A STABLE MATRIX VERSION OF THE WIDEBAND FAST MULTIPOLE METHOD FOR THE 2D HELMHOLTZ KERNEL*

MICHELLE MICHELLE†, XIAOFENG OU†, AND JIANLIN XIA†

Abstract. This paper presents a stable matrix version of the wideband fast multipole method (FMM) for the 2D Helmholtz kernel. It is known that the FMM may experience stability issues in both high-frequency and low-frequency regimes, some of which can be mitigated and others are inherent in nature. Inspired by recent studies, we propose a balancing strategy to overcome the stability challenge that exists in the low-frequency regime. The balancing strategy utilizes some simple properties of Bessel and Hankel functions so as to produce theoretically guaranteed norm bounds for relevant low-rank expansion factors and translation operators. We then present an intuitive and stable matrix version of the wideband FMM, which utilizes two different expansions of the 2D Helmholtz kernels: one that always behaves well in the low-frequency regime based on our balancing strategy, and the other that behaves well (under certain conditions) in the high-frequency regime. The backward stability of this wideband FMM is rigorously justified based on our studies of the norm bounds of the low-rank factors and translation operators. Some numerical experiments demonstrate the effectiveness and the accuracy of the wideband FMM.

Key words. wideband fast multipole method, numerical stability, low-rank approximation, balancing, Bessel functions, Helmholtz kernel

AMS subject classifications. 65F35, 65F55, 15A23, 15A60

1. Introduction.

Given two sets of points $\mathbf{X} = \{x_i\}_{i=1}^M$ and $\mathbf{Y} = \{y_i\}_{i=1}^N$ in the complex plane $\mathbb{C}$, consider the evaluation of the matrix-vector product

$$\phi = Kq \quad \text{with} \quad K := [H_0(\mathsf{k}|x_i - y_j|)]_{x_i\in\mathbf{X}, y_j\in\mathbf{Y}}, \tag{1.1}$$

where K is an $M \times N$ Helmholtz kernel matrix defined by the Hankel function H_0 of the first kind and of order zero, and $\mathsf{k} > 0$ is the wavenumber. A brute-force evaluation of the matrix-vector product costs $O(MN)$. It is well-known that efficient algorithms such as the fast multipole method (FMM) [21, 27, 29, 30, 31] may be used to accelerate such matrix-vector multiplications. These fast multiplications serve as a key component in solving various scattering problems (e.g., see [15, 17, 36]).

The FMM uses degenerate/separable expansions of relevant kernel functions to construct rank-structured matrix (i.e., FMM matrix) approximations to the kernel matrices. A comprehensive study of the separability of the Helmholtz kernel can be found in [18], where the number of terms required for the degenerate expansion of the Green's function of the Helmholtz equation to be within a specific accuracy for a given wavenumber k is shown.

It has long been documented that the FMM may encounter stability issues [4, 8, 12, 22]. Possible causes of these stability issues include artificially large entries in the FMM matrix approximations and mixed products of large and tiny numbers. A heuristic attempt to overcome such an issue involving the use of scaling was discussed in [22]. In subsequent developments, a rigorous scaling strategy with justifications was used in [8] to stabilize the FMM for some 1D kernels, and an improved scaling strategy was given in [26] for 2D non-oscillating kernels such as the generalized Cauchy

*The research of Jianlin Xia was supported in part by NSF grants DMS-1819166 and DMS-2111007. The research of Michelle Michelle was supported in part by an NSERC Postdoctoral Fellowship and the Digital Research Alliance of Canada. Xiaofeng Ou's work was part of a PhD thesis at Purdue University.

†Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA (mmichell@purdue.edu, ou17@purdue.edu, xiaj@purdue.edu).

and Poisson kernels. The work in [26] also confirms the backward stability for the 2D FMM applied to those kernels.

