha,\beta}$ is given by (4.30). The following lemma from [66, Section 3.1] bounds the error of the evaluation of the scalar product in the finite precision arithmetic.

**Lemma 4.4.1.** *Given $x, y \in \mathbb{R}^n$, let $s_n = x^T y$ and $\hat{s}_n = fl(x^T y)$ be the inner product $x^T y$ computed with no overflow or underflow in the finite precision arithmetic compliant with the standard model, i.e. for all floating point numbers $a, b$*

$$
fl(a \circ b) = a \circ b(1+\delta),\ |\delta| < \epsilon_m; \circ = +, -, *, \backslash, \qquad (4.63)
$$

*where $\epsilon_m$ is a machine accuracy. Then,*

$$
\begin{aligned}
|\hat{s}_n - s_n| &\le \gamma_n \sum_{i=1}^{n} |x_i y_i|,\\
\gamma_n &= \frac{n\epsilon_m}{1 - n\epsilon_m}. \qquad (4.64)
\end{aligned}
$$

Questions of the accuracy of the complex floating point arithmetic are considered in [66, Lemma 3.5]. In a nutshell, it is possible to implement the basic arithmetic operations so that

$$\begin{aligned} fl(a \circ b) &= a \circ b(1+\delta),\ |\delta| < \epsilon_m, ; \circ = +, -, \\ fl(ab) &= ab(1+\delta),\ |\delta| < \sqrt{2}\gamma_2, \\ fl\left(\frac{a}{b}\right) &= \frac{a}{b}(1+\delta),\ |\delta| < \sqrt{2}\gamma_4, \end{aligned}$$

where $\gamma_n$ is given by (4.64).

Hence, for complex $s$, the roundoff error of the evaluation (4.62) can be bounded by

$$\begin{aligned} \epsilon_{\text{roundoff}} &\leq \gamma_M \sum_{\ell=1}^{2M^2} |A_\ell||B_\ell| \\ &\leq 2CM^2\gamma_M \mathrm{e}^{\operatorname{Re} sd} \max_{\hat{s}\in\mathbb{S}^2} |M_{\alpha,\beta}(\hat{s})|, \end{aligned} \tag{4.65}$$

where $C$ is a constant coming from the use of the complex arithmetic and $d = \|x - x_\beta\| + \|y - y_\alpha\|$. In the BEM code, the low-frequency (occurring when $|s|d$ is smaller than a fixed value) roundoff error can be controlled by checking if

$$(L-1)\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\,\epsilon_m < \epsilon', \tag{4.66}$$

for a given $\epsilon' > 0$. In case when $\epsilon' \leq 1e-9$ we recommend checking

$$(L-1)\mathrm{e}^{\operatorname{Re} sd}\,|h_{L-1}(is\|c_{\alpha\beta}\|)|\,\epsilon_m < \epsilon'.$$

In the case of high decay, the cancellation errors can occur when performing multipole-to-multipole (local-to-local) transforms. First, recall that the cluster basis of the high-frequency fast multipole method is the matrix of the form

$$V_{kj} = \int_{\tau_\alpha} \mathrm{e}^{-s(y-y_\alpha,\hat{s}_k)}\phi_j(y)d\Gamma_y, \hat{s}_k \in \mathbb{S}^2.$$

In the course of the FMM algorithm this matrix is approximated by performing subsequent multipole-to-multipole transforms. These transforms involve spherical harmonic transforms, which includes many additions and subtractions of these numbers. Since the entries of the cluster basis for large $\operatorname{Re} s$ and $\|y-y_\alpha\|$ can vary greatly in magnitude, such transforms can potentially lead to cancellation errors. For a cluster with a bounding box of diameter $d$ such a cancellation error can be estimated by $\mathrm{e}^{\operatorname{Re} s\frac{d}{2}}\epsilon_m$. This seems to be pessimistic, since in numerous experiments we were not able to detect the influence of these errors on the result. This can be explained as follows. After the multipole-to-multipole transform the corresponding vector is multiplied by $M_{\alpha,\beta}(\hat{s}_\ell), \hat{s}_\ell \in \mathbb{S}^2$. This value is not larger than $L^2\,|h_{L-1}(is\|c_{\alpha\beta}\|)| \sim L^2\mathrm{e}^{-\operatorname{Re} s\|c_{\alpha\beta}\|}|s|^{-1}\|c_{\alpha\beta}\|^{-1}$ (see (4.40) and Remark 4.3.8 for a related discussion). During the local-to-local translation the result is multiplied by $\mathrm{e}^{-s(y_\alpha - y_{\alpha'},\hat{s}_k)}$ (that does not exceed $\mathrm{e}^{\operatorname{Re} s\frac{d}{2}}$), and hence the cancellation error of the final computation is bounded by

$$\epsilon_c \approx \mathrm{e}^{-\operatorname{Re} s(\|c_{\alpha\beta}\|-d)}\epsilon_m.$$

Recall that in this case the *absolute error* is of interest (c.f. (2.9)), hence, due to $\mathrm{e}^{-\operatorname{Re} s(\|c_{\alpha\beta}\|-d)} < 1$, the cancellation errors do not seem to cause problems.

## 4.5 On Choice of the Parameters of the HF FMM

In this section we comment on the choice of the parameters of the HF FMM, particularly, the length of the multipole expansion, the cluster tree and the admissibility condition.

**Cluster Tree**

The cluster tree has to be constructed so that the diameter of the bounding boxes of leaf clusters is $O\left(\frac{1}{|s|}\right)$. Typically, an octree is used, see [27]. In this work we employ a binary cluster tree, similarly to [22, 42, 41]. In [22] it was suggested to make a cluster a leaf if the number of degrees of freedom inside this cluster does not exceed some fixed $n_0$. We use such strategy as well (with $n_0 = 20$), however, with an additional correction: leaf clusters can be located only at the levels $\ell \geq \ell_0$, where $\ell_0$ is given a priori and increases logarithmically with $M$. The reason for the latter requirement is that it may happen that some of clusters with a few boundary elements occur at very coarse levels of the cluster tree. The size of bounding boxes of clusters located at one level of the cluster tree is the same (see Remark 3.2.3), and hence if such clusters were leaves, they would have large non-admissible neighbors and the size of the near-field would increase. This strategy is in agreement with [27]. Concerning the admissibility condition, we suggest the following. If the cluster tree is a binary tree, the admissibility condition should be chosen by Definition 2.2.5:

$$\|c_\tau - c_\sigma\| \geq \frac{\eta}{2}\left(d_\tau + d_\sigma\right),$$

with $\eta \geq \frac{3}{2}$. If the cluster tree is an octree, the clusters that do not touch each other can be considered admissible.

**Lengths of the Multipole Expansions**

To choose the length of multipole and local expansions, we suggest the following scheme. Let $\epsilon > 0$ be the desired accuracy. Let $r_\ell$ be the half-diameter of the bounding box of a cluster at the level $\ell$.

First, let us consider the multipole-to-local operator. Let an admissible block-cluster $b = (\tau_\alpha, \tau_\beta)$ be located at the level $\ell > 0$ of the block-cluster tree. Let the bounding box of the cluster $\tau_\alpha$ be centered at $y_\alpha$ and the bounding box of the cluster $\tau_\beta$ be centered at $x_\beta$. Additionally, $c_{\alpha\beta} = y_\alpha - x_\beta$. The choice of the length $n_b$ of the truncated expansion for the corresponding multipole-to-local translation operator, see also (3.26), can be determined by checking (c.f. Lemma 4.3.4, Section 4.3.2)

$$\left| h_0(is(\|c_{\alpha\beta}\| - 2r_\ell)) - \sum_{m=0}^{n_b-1} (2m+1) j_m(2isr_\ell) h_m(is\|c_{\alpha\beta}\|) \right| \leq \epsilon. \tag{4.67}$$

If

$$(L-1)\left|h_{L-1}(is\|c_{\alpha\beta}\|)\right| \epsilon_m > \epsilon,$$

the low-frequency breakdown happens, see (4.66), and we set formally $n_b = 0$. Such cluster is to be approximated with the help of $\mathcal{H}$-matrix techniques. Similarly, we approximate by zero matrix blocks for which (2.9) holds.

The value $n_b$ depends on $r_\ell$ and $\|c_{\alpha\beta}\|$ only. Since there is a fixed number of different $\|c_{\alpha\beta}\|$ per level, the check (4.67) can be performed once for each different $\|c_{\alpha\beta}\|$. The complexity of this operation is obviously sublinear.

Denoting the set of admissible block clusters located at the level $\ell$ by $\mathcal{L}_+^\ell$, let us introduce an auxiliary quantity:

$$n_\ell = \max_{b \in \mathcal{L}_+^\ell} n_b. \tag{4.68}$$

If $\operatorname{Re} s = 0$, the length of the multipole expansion at the level $\ell$ should be set to $2n_\ell^2$ (and the respective quadrature rule is to be chosen as $(w_k, \hat{q}_k)_{k=1}^{2n_\ell^2}$, see (3.15)). A similar strategy was suggested in [33, 27]. This strategy can be also applied for determining the length of the multipole expansion for the case when $\operatorname{Re} s \neq 0$.

As we discussed before, in some cases it may occur that at levels of the cluster tree where leaves are located there may be no clusters having admissible neighbors (and this can be the case for the applications that are of interest for us, e.g. Runge-Kutta convolution quadrature with the near-field reuse [9]). Therefore, we additionally perform the following checks.

The analysis in Section 4.3 shows that to control the multipole-to-multipole (local-to-local) errors, the length of the expansion has to satisfy conditions of Lemmas 4.3.4 and 4.3.7. Through the multipole-to-multipole (local-to-local) translation the error propagates to the coarser (finer) cluster tree levels. Though we do not present the analysis of the error after several consecutive multipole-to-multipole (local-to-local) translations, we suggest the following strategy.