For the Helmholtz kernel, it is known that the expansion introduced in [28] behaves well in the high-frequency regime, but it becomes fundamentally unstable in the low-frequency regime (i.e., when there is a significant 'sub-wavelength structure') due to the divergent term in the expansion [12]. A few methods have been proposed in attempts to overcome this instability. The work in [5] used results of [25] to identify problematic blocks and used an $\mathcal{H}$-matrix, which is able to be computed in a stable manner, to approximate those blocks. In [11, 12], the wideband FMM was discussed. The method uses a diagonal form expansion in the high-frequency regime and switches to a different expansion in the low-frequency regime. In the later regime, one may encounter instability in the form of overflow. A heuristic attempt has been proposed to deal with it [12, 13], but it lacks a rigorous justification. Moreover, no connection has been made on how this heuristic method affects the overall stability of the FMM.

The aim of the present paper is to address the previous issue encountered in the low-frequency regime and extend the studies of [26] to the 2D Helmholtz kernel. Due to the oscillating nature of the Helmholtz kernel (especially when there is a large k), the underlying tools used to arrive at an appropriate stabilization strategy via scaling are significantly different from those in [26].

The main contributions of this paper include the following aspects. Firstly, we present a stable degenerate expansion of the 2D Helmholtz kernel for the low-frequency regime. To this end, we introduce a *balancing* strategy based on the asymptotic behaviors of Bessel and Hankel functions to eliminate the stability risk caused by artificially large entries in the FMM matrix corresponding to this type of expansions. The strategy applies scaling to the low-rank factors and translation factors in the FMM so as to balance their norms. In fact, these factors can be shown to satisfy some norm bounds that are important to ensure stable FMM matrix-vector multiplications. Similarly, a stable translation relation is also given for the low-frequency regime. Our derivations mainly utilize basic Bessel properties and are convenient to understand.

Secondly, we present an intuitive stable matrix version of the 2D wideband FMM, where we integrate our stable degenerate expansion and translation relation for the low-frequency regime and another expansion from [27] for the high-frequency regime. This stable matrix version of the wideband FMM thus generalize the work in [26] to the Helmholtz kernels and results in a much more elaborate scheme because of the oscillatory kernel and the coexistence of two different types of expansions. The matrix version follows [26] and interprets the FMM matrix in terms of some generators corresponding to basis contributions from point sets as well as translations among the basis contributions. This avoids distinguishing the so-called multipole and local expansions and makes it convenient to understand the wideband FMM.

In the matrix version, we also provide an efficient and stable way to compute the entries of low-rank factors and translation relations in the FMM by using some recurrence relations and by exploiting the underlying symmetry.

Finally, we show the backward stability of the wideband FMM algorithm for the 2D Helmholtz kernel similarly to our previous work [26]. It has been known that rank-structured matrix-vector multiplications, though quite fast, may potentially be unstable for some situations [6, 9]. Here, we can rigorously show the stability by analyzing the growth of the backward error. Such error growth is similar to earlier results for simpler (and essentially one dimensional) hierarchical structured methods in [8, 33, 34]. Here for the 2D wideband FMM, the hierarchical algorithm architecture and our norm bounds for the generators lead to the backward stability.

The organization of this paper is as follows. In section 2, we discuss degenerate expansions of the 2D Helmholtz kernel, including our stable expansion for the low-frequency regime and a well-known expansion for the high-frequency regime. Norm bounds for low-rank factors stemming from these two expansions are discussed. In section 3, we discuss translation relations associated with the two expansions. In section 4, we give the stable matrix form of the wideband FMM and also show recurrence formulas to quickly and stably compute the entries of the some low-rank factors and translations relations. The study of the backward stability of the wideband FMM is given in section 5. Some numerical tests are provided in section 6, followed by some concluding remarks in section 7. Finally, the proofs of some lemmas in section 2 are deferred to Appendix A. The following is a list of notation used throughout the paper.