Let us introduce an auxiliary quantity:

$$H_\ell = \sup_{b\in\mathcal{L}_+^\ell} \left|h_{n_b-1}(is\|c^b_{\alpha\beta}\|)\right| \le \sup_{b\in\mathcal{L}_+^\ell} \left|h_{n_\ell-1}(is\|c^b_{\alpha\beta}\|)\right|, \tag{4.69}$$

where the last inequality is obtained from (4.11). The value $\|c^b_{\alpha\beta}$ is the distance between the centres of bounding boxes of clusters forming an admissible block-cluster $b$. In practice this quantity (at least for large $n_\ell$) can be approximated by

$$\left|h_{n_\ell-1}(is\min_{b\in\mathcal{L}_+^\ell}\|c^b_{\alpha\beta}\|)\right|,$$

c.f. (4.7, 4.8, 4.10), as well as [25], where the dependence of $n_b$ on $\frac{\|c_{\alpha\beta}\|}{\operatorname{diam}\tau_\alpha}$ is discussed for $\operatorname{Re} s = 0$. In our the numerical experiments we employed this approximation at all levels of the block cluster tree.

Given a cluster $\tau_{\alpha,\ell}$ located at the level $\ell$, let $\tau_{\alpha,\ell-1},\ \tau_{\alpha,\ell-2},\ \ldots,\ \tau_{\alpha,k}$ be s.t.

$$\tau_{\alpha,j} \in \operatorname{sons}(\tau_{\alpha,j-1}), \qquad k < j \le \ell. \tag{4.70}$$

Here $k$ is the smallest level at which there is at least one admissible block cluster. Let $r_d^{j,\ell,\alpha}$ be the distance between the centers of the bounding boxes of clusters $\tau_{\alpha,j}$ and $\tau_{\alpha,\ell}$. Given levels $\ell,\ j$, the maximum of $r_d^{j,\ell,\alpha}$ over the pairs of clusters $\tau_{\alpha,\ell}, \tau_{\alpha,j}$ subject to (4.70) we denote by

$$\tilde{r}_d^{j,\ell} = \max_{(\tau_{\alpha,\ell},\tau_{\alpha,j})} r_d^{j,\ell,\alpha}.$$

This quantity can be computed in time not worse than linear (or even $O\left((\operatorname{height}(\mathcal{T}_\mathcal{I}))^2\right) = O\left(\log^2 M\right)$, where we used the fact that $\operatorname{height}(\mathcal{T}_\mathcal{I}) = O(\log M)$), due to the uniform partition of the domain, and hence it does not affect the complexity of the algorithm. Alternatively, it also can be estimated by $\tilde{r}_d^{j,\ell} \le r_j$.

In practice we look for $N_\ell$, such that for all $k \le j < \ell$ (i.e. for all higher levels with respect to the current level),

$$\begin{aligned} N_\ell n_j H_j \left|j_{N_\ell}(isr_\ell)\right| \mathrm{e}^{\operatorname{Re} s(r_j+\tilde{r}_d^{j,\ell})} &< \epsilon, \\ N_\ell n_j H_j \left|j_{N_\ell}(isr_\ell)\right|^2 \mathrm{e}^{2\operatorname{Re} s\tilde{r}_d^{j,\ell}} &< \epsilon, \end{aligned} \tag{4.71}$$

c.f. (4.60). Let $\mathcal{N}_\ell = \max(\max_{q\ge\ell} N_q,\ n_\ell)$ (c.f. Lemma 4.3.7, Lemma 4.3.4). Then we choose the length of the expansion at the level $\ell$ to be equal to $2\mathcal{N}_\ell^2$.

**Remark 4.5.1.** *Recall that $j_n(x)$ has real zeros. Hence in the case when the above inequalities hold true for some $N = N_0$, we suggest checking additionally if they are also valid for several consecutive values of $N$, namely $N_0 + 1, \ldots N_0 + q$ (with $q = 5$ for example).*

We would like to underline that in practice $\max_{q\ge\ell} N_q$ is rarely larger than $n_\ell$, and hence the choice $n_\ell$ as (4.68) may suffice.

All these checks are of the complexity not larger than linear, and hence do not affect the asymptotic complexity of the fast multipole algorithm.

**Remark 4.5.2.** *In this section we provided the heuristic to choose the length of the multipole expansion for $h_0(is\|x-y\|)$. A verbatim application of this strategy to the BEM approximation of the single layer potential (whose kernel is $\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|} = -\frac{s}{4\pi}h_0(is\|x-y\|)$) does not seem to lead to a noticeable increase of the error with $|s|$. Nonetheless for larger values of $|s|$ we suggest to adapt this heuristic for the case $\frac{\mathrm{e}^{-s\|x-y\|}}{4\pi\|x-y\|}$.*

# Chapter 5

# Numerical Comparison of $\mathcal{H}$-Matrix Techniques and the High-Frequency Fast Multipole Method

In this section we compare the performance of $\mathcal{H}$-matrices and the fast multipole method for the Helmholtz equation with decay. We present results of numerical experiments for $\mathcal{H}$-matrices and $\mathcal{H}^2$-matrices, which were constructed with the help of the expansions coming from the fast multipole method, approximating the Helmholtz boundary single-layer operator on the unit sphere and on the surface of a NASA almond-like domain (see Figure 5.1), both for complex and real wavenumbers. The domain and the mesh for it were generated with the help of Gmsh [45]. The length of this domain is 2.5, width 1 and height 0.32.

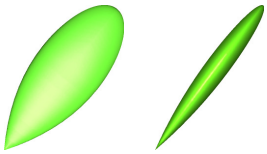

*Figure 5.1: The domain that we use in experiments. The domain is oriented parallel to $x$-axis; the incoming wave first hits the tip of the domain.*

We report the matrix construction times and matrix-vector multiplication times, as well as errors of the approximation. The matrix construction time for $\mathcal{H}^2$-matrices includes the time needed to construct an $\mathcal{H}$-matrix part, transfer matrices, multipole-to-local operators as well as the leaf cluster basis. Let us also remark that the time for the construction of the latter is the time to evaluate the integrals (3.17) using the precomputed quadrature points, weights and cluster centers (see the end of Section 3.2.1), and the precomputation itself is included into the matrix construction time as well (in practice it is negligible compared to the time needed to construct the leaf cluster basis).

Matrix-vector multiplication errors are measured as

$$\epsilon_r = \max_v \frac{\|\mathcal{M}v - Dv\|_2}{\|Dv\|_2},$$

based on results of 100 matrix-vector multiplications with random complex vectors with entries having real and imaginary part lying in $[-1,\ 1]$. Here $\mathcal{M}$ is an $\mathcal{H}-$matrix or an $\mathcal{H}^2-$matrix, which

was constructed with the help of the high-frequency fast multipole expansions. The matrix $\mathcal{M}$ is assembled with a given accuracy $\epsilon$. The matrix $D$ is a dense matrix (or a highly accurate $\mathcal{H}$-matrix) constructed using the quadrature rule (for the evaluation of Galerkin integrals) with the same number of quadrature points as for an $\mathcal{H}$-matrix. We also measure an average matrix-vector multiplication time.

Typically, the assembly time of $\mathcal{H}$-matrices is significantly larger than the assembly time of $\mathcal{H}^2$-matrices, see also [22], while the $\mathcal{H}$-matrix-vector multiplication time is much smaller, even for discretizations with approximately $10^5$ degrees of freedom. If many matrix-vector multiplications are needed, $\mathcal{H}$-matrices can outperform the HF FMM approximation. In this work we compute $N_{mv}$, the number of matrix-vector multiplications after which it makes sense to construct an $\mathcal{H}$-matrix based approximation. We use the formula

$$N_{mv} = \left\lceil \frac{T_c\left(\mathcal{H}\right) - T_c\left(\mathcal{H}^2\right)}{T_{mv}\left(\mathcal{H}^2\right) - T_{mv}\left(\mathcal{H}\right)} \right\rceil,$$

where $T_c$ is the matrix construction times and $T_{mv}$ is a time needed for one matrix-vector multiplication.

In the case when a matrix-vector multiplication with a given $\mathcal{H}^2$-matrix has to be done only a few times and it is necessary to store the matrix, the strategy of Section 3.2.1 can be employed. In this case it may be necessary to construct the leaf cluster basis as in Section 3.2.1 every time a matrix-vector multiplication is performed, and the value $T_{mv}$ thus has to include the time needed to construct the leaf cluster basis.

All experiments were done on the cluster of the Max Planck Institute for Mathematics in the Sciences, on a single processor of the hexa-core CPU Intel Xeon X5650 with 2.67 GHz. For the computation we used $\mathcal{H}$LIBpro library, see [70].

Spherical Bessel and Hankel functions were computed with the help of the Amos library [4].

Indeed, there is still a room for improving the FMM code and many FMM parameters, e.g. the quadrature order of near-singular and singular integrals, the number of quadrature points to construct the leaf cluster basis[1], or the accuracy of the $\mathcal{H}$-matrix approximations involved in spherical harmonic transforms. Additionally, one could make use of the strategy of [43] to control the HF FMM error.

The experiments are done with three accuracy settings: $\epsilon = 10^{-4}$, $\epsilon = 10^{-6}$ and $\epsilon = 10^{-9}$. The parameter $\epsilon$ used to determine the length of the multipole expansions, see Section 4.5, is set to one of these values. The relative accuracy for $\mathcal{H}$-matrices is set correspondingly to $10^{-3}$, $10^{-5}$ and $10^{-8}$ (such a choice is based on the observation that setting an ACA+ accuracy to some parameter often guarantees an accuracy several magnitudes higher).

In all the experiments with $\mathcal{H}^2$-matrices we used an admissibility condition according to Definition 2.2.5, with $\eta = \frac{3}{2}$ (for accuracies $\epsilon > \epsilon = 10^{-9}$) or $\eta = 2$ (for accuracies $10^{-9}$). For high accuracies, the latter choice appeared advantageous for the time of the construction of the cluster basis in the case of decay, which is due to the strategy adapted for choosing lengths of multipole expansions.

We would like to make an additional remark concerning the complexity of the algorithm. Further numerical experiments show that the complexity of the HF FMM is close to $O(M \log M)$, though the theoretical estimate is $O(M \log^2 M)$. The logarithmic factor in the latter comes from multipole-to-multipole and local-to-local transforms. Unlike multipole-to-local transforms that are performed many times per a cluster and total complexity of which is $O(M \log M)$, the M2M and L2L transforms are done only once per a cluster. Hence, the complexity of the FMM is $O(c_1 M \log M + c_2 M \log^2 M)$, where typically $c_2 \ll c_1$. Therefore, for a range of $M$ the time of the FMM accelerated matrix-vector multiplication scales closer to $O(M \log M)$ rather than $O(M \log^2 M)$. This can also be noticed in numerical experiments of [27].

[1]For the construction of the leaf cluster basis we suggest to use the following number of quadrature points in every triangular mesh element: for the accuracy $\epsilon = 10^{-4}$ $q = 2^2$, for $\epsilon = 10^{-6}$ $q = 3^2$ and for $\epsilon = 10^{-9}$ $q = 4^2$. Although for many algorithms this parameter does not play a significant role, it may be advantageous to adjust it in implementations of convolution quadrature that utilize the FMM (e.g to improve the performance of the algorithm of [9])

## 5.1 Real Wavenumber

In this section we present results of numerical experiments for the Helmholtz equation without decay. Our goal is to validate the correctness of the fast multipole implementation as well as to compare efficiency of the FMM and $\mathcal{H}$-matrices for this simple case. Similar experiments have been already performed in [22], but for the Burton-Miller integral formulation rather than for the Helmholtz single layer boundary operator, and using an optimized for the half-space problem HF FMM [23].

### 5.1.1 Accuracy $10^{-4}$

The results for the unit sphere are presented in Table 5.1 and for the elongated domain in Table 5.2.

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2048 | $-8i$ | 7.1 | 4(0.3) | 0.014 | 0.4 | 2.7e-4 | 3.2e-4 | 9 |
| 4232 | $-11.3i$ | 18.7 | 9.8 (0.8) | 0.04 | 0.84 | 3.2e-4 | 2.6e-4 | 12 |
| 8192 | $-16i$ | 41.6 | 21 (1.4) | 0.09 | 1.3 | 4e-4 | 2.4e-4 | 18 |
| 16200 | $-22.6i$ | 89 | 41 (2.6) | 0.2 | 3 | 4.6e-4 | 2.3e-4 | 18 |
| 32768 | $-32i$ | 225 | 90 (5.8) | 0.46 | 5.2 | 5e-4 | 5.6e-4 | 29 |
| 65448 | $-45.3i$ | 1481 | 414 (11) | 1 | 13.2 | 5.2e-4 | 5.8e-4 | 88 |
| 129970 | $-64.0i$ | 3254 | 847 (22.5) | 2.5 | 26.4 | - | - | 101 |

*Table 5.1: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-4}$ for the unit sphere. The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ is the number of matrix-vector multiplications when it makes sense to construct an $\mathcal{H}$-matrix instead of an FMM approximation of a BEM matrix. For the last experiment we did not construct a highly accurate matrix: it appeared to be too expensive. For the last two experiments we increased the Galerkin quadrature order: we observed the deterioration of accuracy due to the insufficient number of quadrature points.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2152 | $-10.6i$ | 8.8 | 6.4 (0.4) | 0.012 | 0.25 | 11 |
| 4096 | $-15i$ | 19.9 | 9 (0.7) | 0.027 | 0.9 | 13 |
| 8780 | $-21.2i$ | 51 | 23.8 (1.6) | 0.07 | 1.4 | 21 |
| 16072 | $-30i$ | 302 | 112.4 (3) | 0.15 | 2.4 | 85 |
| 64230 | $-60i$ | 1625 | 431.2 (11.5) | 0.87 | 8.7 | 153 |
| 144092 | $-90i$ | 4067 | 1047 (26.1) | 2.3 | 21.4 | 159 |

*Table 5.2: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-4}$, for the NASA almond-like domain. The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. The relative error of $\mathcal{H}$-matrix approximations does not exceed $6 \cdot 10^{-4}$ and of the HF FMM $5 \cdot 10^{-4}$. For the last three experiments we increased the Galerkin quadrature order.*

As expected, the number of matrix-vector multiplications for which $\mathcal{H}$-matrices outperform the HF FMM increases with the discretization size. For problems with $5 \cdot 10^4 - 10^5$ unknowns the $\mathcal{H}$-matrices should be constructed only if more than $100 - 150$ matrix-vector multiplications are needed. For smaller discretizations $\mathcal{H}$-matrices have to be constructed if more than 10-20 matrix vector multiplications are needed.

### 5.1.2 Accuracy $10^{-6}$

The results for the accuracy setup $\epsilon = 10^{-6}$ for the sphere are shown in Table 5.3 and for the elongated domain in Table 5.4. As before, the number of matrix-vector multiplications with which $\mathcal{H}$-matrices outperform the HF FMM increases with the discretization size. However, in this case the

HF FMM is more efficient than $\mathcal{H}$-matrices already for quite small discretizations ($10^4$ unknowns): for both domains if less than 40 matrix vector multiplications are needed, there is no sense to construct $\mathcal{H}$-matrices.

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2048 | $-8i$ | 22.6 | 14.5 (0.9) | 0.016 | 0.6 | 5e-6 | 1.5e-6 | 14 |
| 4232 | $-11.3i$ | 61.7 | 33.7 (2) | 0.046 | 1.1 | 3.9e-6 | 1.5e-6 | 27 |
| 8192 | $-16i$ | 150.7 | 59.5 (4.5) | 0.1 | 2.24 | 4.5e-6 | 1.1e-6 | 43 |
| 16200 | $-22.6i$ | 363 | 117 (7.9) | 0.27 | 4.3 | 4.9e-6 | 1.1e-6 | 62 |
| 32768 | $-32i$ | 890 | 241.7 (19.1) | 0.64 | 8.5 | 6.2e-6 | 1.2e-6 | 83 |
| 65448 | $-45.3i$ | 7612 | 491 (35.2) | 1.7 | 19.4 | 1.8e-5 | 2.4e-5 | 403 |
| 129970 | $-64.0i$ | - | 972 (79) | - | 42.3 | - | - | - |

*Table 5.3: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-6}$ (unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. The deterioration of the accuracy for large matrices is due to insufficient accuracy of Galerkin quadrature. $N_{mv}$ is the number of matrix-vector multiplications when it makes sense to construct an $\mathcal{H}$-matrix instead of an FMM matrix. For the largest experiment we did not construct $\mathcal{H}$-matrices.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2152 | $-10.6i$ | 33.1 | 23.4 (0.6) | 0.013 | 0.34 | 1e-5 | 2.3e-6 | 30 |
| 4096 | $-15i$ | 79 | 38.7 (1.6) | 0.034 | 0.73 | 3.4e-6 | 2.3e-6 | 58 |
| 8780 | $-21.2i$ | 199.3 | 87.5 (3.2) | 0.09 | 1.98 | 4.5e-6 | 2.6e-6 | 60 |
| 16072 | $-30i$ | 450 | 156 (6.3) | 0.21 | 3.4 | 7.5e-6 | 1.3e-5 | 93 |
| 64230 | $-60i$ | 4347 | 535 (27.7) | 1.46 | 15.2 | 1.8e-5 | 6.8e-5[2] | 278 |
| 144092 | $-90i$ | 16885.6 | 1276.3 (68.2) | 4.6 | 40.9 | - | - | 431 |

*Table 5.4: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-6}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation.*

The results in Table 5.5 demonstrate that the high-frequency fast multipole method is of almost linear complexity, while the complexity of $\mathcal{H}$-matrices scales somewhat worse, though better than predicted theoretically. This is connected to the fact that low-rank approximations constructed using $\mathcal{H}$-matrix techniques are close to optimal and take into account the geometry of a problem.

| $N_n$ | $s_n$ | $\log_2 \frac{T_c^n}{T_c^{n-1}}(\mathcal{H})$ | $\log_2 \frac{T_{mv}^n}{T_{mv}^{n-1}}(\mathcal{H})$ | $\log_2 \frac{T_c^n}{T_c^{n-1}}(\mathcal{H}^2)$ | $\log_2 \frac{T_{mv}^n}{T_{mv}^{n-1}}(\mathcal{H}^2)$ |
|---|---|---|---|---|---|
| 4232 | $-11.3i$ | 1.45 | 1.5 | 1.22 (1.15) | 0.9 |
| 8192 | $-16i$ | 1.29 | 1.12 | 0.82 (1.17) | 1 |
| 16200 | $-22.6i$ | 1.27 | 1.4 | 0.98 (0.82) | 0.94 |
| 32768 | $-32i$ | 1.29 | 1.25 | 1.05 (1.27) | 1 |
| 65448 | $-45.3i$ | 3.1 | 1.4 | 1.03 (0.89) | 1.2 |
| 129970 | $-64.0i$ | - | - | 1 (1.17) | 1.12 |

*Table 5.5: The rate of times for matrix assembly and times for matrix-vector multiplication for a current discretization and the twice coarser one for different techniques with the accuracy setup $\epsilon = 10^{-6}$.*

[2]The deterioration of the accuracy is due to the insufficient Galerkin quadrature order. After its increase, the error of $\mathcal{H}$-matrix approximation does not exceed $6e-6$ and of $\mathcal{H}^2$-matrix approximation $2.4e-6$. The construction times increase correspondingly to 5840 and 1157 seconds.