- $A = [a_{ij}]_{1\le i\le m,1\le j\le n}$ or $A = [a_{ij}]_{m\times n}$ denotes an $m \times n$ matrix A.
- $[\kappa(x_i, y_j)]_{x_i\in\mathbf{x},y_j\in\mathbf{y}}$ denotes the matrix given by the evaluation of $\kappa(x, y)$ at all the points $x \in \mathbf{x}, y \in \mathbf{y}$. Sometimes, the definition of a matrix may use a mixture of index sets and point sets.
- For $A = [a_{ij}]_{m\times n}$, the following norms will be frequently used in the paper:

$$\|A\|_{\max} := \max_{1\le i\le m,1\le j\le n} |a_{ij}|, \quad \|A\|_{1,1} := \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|,$$

$$\|A\|_1 := \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|, \quad \|A\|_\infty := \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|.$$

- $\mathrm{diag}(a_{1,1}, \ldots, a_{n,n})$ denotes a diagonal matrix with the specified diagonal entries.
- $\Re(z)$ and $\Im(z)$ denote the real and imaginary parts of a complex number z, respectively.
- For a complex number ξ, θ_ξ denotes its principal angle, where $\theta_\xi \in (-\pi, \pi]$.
- If $j \le k$, define $\prod_{i=k}^{\ge j}$ or $\prod\limits_{i=k}^{\ge j} A_i := A_k A_{k-1} \cdots A_j$, where A_i, $i = j, \ldots, k$, are matrices of suitable sizes. If $k < j$, then $\prod_{i=k}^{\ge j} A_i := I$ (the identity matrix).

2. 2D Helmholtz kernel: stable degenerate expansions and low-rank approximations. In this section, we discuss stable degenerate expansions of the 2D Helmholtz kernel $\kappa(x, y) = H_0(\mathsf{k}|x - y|)$ for both the low-frequency regime and the high-frequency regime. We will mainly focus on the former case. The latter is well-studied, but we will revisit some essential ideas so as to connect it with our stable matrix version of the wideband 2D FMM and its backward stability analysis.

The degenerate expansion of $\kappa(x, y)$ involves the concept of *well-separated* sets. We follow the one used in [8, 26, 29]. Two sets of points $\mathbf{x} \subset \mathbf{X}$ and $\mathbf{y} \subset \mathbf{Y}$ are well-separated with a separation ratio $\tau \in (0, 1)$ if

$$\delta_{\mathbf{x}} + \delta_{\mathbf{y}} \le \tau |o_{\mathbf{x}} - o_{\mathbf{y}}|, \tag{2.1}$$

where $\delta_{\mathbf{x}}$ and $\delta_{\mathbf{y}}$ are the radii of $\mathbf{x}$ and $\mathbf{y}$ corresponding to given centers $o_{\mathbf{x}}$ and $o_{\mathbf{y}}$, respectively. With well-separated $\mathbf{x}$ and $\mathbf{y}$, we seek to find a degenerate expansion of $\kappa(x, y)$ for each $x \in \mathbf{x}, y \in \mathbf{y}$ so as to obtain a low-rank approximation of the kernel matrix $K_{\mathbf{x},\mathbf{y}} := [\kappa(x_i, y_j)]_{x_i\in\mathbf{x},y_j\in\mathbf{y}}$ of the following form:

$$K_{\mathbf{x},\mathbf{y}} = UBV^{\mathsf{T}} + \mathcal{E}_r \approx UBV^{\mathsf{T}}, \tag{2.2}$$

where B is an $r \times r$ matrix with r a positive integer determined by the accuracy requirement, and $\mathcal{E}_r$ is the error matrix.

2.1. Basic properties of Bessel functions. Before we present our stable degenerate expansion of the 2D Helmholtz kernel, we first discuss some basic properties and bounds of Bessel functions, which will be used in our subsequent analysis. Even though the derivations of the properties are relatively elementary, we are not aware of any references that explicitly record these results.

Let $H_p(z)$ be the Hankel function of the first kind:

$$H_p(z) = J_p(z) + \mathrm{i}Y_p(z), \tag{2.3}$$

where $z > 0$, $p \in \mathbb{Z}$, and $J_p(z)$ and $Y_p(z)$ are respectively the pth order Bessel functions of the first and the second kinds. It was shown in [3, Lemma 2] that, for fixed $z > 0$,

$$|H_{p_1}(z)| < |H_{p_2}(z)| \quad \text{for all } p_1 \leq p_2 \text{ and } p_1, p_2 \in \mathbb{N} \cup \{0\}. \tag{2.4}$$

Meanwhile, for fixed $p \in \mathbb{N} \cup \{0\}$,

$$|H_p(z_1)| > |H_p(z_2)| \quad \text{for all } z_1 \leq z_2 \text{ and } z_1, z_2 > 0. \tag{2.5}$$

The next lemma studies the monotonicity of $|H_p(z)|$ and is proved in Appendix A.

LEMMA 2.1. *The Hankel function $H_p(z)$ has the following properties.*
(i) If $z \geq 1/2$ and $0 \leq p \leq z$, then $|H_p(z)| < \sqrt{4/\pi}$.
(ii) If $z > 0$, $p \geq 0$, and $0 < \lambda < 1$, then $|H_p(\lambda z)| \leq \lambda^{-1/2} |H_{p/\lambda}(z)|$.

Next, following [3, (6)] and using $Y_p(z)$ in (2.3), define

$$C_p(z) := -Y_p(z) \sqrt{\frac{\pi p}{2}} \left(\frac{ez}{2p}\right)^p, \quad p \in \mathbb{N} \cup \{0\}, z > 0. \tag{2.6}$$

By [3, Corollary 6], we have $C_p(z) > C_{p+1}(z) > \cdots > 1$ for all $p \geq 2$ and $p \geq z > 0$. The next lemma includes some basic properties of (2.6) and is proved in Appendix A.

LEMMA 2.2. *Let $p \in \mathbb{N} \cup \{0\}$ and $z > 0$. If $p \geq 2$ and $0 < z \leq p$, then the function $C_p(z)$ in (2.6) has the following properties:*
(i) $C_p(z) < \frac{4\sqrt{2}}{e} C_{p+1}(z)$,
(ii) $C_p(z)$ is a strictly increasing function of z,
(iii) $C_p(z) \leq |H_p(p)| \sqrt{\frac{\pi p}{2}} \left(\frac{e}{2}\right)^p$.

Also, we recall *Graf's addition formulas*, which are essential to the theory of the FMM for the Helmholtz kernel (see, e.g., [1, (9.1.79)], [4, Theorem 3.1], and [24, (1)]). Suppose $z_1, z_2 \in \mathbb{C}$. We have

$$J_p(\mathsf{k}|z_1 - z_2|) e^{\pm \mathrm{i} p \theta_{\mathsf{k}(z_1 - z_2)}} = \sum_{l=-\infty}^{\infty} J_{p+l}(\mathsf{k}|z_1|) e^{\pm \mathrm{i}(p+l)\theta_{\mathsf{k}z_1}} J_l(\mathsf{k}|z_2|) e^{\mp \mathrm{i} l \theta_{\mathsf{k}z_2}}. \tag{2.7}$$

Furthermore, if $|z_1| > |z_2|$, we have

$$H_p(\mathsf{k}|z_1 - z_2|) e^{\pm \mathrm{i} p \theta_{\mathsf{k}(z_1 - z_2)}} = \sum_{l=-\infty}^{\infty} H_{p+l}(\mathsf{k}|z_1|) e^{\pm \mathrm{i}(p+l)\theta_{\mathsf{k}z_1}} J_l(\mathsf{k}|z_2|) e^{\mp \mathrm{i} l \theta_{\mathsf{k}z_2}}. \tag{2.8}$$

Given $r \in \mathbb{N}$, we provide the truncation errors of (2.7) and (2.8), as they are used to derive truncation error bounds of degenerate expansions of the 2D Helmholtz kernel. Some relevant analysis was performed in [2, 3, 4, 5, 23, 24, 28]. Various forms of truncation bounds are available. For example, exponentially decaying truncation error bounds of (2.7) and (2.8) were provided in [5, Lemma 4.2] by following the analysis of [2] and putting some conditions on the separation parameter and the expansion order. Meanwhile, computable error bounds were given in [3].