### 5.1.3 Accuracy $10^{-9}$

The results for the unit sphere are shown in Table 5.6 and for the NASA almond-like domain in Table 5.7. Besides showing that $\mathcal{H}$-matrices are less efficient than the HF FMM already for discretizations with $10^4$ unknowns in the case when 100 matrix-vector multiplications is needed, these results demonstrate that matrix-vector multiplication times of $\mathcal{H}^2$-matrices constructed with a high accuracy can be slightly smaller (or do not increase significantly) compared to that of $\mathcal{H}^2$-matrices constructed with a lower accuracy (c.f. Tables 5.6 and 5.3). This happens because of two reasons. First, the increase of the accuracy requires more matrix blocks to be approximated with the help of $\mathcal{H}$-matrix techniques (due to the low-frequency breakdown), and, as it can be seen from the numerical results, the matrix-vector multiplication times of $\mathcal{H}$-matrices are in practice much smaller compared to that of $\mathcal{H}^2$-matrices. Second, for large values of $|s|$ the increase of the length of the multipole expansion with the accuracy is relatively insignificant, c.f. [25, Table 3].

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2048 | $-8i$ | 155 | 94.6 (1.6) | 0.017 | 0.58 | 3.2e-9 | 1.2e-9 | 108 |
| 4232 | $-11.3i$ | 466 | 198.5 (3.8) | 0.06 | 1.1 | 4.1e-9 | 1.2e-9 | 258 |
| 8192 | $-16i$ | 1219 | 374.5 (7.8) | 0.16 | 2.8 | 4.7e-9 | 1.2e-9 | 320 |
| 16200 | $-22.6i$ | 3212 | 720 (14.6) | 0.42 | 5.6 | 5.2e-9 | 1.6e-9 | 482 |
| 32768 | $-32i$ | 8259 | 1570 (35) | 1 | 11.4 | 5.9e-9 | 1.8e-9 | 644 |
| 65448 | $-45.3i$ | 19270 | 3142 (68.7) | 2.7 | 25.3 | 1.9e-8 | 2e-9 | 714 |
| 129970 | $-64i$ | - | 6225 (140.3) | - | 57.2 | - | - | - |

*Table 5.6: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-9}$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. For the last two experiments we did not construct an $\mathcal{H}$-matrix approximation: for the given accuracy setting, it appears to be too expensive.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2152 | $-10.6i$ | 194 | 121 (1.1) | 0.018 | 0.45 | 167 |
| 4096 | $-15i$ | 502 | 291 (2.6) | 0.06 | 0.92 | 246 |
| 8780 | $-21.2i$ | 1518 | 634 (5.9) | 0.18 | 3 | 349 |
| 16072 | $-30i$ | 3556 | 1226.7 (10.6) | 0.41 | 4.2 | 615 |
| 64230 | $-60i$ | 22077 | 4342.5 (46.9) | 2 | 21.8 | 896 |
| 144092 | $-90i$ | 57817 | 9100 (118.3) | 5.4 | 57.1 | 943 |

*Table 5.7: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-9}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation.*
*For the last two experiments we did not construct a highly accurate approximation, but compared the results of FMM-accelerated to $\mathcal{H}$-matrix approximated matrix-vector products. In both cases the relative error did not exceed $2e-8$.*
*For the rest of experiments the relative errors of $\mathcal{H}^2$-matrices did not exceed $2e-9$ and for the $\mathcal{H}$-matrices they varied from $8e-7$ to $3e-8$.*

The results of our numerical experiments confirm the conclusion of [22]: if many matrix-vector multiplications are needed (for discretizations with $10^5$ unknowns more than 100-150 for $\epsilon = 10^{-4}$, 200-400 for $\epsilon = 10^{-6}$ and $> 700$ for $\epsilon = 10^{-9}$), $\mathcal{H}$-matrices are advantageous over $\mathcal{H}^2$-matrices.

## 5.2 Complex Wavenumber

In this section we present the results of the numerical experiments for the Helmholtz equation with decay. First, we fix $M$ and $|s|$ and study how the efficiency of $\mathcal{H}$- and $\mathcal{H}^2$-matrix approximations

changes with the argument of $s = |s|e^{i\phi}$, $\phi \in \left[-\frac{\pi}{2},\ 0\right]$.

### 5.2.1 Accuracy $10^{-4}$

The results of the experiments with this accuracy setting for the NASA almond-like domain are shown in Table 5.9 and for the unit sphere in Table 5.8. In both cases the number of matrix-vector multiplications when $\mathcal{H}$-matrices outperform the FMM reduces with $|\phi|$. If $|\phi| \leq \frac{\pi}{4}$, only 5-12 matrix vector multiplications are sufficient for $\mathcal{H}$-matrices to outperform the HF FMM.

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 89 | 41 (2.6) | 0.2 | 3 | 18 |
| $-\frac{6\pi}{14}$ | 109 | 53.5 (5.5) | 0.14 | 2.8 | 21 |
| $-\frac{5\pi}{14}$ | 81.1 | 57.4 (5.2) | 0.078 | 2.5 | 10 |
| $-\frac{4\pi}{14}$ | 80 | 54.9 (4.5) | 0.055 | 2.3 | 12 |
| $-\frac{3\pi}{14}$ | 68 | 58.9 (4.6) | 0.054 | 2.2 | 5 |
| $-\frac{2\pi}{14}$ | 61 | 48.4 (4.4) | 0.05 | 1.8 | 8 |
| $-\frac{\pi}{14}$ | 68 | 58.4 (4.5) | 0.05 | 1.4 | 8 |
| 0 | 52 | 46.2 (3.3) | 0.04 | 1.3 | 5 |

*Table 5.8: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-4}$, $M = 16200$, $|s| = 22.6$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. The relative error of $\mathcal{H}$-matrix approximations does not exceed $2.4 \cdot 10^{-4}$ and of the HF FMM $8 \cdot 10^{-5}$ in the case of decay and correspondingly $4.6 \cdot 10^{-4}$ and $2.3 \cdot 10^{-4}$ in the no-decay case.*

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 146.6 | 45 (3.3) | 0.16 | 2.4 | 46 |
| $-\frac{6\pi}{14}$ | 125 | 71 (6.6) | 0.12 | 2.6 | 22 |
| $-\frac{5\pi}{14}$ | 106 | 66.7 (6.4) | 0.09 | 2.2 | 19 |
| $-\frac{4\pi}{14}$ | 100 | 67.3 (6.2) | 0.075 | 2.2 | 16 |
| $-\frac{3\pi}{14}$ | 93 | 67.5 (6.2) | 0.07 | 2.2 | 12 |
| $-\frac{2\pi}{14}$ | 87 | 68.1 (5.9) | 0.06 | 2 | 10 |
| $-\frac{\pi}{14}$ | 83.3 | 68.9 (5.9) | 0.05 | 2.3 | 7 |
| 0 | 68.4 | 55.5 (4.2) | 0.05 | 2.1 | 7 |

*Table 5.9: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-4}$, $M = 16072$, $|s| = 30$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of the matrix-vector multiplications needed for the $\mathcal{H}$-matrix approximation to outperform the $\mathcal{H}^2$-approximation. We did not increase the Galerkin quadrature order with $M$ (unlike the experiments without the decay). The relative error of $\mathcal{H}$- and $\mathcal{H}^2$-matrix approximations does not exceed $3.5 \cdot 10^{-4}$ in the case with decay and $4 \cdot 10^{-3}$ in the no-decay case (with the error of FMM approximation not exceeding $3 \cdot 10^{-4}$).*

To check whether $N_{mv}$ changes with the discretization size, we conduct the experiment for $|\phi| = \frac{\pi}{4}$, different $|s|$ and $M$, as well as different ratios $\frac{|s|^2}{M}$. The results of the experiment for the unit sphere are shown in Table 5.10 and for the NASA almond-like domain in Tables 5.11 and 5.12. In all the cases the number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform the FMM does not seem to increase, remaining bounded by 9 for the unit sphere and by 18 for the NASA almond-like domain. The increased absolute value of decay (preserving $\phi = \text{const} = \frac{\pi}{4}$) favorably affects the complexity of approximation, as Tables 5.11 and 5.12 show.

Numerical results show the following effects of the presence of decay. First, the construction time of FMM matrix approximations increases significantly. This is due to the fact that the evaluation of the Helmholtz kernel with decay $\frac{e^{-s\|x-y\|}}{4\pi\|x-y\|}$, $s \in \mathbb{C}$, is more computationally expensive compared

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2048 | 8-8i | 8.3 | 4.65 (0.44) | 0.004 | 0.33 | 2.4e-5 | 2.4e-5 | 13 |
| 4232 | 11.3-11.3i | 18 | 11.8 (1.1) | 0.01 | 0.53 | 2.6e-5 | 2.5e-5 | 12 |
| 8192 | 16-16i | 32.2 | 25.7 (2) | 0.017 | 0.85 | 3.3e-5 | 3.3e-5 | 8 |
| 16200 | 22.6-22.6i | 60 | 51.4 (3.8) | 0.036 | 1.5 | 4.8e-5 | 4e-5 | 6 |
| 32768 | 32-32i | 122 | 112.3 (8) | 0.068 | 2.2 | 6.7e-5 | 5.6e-5 | 5 |
| 65448 | 45.3-45.3i | 254.7 | 215 (15.4) | 0.13 | 4.6 | 9.8e-5 | 8.6e-5 | 9 |
| 129970 | 64-64.0i | 510.7 | 456 (30.6) | 0.25 | 7.6 | 1.3e-4 | 1.2e-4 | 8 |

*Table 5.10: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-4}$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ is the number of matrix-vector multiplications when it makes sense to construct an $\mathcal{H}$-matrix instead of an $\mathcal{H}^2$-matrix.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2152 | $10.6 - 10.6i$ | 11.6 | 9.4 (0.8) | 0.007 | 0.2 | 12 |
| 4096 | $15 - 15i$ | 24.5 | 11.8 (1.3) | 0.012 | 0.6 | 22 |
| 8780 | $21.2 - 21.2i$ | 45.7 | 33 (2.9) | 0.03 | 1.12 | 12 |
| 16072 | $30 - 30i$ | 75.6 | 65.8 (4.9) | 0.045 | 1.61 | 7 |
| 64230 | $60 - 60i$ | 291.2 | 229.3 (18.8) | 0.16 | 5.45 | 12 |
| 144092 | $90 - 90i$ | 657.4 | 539 (43.6) | 0.35 | 12.2 | 10 |

*Table 5.11: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-4}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. For the last experiment we did not construct a highly accurate matrix approximation but compare the results of $\mathcal{H}$-matrix accelerated and FMM accelerated matrix-vector products (this relative error did not exceed $9.5e-4$). The relative error of the rest of accelerated matrix-vector products did not exceed $6e-4$.*

to the evaluation $\frac{e^{i\kappa\|x-y\|}}{4\pi\|x-y\|}$, $\kappa \in \mathbb{R}$. Similar arguments apply to the leaf cluster basis. In the latter sections we will see that this can affect $\mathcal{H}$-matrices as well, though not being always the case: the presence of sufficiently large decay can also reduce the time of the $\mathcal{H}$-matrix construction (due to the drastic decrease of ranks of $\mathcal{H}$-matrix blocks).

The results in Tables 5.1, 5.8 and 5.10 show that matrix-vector multiplication times of the HF FMM in the presence of decay are smaller than in the case of no-decay. This can be explained by the reduction of the length of multipole expansions. Similarly, the matrix-vector multiplication costs for $\mathcal{H}$-matrices are reduced compared to the no-decay case.

In cases of prevailing decay (i.e. $\operatorname{Re} s \geq \operatorname{Im} s$), the memory (RAM) requirements for $\mathcal{H}$-matrices can be several (5-6) times smaller than that of $\mathcal{H}^2$-matrices, even for large discretizations. This is connected to the fact that the cluster basis should be precomputed and stored in memory (RAM) and the respective HF FMM multipole expansions are relatively long, c.f. [27, Table 10]. In a nutshell, the length of the multipole expansion for a cluster at the level $\ell$ of the cluster tree with a bounding box of diameter $d$ can be determined by examining the convergence of the Gegenbauer's series. Namely, it is sufficient to find $n$ s.t. $|j_n(isd)h_n(isc)| < \epsilon$, where $c$ is the distance between the centers of the bounding boxes of two closest admissible clusters at the level $\ell$ (see e.g. [33] or [27] where a similar criterion is used). For small clusters the actual values produced by such criterion may be large ($n \gg |s|d$) due to the superexponential growth of spherical Hankel functions of a complex argument, c.f. [27]. This effect is enhanced by the fact that the length of the multipole expansion has to be chosen as $O\left(n^2\right)$. We did not observe this behavior for large discretizations in the no-decay case.

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2152 | $5-5i$ | 11.1 | 9.1 (0.5) | 0.007 | 0.26 | 8 |
| 4096 | $7.1-7.1i$ | 25.9 | 16.5 (1.1) | 0.016 | 0.8 | 12 |
| 8780 | $10-10i$ | 53.5 | 35.3 (2.8) | 0.035 | 1.35 | 14 |
| 16072 | $14-14i$ | 93.9 | 65.2 (5.2) | 0.066 | 2.27 | 14 |
| 64230 | $28-28i$ | 375.3 | 240.3 (19.9) | 0.25 | 7.9 | 18 |
| 144092 | $42-42i$ | 849 | 565.8 (43.4) | 0.53 | 17.73 | 17 |

*Table 5.12: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-4}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. For the last two experiments we did not construct a highly accurate matrix approximation but compare the results of $\mathcal{H}$-matrix accelerated and FMM accelerated matrix-vector products (this relative error did not exceed $7.2e-4$ for the experiment with 64230 unknowns and $1.1e-3$ for the experiment with 144092 unknowns).*
*For other discretizations the relative error of $\mathcal{H}$-matrices and the FMM did not exceed $3.3e-4$.*

### 5.2.2 Accuracy $10^{-6}$

Similarly to the previous section, we study the effect of the change of the argument of $s$ on the efficiency of different techniques with the accuracy setting $\epsilon = 10^{-6}$. These results are shown in Table 5.14 for the thin, elongated domain and in 5.13 for the unit sphere.

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 363 | 117 (7.8) | 0.27 | 4.3 | 62 |
| $-\frac{6\pi}{14}$ | 425.5 | 173 (19.4) | 0.2 | 4.5 | 59 |
| $-\frac{5\pi}{14}$ | 323.4 | 166.9 (18.3) | 0.11 | 4.5 | 36 |
| $-\frac{4\pi}{14}$ | 268.6 | 156 (16.6) | 0.086 | 4.4 | 27 |
| $-\frac{3\pi}{14}$ | 236.2 | 148 (15.2) | 0.072 | 4.8 | 19 |
| $-\frac{2\pi}{14}$ | 221.1 | 146.3 (14.8) | 0.066 | 4 | 20 |
| $-\frac{\pi}{14}$ | 215.4 | 145.8 (13.9) | 0.065 | 3.8 | 19 |
| 0 | 174 | 118.2 (10.2) | 0.06 | 3.8 | 15 |

*Table 5.13: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-6}$, $M = 16200$, $|s| = 22.6$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. In all the experiments the relative error of $\mathcal{H}$- and $\mathcal{H}^2$-approximations did not exceed $1.6 \cdot 10^{-6}$.*

As before, with decreasing $|\phi|$ the efficiency of $\mathcal{H}$-matrix approximation is improved compared to $\mathcal{H}^2$-matrix approximation. In the case of the unit sphere $\mathcal{H}$-matrix approximation is more efficient than the HF FMM if $|\phi| \leq \frac{\pi}{4}$ and more than 20 matrix-vector multiplications are needed. In the case of the NASA almond-like domain $\mathcal{H}^2$-matrix approximation outperforms $\mathcal{H}$-matrix approximation for the whole range of $\phi$ if less than 25 matrix-vector multiplications are needed.

To check how $N_{mv}$ changes with the discretization size, we assemble approximations of the matrices $\boldsymbol{V}(s)$ for $s = |s|\mathrm{e}^{-\frac{\pi}{4}i}$. The results of these experiments are shown in Tables 5.15 for the unit sphere and 5.16 for the thin NASA almond-like domain.

Numerical results in Table 5.15 suggest that, similarly to the case of the Helmholtz equation without decay, the assembly time of $\mathcal{H}$-matrices is larger than that of $\mathcal{H}^2$-matrices. However, if in the case of purely real wavenumber for the matrices of size $10^4 - 10^5$ the difference varies from 1.5 to 4 (and even more) times, in the case of prevailing decay (i.e. for $s = |s|\mathrm{e}^{i\alpha}$, $\alpha \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$) the difference is not that significant. In our experiments it never exceeded 2 times for matrices of size $10^4 - 10^5$. Therefore, $\mathcal{H}$-matrix approximations in this case are more efficient than $\mathcal{H}^2$-matrix approximations, even if a small number of matrix-vector multiplications is needed.

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 450 | 156 (6.3) | 0.21 | 3.4 | 93 |
| $-\frac{6\pi}{14}$ | 512.3 | 222 (13.6) | 0.13 | 3.5 | 87 |
| $-\frac{5\pi}{14}$ | 426.3 | 218.6 (13.3) | 0.11 | 3.23 | 67 |
| $-\frac{4\pi}{14}$ | 373.4 | 214.1 (12.7) | 0.09 | 3 | 55 |
| $-\frac{3\pi}{14}$ | 334 | 201 (13.7) | 0.08 | 3.2 | 43 |
| $-\frac{2\pi}{14}$ | 312.5 | 207.4 (13) | 0.077 | 2.9 | 38 |
| $-\frac{\pi}{14}$ | 301 | 193 (15.9) | 0.076 | 3.2 | 35 |
| 0 | 245.2 | 169.1 (11.7) | 0.07 | 3.1 | 26 |

*Table 5.14: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-6}$, $M = 16072$, $|s| = 30$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. In the case of non-zero decay the relative error of $\mathcal{H}$-matrix approximations does not exceed $4.3e-6$ and of the HF FMM $3e-6$.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2048 | 8-8i | 29.8 | 17.7 (2.1) | 0.0075 | 0.75 | 17 |
| 4232 | 11.3-11.3i | 73.3 | 47.9 (5.1) | 0.015 | 0.92 | 29 |
| 8192 | 16-16i | 119.5 | 91 (11.2) | 0.03 | 2.1 | 14 |
| 16200 | 22.6-22.6i | 240.8 | 171.8 (18.1) | 0.056 | 3.24 | 22 |
| 32768 | 32-32i | 458.7 | 331 (45.6) | 0.11 | 6.5 | 20 |
| 65448 | 45.3-45.3i | 923.7 | 697.1 (78.2) | 0.22 | 10.2 | 23 |
| 129970 | 64-64.0i | 1856.9 | 1375 (177.2) | 0.44 | 21.9 | 23 |

*Table 5.15: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-6}$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. In all experiments the computed relative error of approximations did not exceed $2.5e-6$.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2152 | $10.6-10.6i$ | 34.6 | 27.2 (1.6) | 0.007 | 0.43 | 49 |
| 4096 | $15-15i$ | 72.9 | 44.2 (4.2) | 0.019 | 1.2 | 25 |
| 8780 | $21.2-21.2i$ | 173.6 | 102.3 (9) | 0.038 | 2.35 | 31 |
| 16072 | $30-30i$ | 300.3 | 191.7 (19.9) | 0.075 | 3.74 | 30 |
| 64230 | $60-60i$ | 1167 | 716 (81.1) | 0.25 | 12.95 | 36 |
| 144092 | $90-90i$ | 2546 | 1619 (163.4) | 0.56 | 31.1 | 31 |

*Table 5.16: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-6}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. In all experiments but the last one the relative error of accelerated matrix-vector products did not exceed $7.3e-6$. In the last experiment the relative error increased to $3.5e-5$, which is connected to the insufficient precision of Galerkin quadrature.*

### 5.2.3 Accuracy $10^{-9}$

The results of the computation with the accuracy setup $\epsilon = 10^{-9}$ for the sphere are shown in Tables 5.17 and 5.19, while for the elongated domain in Tables 5.18 and 5.20. When a higher accuracy is required $\mathcal{H}$-matrices outperform $\mathcal{H}^2$-matrices for larger problems only when more than 200-400 matrix-vector multiplication is needed. This remains true even in the case of high decay, though indeed, with decreasing $|\phi|$ in $s = |s|e^{i\phi}$ the performance of $\mathcal{H}$-matrices improves compared to the performance of the HF FMM.

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 3212 | 720 (14.4) | 0.42 | 5.6 | 482 |
| $-\frac{6\pi}{14}$ | 4558 | 1290 (30.5) | 0.27 | 4.8 | 722 |
| $-\frac{5\pi}{14}$ | 3558 | 1242 (28.3) | 0.18 | 5.1 | 471 |
| $-\frac{4\pi}{14}$ | 2774 | 1145 (28.2) | 0.13 | 5.1 | 328 |
| $-\frac{3\pi}{14}$ | 2379 | 1082 (27.7) | 0.1 | 4.8 | 276 |
| $-\frac{2\pi}{14}$ | 2189 | 1107 (26.1) | 0.1 | 4.12 | 270 |
| $-\frac{\pi}{14}$ | 1920 | 1047 (29) | 0.09 | 4.1 | 218 |
| 0 | 1592 | 796 (20.7) | 0.09 | 4.6 | 177 |

*Table 5.17: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-9}$, $M = 16200$, $|s| = 22.6$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation. In all the experiments the relative error of $\mathcal{H}$-matrix approximations did not exceed $5.2e-9$ and of $\mathcal{H}^2$-approximations $1.6e-9$.*

| $\phi$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|
| $-\frac{\pi}{2}$ | 3556 | 1226.7 (10.6) | 0.41 | 4.2 | 615 |
| $-\frac{6\pi}{14}$ | 5074 | 2007 (22.9) | 0.3 | 4.24 | 779 |
| $-\frac{5\pi}{14}$ | 4270 | 1945 (22.2) | 0.25 | 4.23 | 585 |
| $-\frac{4\pi}{14}$ | 3763 | 1730 (26.9) | 0.22 | 3.7 | 585 |
| $-\frac{3\pi}{14}$ | 3408 | 1688 (26.8) | 0.2 | 3.9 | 465 |
| $-\frac{2\pi}{14}$ | 3115 | 1601 (26.7) | 0.17 | 3.9 | 406 |
| $-\frac{\pi}{14}$ | 2946 | 1567 (26.9) | 0.16 | 3.3 | 440 |
| 0 | 2392 | 1274 (20) | 0.17 | 3.5 | 336 |

*Table 5.18: Construction times $T_c$ and matrix-vector multiplication times $T_{mv}$ for the accuracy setup $\epsilon = 10^{-9}$, $M = 16072$, $|s| = 30$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for number of the matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation . In all the experiments the relative error of $\mathcal{H}$-matrix approximations did not exceed $5.3 \cdot 10^{-8}$ and of $\mathcal{H}^2$-approximations $3.5 \cdot 10^{-9}$.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|
| 2048 | 8-8i | 189.3 | 84.7 (3.8) | 0.011 | 0.8 | 152 |
| 4232 | 11.3-11.3i | 466 | 201 (9.6) | 0.024 | 1.5 | 180 |
| 8192 | 16-16i | 921.3 | 496 (17.3) | 0.044 | 2.7 | 161 |
| 16200 | 22.6-22.6i | 1824 | 868 (36) | 0.091 | 4.8 | 204 |
| 32768 | 32-32i | 1957 | 3800 (69) | 0.18 | 7.2 | 263 |
| 65448 | 45.3-45.3i | 7469 | 4300 (133.5) | 0.35 | 13.4 | 243 |
| 129970 | 64-64.0i | 15151 | 8503 (282) | 0.72 | 26.6 | 257 |

*Table 5.19: Times of matrix construction and matrix-vector multiplication for the accuracy setup $\epsilon = 10^{-9}$ (the unit sphere). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation.*
*In all cases the relative error of $\mathcal{H}^2$-matrices did not exceed $1.2e-9$, and of $\mathcal{H}$-matrices varied from $2e-9$ to $5.5e-8$.*

| $M$ | $s$ | $T_c(\mathcal{H})$ | $T_c(\mathcal{H}^2)$ | $T_{mv}(\mathcal{H})$ | $T_{mv}(\mathcal{H}^2)$ | $\epsilon_r(\mathcal{H})$ | $\epsilon_r(\mathcal{H}^2)$ | $N_{mv}$ |
|---|---|---|---|---|---|---|---|---|
| 2152 | $10.6-10.6i$ | 237 | 172 (3) | 0.015 | 0.49 | 1.4e-6 | 8.9e-10 | 137 |
| 4096 | $15-15i$ | 555.5 | 291.4 (6.7) | 0.03 | 1.2 | 5e-8 | 8.2e-9 | 226 |
| 8780 | $21.2-21.2i$ | 1468 | 696 (18.3) | 0.09 | 2.9 | 1.1e-7 | 1.8e-9 | 282 |
| 16072 | $30-30i$ | 2792 | 1282.6 (32.8) | 0.15 | 4 | 5.3e-8 | 2.6e-9 | 393 |
| 64230 | $60-60i$ | 11250 | 4800 (125.9) | 0.5 | 16.2 | 3.3e-8 | 4.7e-9 | 411 |
| 144092 | $90-90i$ | 24217 | 12052 (300) | 1.08 | 35.6 | 4.3e-8 | 7.7e-9 | 353 |

*Table 5.20: Construction times $T_c$, matrix-vector multiplication times $T_{mv}$ and computed relative errors for the accuracy setup $\epsilon = 10^{-9}$ (the elongated domain). The times are given in seconds. In brackets the time to construct the leaf cluster basis is shown. $N_{mv}$ stands for the number of matrix-vector multiplications needed for $\mathcal{H}$-matrix approximation to outperform $\mathcal{H}^2$-approximation.*
*The decrease in the accuracy of $\mathcal{H}$-matrices is likely to be connected to the insufficient Galerkin quadrature order.*

# Conclusions

In this work we provided the description of the HF FMM with important technical details, as well as analyzed the error of the multilevel FMM approximation. Additionally, we reviewed known facts about $\mathcal{H}$-matrix techniques.

We have shown that the presence of decay allows to decrease the ranks of multipole transforms, thus improving the performance of the HF FMM. In this regime $\mathcal{H}$-matrix approximation is of almost linear complexity, namely $O(M \log^a M)$, $a \geq 0$. In practice, the choice between $\mathcal{H}$-matrices and the HF FMM is quite difficult. The efficiency of both methods depends on the geometry of the domain, accuracy and discretization size.

In [22] it was shown that for the no-decay case the time to construct an $\mathcal{H}$-matrix is significantly larger than the time to construct an HF FMM matrix, while the opposite holds for the matrix-vector multiplication times. Hence, $\mathcal{H}$-matrices have to be constructed only when 300-600 matrix-vector multiplications are needed. Our experiments show that the presence of decay allows to decrease the $\mathcal{H}$-matrix construction times, and thus they perform in practice better than the HF FMM even if a small number of matrix-vector multiplications is needed. Based on the results of the previous sections, we suggest the following simple heuristic to perform the choice between $\mathcal{H}$-matrices and the high-frequency fast multipole method.

First, $\mathcal{H}$-matrices have to be constructed when at the first few (2-3) levels of the admissible block-cluster tree there are admissible block-clusters that cannot be approximated by FMM expansions because of the low-frequency breakdown. This is a clear indicator that $|\kappa| \operatorname{diam} \Omega$ is small enough, i.e. this is the case of the low-frequency regime.

Next, depending on the accuracy and the angle $\phi$ in $s = |s|\mathrm{e}^{i\phi}$, we suggest that $\mathcal{H}$-matrices are constructed only when more than $N_{mv}$ matrix-vector multiplications are needed, where $N_{mv}$ is given in Table 5.21. These values are valid for discretizations with $5 \cdot 10^4 - 10^5$ unknowns. For moderate

| | $1e-4$ | $1e-6$ | $1e-9$ |
|---|---|---|---|
| $\lvert\phi\rvert \geq \frac{\pi}{4}$ | 100-150 | 250-400 | 700-900 |
| $\lvert\phi\rvert \leq \frac{\pi}{4}$ | 8-12 | 20-30 | 90-400 |

*Table 5.21: The number of matrix-vector multiplications depending on the desired accuracy $\epsilon$ and the argument $\phi$ in $s = |s|\mathrm{e}^{i\phi}$ for which the use of $\mathcal{H}$-matrix based techniques would be advantageous compared to HF FMM matrices.*

and small accuracies ($\epsilon \geq 10^{-6}$) and small decay in practice $\mathcal{H}$-matrices perform worse than the HF FMM, while in the case of prevailing decay they are advantageous over the HF FMM, as Table 5.21 shows. For high accuracies it makes sense to use the HF FMM for the whole range of $\phi$.

Nevertheless, as noted in [22] for non-decay case, the application of $\mathcal{H}$-matrices to the solution of boundary-value problems for the Helmholtz equation may still be advantageous, due to the availability of efficient $\mathcal{H}$-matrix based preconditioners, which require a corresponding $\mathcal{H}$-matrix to be constructed in advance.

# Appendix A

# Complexity of the HF FMM

Let us analyze the complexity of the HF FMM when applied to the approximation of

$$f_\ell = \sum_{k=1}^{M} h_0(is\|x_\ell - x_k\|)q_k, \qquad x_\ell \in \Gamma,\ q_\ell \in \mathbb{C},\ \ell = 1, \dots, M.$$

Let us assume that the diameter of the domain is 1, fix an admissible block-cluster tree and introduce the following quantities. Let $\mathcal{C}^j_{M2M}(\mathcal{C}^j_{L2L})$ be the complexity of one multipole-to-multipole (local-to-local) transform on the level $j$, $\mathcal{C}^j_{M2L}$ be the complexity of one multipole-to-local transform at the level $j$. Let us assume that an octree based partitioning of the domain is employed; however, since we consider a 2-dimensional surface $\Gamma$, at each level $j$ of a cluster tree there are (in average) $4^j$ non-empty clusters (see Remark 2.2.6). Importantly, the diameter of the bounding box of a cluster at the level $j$ can be bounded by $2^{-j}$. The cluster basis for a leaf cluster $\tau$ is a matrix of size $(N_\tau \times \#\tau)$, where $N_\tau$ is the length of the multipole expansion. If for all leaf clusters $\sigma$

$$\operatorname{diam}(\Omega_\sigma|s|) \le \text{const},$$

then $N_\sigma$ can be bounded uniformly (for all leaf clusters $\sigma$) by constant. Hence, the number of levels of the cluster tree has to be chosen so that this condition holds true. More precisely, the depth of the cluster tree $L$ should satisfy

$$2^{-L}|s| \le \text{const},$$

i.e. $L \ge \log|s| + c$, for some $c \in \mathbb{R}$. Hence, forming the multipole expansion (performing the forward transformation involving the leaf cluster basis, see Section 2.5) requires $O(M)$ steps.

We additionally assume that the number of admissible neighbors of a cluster at level $\ell$ can be bounded by a constant independent on $\ell, M$, i.e. the sparsity constant of the block-cluster tree $C_{sp} \le \text{const}$, see Section 2.3. Then the complexity of the FMM is bounded by the sum of the complexities of all the transforms over all the levels of the block-cluster tree:

$$O(M) + O\left(\sum_{k=0}^{L}\left(\mathcal{C}^k_{M2M}4^k + \mathcal{C}^k_{L2L}4^k\right)\right) + O\left(\sum_{k=0}^{L}\mathcal{C}^k_{M2L}4^k\right).$$

In the case when $s = |s|\mathrm{e}^{i\alpha}$, $\alpha = \text{const}$ and $|s| \to +\infty$, the complexity of the transforms can be bounded by a (possibly large) constant. Hence, the complexity of the HF FMM is

$$O\left(M + \sum_{k=0}^{\log|s|+c} 4^k\right) = O(M),$$

under assumption $M = O(|s|^2)$.

Let us remark that in the no-decay case

$$C^k_{M2M} = O(4^{-k}|s|^2\log|2^{-k}s|), \qquad C^k_{L2L} = O(4^{-k}|s|^2\log|2^{-k}s|), \qquad C^k_{M2L} = O(4^{-k}|s|^2),$$

and hence the total complexity of the HF FMM is $O(M\log^2 M)$.

# Bibliography


[1] *NIST digital library of mathematical functions. http://dlmf.nist.gov/, release 1.0.5 of 2012-10-01.*

[2] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of National Bureau of Standards Applied Mathematics Series, 1972.

[3] S. Amini and A. Profit, *Multi-level fast multipole solution of the scattering problem*, Engineering Analysis with Boundary Elements, 27 (2003), pp. 547 – 564.

[4] D. E. Amos, *Algorithm 644: a portable package for Bessel functions of a complex argument and nonnegative order*, ACM Trans. Math. Software, 12 (1986), pp. 265–273.

[5] A. Bamberger and T. Ha-Duong, *Formulation variationnelle espace-temps pour le calcul par potentiel retardé de la diffraction d'une onde acoustique*, Mathematical Methods in the Applied Sciences, 8 (1986), pp. 405–435.

[6] L. Banjai, *A boundary element method for the solution of Helmholtz problems for a large range of complex wavenumbers.* Presentation at the 23rd Annual GAMM Seminar Leipzig 'Integral Equation Methods for High-Frequency Scattering Problems'.

[7] L. Banjai, *Multistep and multistage convolution quadrature for the wave equation: Algorithms and experiments*, SIAM Journal on Scientific Computing, 32 (2010), pp. 2964–2994.

[8] L. Banjai and W. Hackbusch, *Hierarchical matrix techniques for low- and high-frequency Helmholtz problems*, IMA Journal of Numerical Analysis, 28 (2008), pp. 46–79.

[9] L. Banjai and M. Kachanovska, *Fast convolution quadrature for the wave equation in three dimensions.* Submitted.

[10] R. W. Barnard, K. Pearce, and L. Schovanec, *Inequalities for the perimeter of an ellipse*, J. Math. Anal. Appl., 260 (2001), pp. 295–306.

[11] J. Barnes and P. Hut, *A hierarchical $o(n \log n)$ force calculation algorithm*, Nature, 324 (1986).

[12] M. Bebendorf, *Approximation of boundary element matrices*, Numer. Math., 86 (2000), pp. 565–589.

[13] ———, *Hierarchical LU decomposition-based preconditioners for BEM*, Computing, 74 (2005), pp. 225–247.

[14] M. Bebendorf, *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, vol. 63 of Lecture Notes in Computational Science and Engineering (LNCSE), Springer-Verlag, 2008. ISBN 978-3-540-77146-3.

[15] M. Bebendorf and R. Grzhibovskis, *Accelerating Galerkin BEM for linear elasticity using adaptive cross approximation*, Math. Methods Appl. Sci., 29 (2006), pp. 1721–1747.

[16] M. Bebendorf and W. Hackbusch, *Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with $L^\infty$-coefficients*, Numer. Math., 95 (2003), pp. 1–28.

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

[18] S. Börm, *Efficient numerical methods for non-local operators*, vol. 14 of EMS Tracts in Mathematics, European Mathematical Society (EMS), Zürich, 2010.

[19] S. Börm and L. Grasedyck, *Hybrid cross approximation of integral operators*, Numer. Math., 101 (2005), pp. 221–249.

[20] S. Börm and W. Hackbusch, *Data-sparse approximation by adaptive $\mathcal{H}^2$-matrices*, Computing, 69 (2002), pp. 1–35.

[21] J. Breuer, *Schnelle Randelementmethoden zur Simulation von elektrischen Wirbelstromfeldern sowie ihrer Wärmeproduktion und Kühlung*, PhD thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, 2005.

[22] D. Brunner, M. Junge, P. Rapp, M. Bebendorf, and L. Gaul, *Comparison of the fast multipole method with hierarchical matrices for the Helmholtz-BEM*, CMES: Computer Modeling in Engineering & Sciences, 58 (2010), pp. 131–160.

[23] D. Brunner, G. Of, M. Junge, O. Steinbach, and L. Gaul, *A fast BE-FE coupling scheme for partly immersed bodies*, Internat. J. Numer. Methods Engrg., 81 (2010), pp. 28–47.

[24] Q. Carayol and F. Collino, *Error estimates in the fast multipole method for scattering problems. I. Truncation of the Jacobi-Anger series*, M2AN Math. Model. Numer. Anal., 38 (2004), pp. 371–394.

[25] ———, *Error estimates in the fast multipole method for scattering problems. II. Truncation of the Gegenbauer series*, M2AN Math. Model. Numer. Anal., 39 (2005), pp. 183–221.

[26] C. Cecka and E. Darve, *Fourier-Based Fast Multipole Method for the Helmholtz Equation*, SIAM J. Sci. Comput., 35 (2013), pp. A79–A103.

[27] H. Cheng, W. Y. Crutchfield, Z. Gimbutas, L. F. Greengard, J. F. Ethridge, J. Huang, V. Rokhlin, N. Yarvin, and J. Zhao, *A wideband fast multipole method for the Helmholtz equation in three dimensions*, J. Comput. Phys., 216 (2006), pp. 300–325.

[28] W. C. Chew, J.-M. Jin, E. Michielssen, and J. Song, eds., *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, 2001.

[29] C. W. Clenshaw, *A note on the summation of Chebyshev series*, Math. Tables Aids Comput., 9 (1955), pp. 118–120.

[30] R. Coifman, V. Rokhlin, and S. Wandzura, *The fast multipole method for the wave equation: a pedestrian prescription*, Antennas and Propagation Magazine, IEEE, 35 (1993), pp. 7 –12.

[31] E. Darrigrand, *Coupling of fast multipole method and microlocal discretization for the 3-D Helmholtz equation*, J. Comput. Phys., 181 (2002), pp. 126–154.

[32] E. Darve, *The fast multipole method I: error analysis and asymptotic complexity*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 98–128.

[33] E. Darve, *The fast multipole method: numerical implementation*, Journal of Computational Physics, 160 (2000), pp. 195 – 240.

[34] E. Darve and P. Havé, *A fast multipole method for Maxwell equations stable at all frequencies*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 362 (2004), pp. 603–628.

[35] P. J. Davis, *Interpolation and approximation*, Dover Publications Inc., New York, 1975. Republication, with minor corrections, of the 1963 original, with a new preface and bibliography.

[36] B. Engquist and L. Ying, *Fast directional multilevel algorithms for oscillatory kernels*, SIAM J. Sci. Comput., 29 (2007), pp. 1710–1737 (electronic).

[37] ———, *A fast directional algorithm for high frequency acoustic scattering in two dimensions*, Commun. Math. Sci., 7 (2009), pp. 327–345.

[38] M. A. Epton and B. Dembart, *Multipole translation theory for the three-dimensional Laplace and Helmholtz equations*, SIAM J. Sci. Comput., 16 (1995), pp. 865–897.

[39] S. Erichsen and S. A. Sauter, *Efficient automatic quadrature in 3-d Galerkin BEM*, Comput. Methods Appl. Mech. Engrg., 157 (1998), pp. 215–224. Seventh Conference on Numerical Methods and Computational Mechanics in Science and Engineering (NMCM 96) (Miskolc).

[40] W. N. Everitt and D. S. Jones, *On an integral inequality*, Proc. Roy. Soc. London Ser. A, 357 (1977), pp. 271–288.

[41] M. Fischer, *The Fast Multipole Boundary Element Method and its Application to Structure-Acoustic Field Interaction*, PhD thesis, University of Stuttgart, 2004.

[42] M. Fischer, H. Perfahl, and L. Gaul, *Approximate inverse preconditioning for the fast multipole BEM in acoustics*, Comput. Vis. Sci., 8 (2005), pp. 169–177.

[43] A. Frangi and M. Bonnet, *On the application of the fast multipole method to Helmholtz-like problems with complex wavenumber*, CMES. Computer Modeling in Engineering & Sciences, 58 (2010), pp. 271–295.

[44] N. Geng, A. Sullivan, and L. Carin, *Fast multipole method for scattering from an arbitrary PEC target above or buried in a lossy half space*, IEEE Trans. Antennas and Propagation, 49 (2001), pp. 740–748.

[45] C. Geuzaine and J.-F. Remacle, *Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering, 79 (2009), pp. 1309–1331.

[46] S. Goreinov, *Mosaic-skeleton approximations of matrices, generated by asymptotically smooth and oscillatory kernels*, in Matrix Methods and Computations, E. Tyrtyshnikov, ed., INM RAS, Moscow, 1999, pp. 42–76. (in Russian).

[47] S. Goreinov, E. Tyrtyshnikov, and N. Zamarashkin, *A theory of pseudoskeleton approximations*, Linear Algebra Appl., 261 (1997), pp. 1–21.

[48] L. Grasedyck, *Adaptive recompression of $\mathcal{H}$-matrices for BEM*, Computing, 74 (2005), pp. 205–223.

[49] L. Grasedyck and W. Hackbusch, *Construction and arithmetics of $\mathcal{H}$-matrices*, Computing, 70 (2003), pp. 295–334.

[50] L. Grasedyck, R. Kriemann, and S. Le Borne, *Domain decomposition based $\mathcal{H}$-LU preconditioning*, Numer. Math., 112 (2009), pp. 565–600.

[51] E. Grasso, S. Chaillat, M. Bonnet, and J.-F. Semblat, *Application of the multi-level time-harmonic fast multipole BEM to 3-D visco-elastodynamics*, Eng. Anal. Bound. Elem., 36 (2012), pp. 744–758.

[52] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, Journal of Computational Physics, 73 (1987), pp. 325 – 348.

[53] L. F. Greengard and J. Huang, *A new version of the fast multipole method for screened Coulomb interactions in three dimensions*, J. Comput. Phys., 180 (2002), pp. 642–658.

[54] H. Groemer, *Geometric Applications of Fourier Series and Spherical Harmonics*, Cambridge University Press, 1996.

[55] N. Gumerov and R. Duraiswami, *Fast multipole accelerated boundary element methods for the 3D Helmholtz equation*, tech. rep., University of Maryland Computer Science Department, 2008.

[56] N. A. Gumerov and R. Duraiswami, *Recursions for the computation of multipole translation and rotation coefficients for the 3-D Helmholtz equation*, SIAM J. Sci. Comput., 25 (2003/04), pp. 1344–1381.

[57] W. Hackbusch, *A sparse matrix arithmetic based on $\mathcal{H}$-matrices. I. Introduction to $\mathcal{H}$-matrices*, Computing, 62 (1999), pp. 89–108.

[58] ———, *Panel clustering techniques and hierarchical matrices for BEM and FEM*, (2004).

[59] ———, *Hierarchische Matrizen: Algorithmen und Analysis*, Springer-Verlag Berlin Heidelberg, 2009.

[60] W. Hackbusch and S. Börm, *Data-sparse approximation by adaptive $\mathcal{H}^2$-matrices*, Computing, 69 (2002), pp. 1–35.

[61] ———, *$\mathcal{H}^2$-matrix approximation of integral operators by interpolation*, Appl. Numer. Math., 43 (2002), pp. 129–143. 19th Dundee Biennial Conference on Numerical Analysis (2001).

[62] W. Hackbusch, B. Khoromskij, and S. A. Sauter, *On $\mathcal{H}^2$-matrices*, Springer, Berlin, 2000, pp. 9–29.

[63] W. Hackbusch and B. N. Khoromskij, *A sparse $\mathcal{H}$-matrix arithmetic. II. Application to multi-dimensional problems*, Computing, 64 (2000), pp. 21–47.

[64] W. Hackbusch and Z. P. Nowak, *On the fast matrix multiplication in the boundary element method by panel clustering*, Numer. Math., 54 (1989), pp. 463–491.

[65] W. Hackbusch and S. A. Sauter, *On numerical cubatures of nearly singular surface integrals arising in BEM collocation*, Computing, 52 (1994), pp. 139–159.

[66] N. J. Higham, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[67] R. Jakob-Chien and B. K. Alpert, *A fast spherical filter with uniform resolution*, Journal of Computational Physics, 136 (1997), pp. 580–584.

[68] M. Kachanovska, *Fast, parallel techniques for time-domain boundary integral equations*, PhD thesis, University of Leipzig, 2014.

[69] S. Koc, J. Song, and W. C. Chew, *Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem*, SIAM J. Numer. Anal., 36 (1999), pp. 906–921.

[70] R. Kriemann, *HLIBpro user manual. Technical Report 9/2008*, MPI for Mathematics in the Sciences, Leipzig, 2008.

[71] W. McLean, *Strongly elliptic systems and boundary integral equations*, (2000), pp. xiv+357.

[72] F. W. J. Olver, *Asymptotics and special functions*, AKP Classics, A K Peters Ltd., Wellesley, MA, 1997. Reprint of the 1974 original [Academic Press, New York].

[73] J. Rahola, *Diagonal forms of the translation operators in the fast multipole algorithm for scattering problems*, BIT, 36 (1996), pp. 333–358.

[74] V. Rokhlin, *Rapid solution of integral equations of classical potential theory*, J. Comput. Phys., 60 (1985), pp. 187–207.

[75] ———, *Diagonal forms of translation operators for the Helmholtz equation in three dimensions*, Applied and Computational Harmonic Analysis, 1 (1993), pp. 82 – 93.

[76] L. G. S. Börm and W. Hackbusch, *Hierarchical Matrices*, Lecture notes for a winter school on the hierarchical matrices, 2003. http://www.mis.mpg.de/publications/other-series/ln/lecturenote-2103.html.

[77] W. H. S. Börm, L. Grasedyck, *Introduction to hierarchical matrices with applications*, Engineering Analysis with Boundary Elements, 27 (2003), pp. 405–422.

[78] T. Sakuma, S. Schneider, and Y. Yasuda, *Fast solution methods*, in Computational Acoustics of Noise Propagation in Fluids - Finite and Boundary Element Methods, S. Marburg and B. Nolte, eds., Springer Berlin Heidelberg, 2008, pp. 333–366.

[79] H. E. Salzer, *Lagrangian interpolation at the Chebyshev points $X_{n,\nu} \equiv \cos(\nu\pi/n)$, $\nu = 0(1)n$; some unnoted advantages*, Comput. J., 15 (1972), pp. 156–159.

[80] J. Sarvas, *Performing interpolation and anterpolation entirely by fast Fourier transform in the 3-D multilevel fast multipole algorithm*, SIAM J. Numer. Anal., 41 (2003), pp. 2180–2196.

[81] S. A. Sauter, *Cubature techniques for 3-D Galerkin BEM*, in Boundary elements: implementation and analysis of advanced algorithms (Kiel, 1996), vol. 54 of Notes Numer. Fluid Mech., Vieweg, Braunschweig, 1996, pp. 29–44.

[82] S. A. Sauter and A. Krapp, *On the effect of numerical integration in the Galerkin boundary element method*, Numer. Math., 74 (1996), pp. 337–359.

[83] S. A. Sauter and C. Schwab, *Boundary element methods*, vol. 39 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2011. Translated and expanded from the 2004 German original.

[84] J. Song, C.-C. Lu, W. Chew, and S. W. Lee, *Fast Illinois solver code (FISC)*, Antennas and Propagation Magazine, IEEE, 40 (1998), pp. 27–34.

[85] J. Song, C.-C. Lu, and W. C. Chew, *Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects*, Antennas and Propagation, IEEE Transactions on, 45 (1997), pp. 1488–1493.

[86] O. Steinbach, *Numerical approximation methods for elliptic boundary value problems*, Springer, New York, 2008. Finite and boundary elements, Translated from the 2003 German original.

[87] M. S. Tong and W. C. Chew, *Multilevel fast multipole acceleration in the Nyström discretization of surface electromagnetic integral equations for composite objects*, IEEE Trans. Antennas and Propagation, 58 (2010), pp. 3411–3416.

[88] H. Wang and S. Xiang, *On the convergence rates of Legendre approximation*, Math. Comp., 81 (2012), pp. 861–877.

[89] G. Watson, *A Treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, England, 1944.

[90] N. Yarvin and V. Rokhlin, *A generalized one-dimensional fast multipole method with application to filtering of spherical harmonics*, J. Comput. Phys., 147 (1998), pp. 594–609.

[91] Y. Yasuda and T. Sakuma, *Analysis of sound fields in porous materials using the fast multipole BEM*, in 37th International Congress and Exposition on Noise Control (Inter-noise 2008), 2008.

[92] L. Ying, G. Biros, and D. Zorin, *A kernel-independent adaptive fast multipole algorithm in two and three dimensions*, J. Comput. Phys., 196 (2004), pp. 591–626.

[93] K. Yoshida, *Applications of Fast Multipole Method to Boundary Integral Equation Method*, PhD thesis, Kyoto University, Japan, 2001.