Even though we can use the relatively sharp truncation error bounds of (2.7) and (2.8) provided in [23, 24], these bounds have rather complicated expressions. The bounds presented in this paper have simpler expressions and can be derived only by using the previously discussed basic properties of Bessel functions. These basic properties of Bessel functions make it more convenient to study the backward errors later in the paper since they can be immediately used to bound relevant matrix norms.

We now present the following two lemmas on truncation error bounds of (2.7) and (2.8) and their proofs are given in Appendix A.

LEMMA 2.3. *Define*

$$E_r^{J_p}(\mathsf{k}z_1, \mathsf{k}z_2) := \sum_{|l|\geq r \text{ or } |p+l|\geq r} J_{p+l}(\mathsf{k}|z_1|)e^{\pm \mathrm{i}(p+l)\theta_{\mathsf{k}z_1}} J_l(\mathsf{k}|z_2|)e^{\mp \mathrm{i}l\theta_{\mathsf{k}z_2}}.$$

Suppose $z_{\max} \geq \max(|z_1|, |z_2|)$. *Let* $r \geq \mathsf{k}z_{\max}$ *such that* $(\mathsf{k}z_{\max})^{l-r} \leq l!/r!$ *for* $l \geq r$. *Then,*

$$|E_r^{J_p}(\mathsf{k}z_1, \mathsf{k}z_2)| \leq \frac{8}{r!}\left(\frac{\mathsf{k}z_{\max}}{2}\right)^r.$$

LEMMA 2.4. *Define*

$$E_r^{H_0}(\mathsf{k}w, \mathsf{k}t) := \sum_{|l|\geq r} H_l(\mathsf{k}|w|)e^{\pm \mathrm{i}l\theta_{\mathsf{k}w}} J_l(\mathsf{k}|t|)e^{\mp \mathrm{i}l\theta_{\mathsf{k}t}}.$$

Suppose $|t| \leq t_{\max} \leq \tau|w|$ *for some* $\tau \in (0,1)$. *If* $r \geq \max(\mathsf{k}t_{\max}/\tau, 2)$, *then*

$$|E_r^{H_0}(\mathsf{k}w, \mathsf{k}t)| \leq \frac{2\sqrt{2}}{\pi r}\frac{\tau^r}{1-\tau}C_r\left(\frac{\mathsf{k}t_{\max}}{\tau}\right),$$

where $C_r(\cdot)$ *is defined in* (2.6).

2.2. Stable degenerate expansion for the low-frequency regime. In this subsection, we present a stable expansion for the 2D Helmholtz kernel in the low-frequency regime and describe a balancing strategy that leads to such an expansion. A similar expansion (without the balancing strategy) was implicitly derived in [28]. In a later study, a matrix representation of such an expansion was provided in [4, Theorem 3.2] and is called a non-diagonal form (since the factor associated with the multipole-to-local expansion is not a diagonal matrix). The authors of [4] left the truncation error of the degenerate expansion as an infinite series involving Hankel and Bessel functions. For the sake of completeness, we provide a bound for the truncation error using some basic Bessel properties presented earlier, which shows that the truncation error diminishes as the number of terms in the expansion increases.

Suppose $\mathbf{x} = \{x_i\}_{i=1}^n$ and $\mathbf{y} = \{y_j\}_{j=1}^m$ are well-separated with separation ratio τ. As before, use $o_{\mathbf{x}}$ and $\delta_{\mathbf{x}}$ to mean the center and radius of $\mathbf{x}$. Similarly understand $o_{\mathbf{y}}$ and $\delta_{\mathbf{y}}$. For any $x \in \mathbf{x}$, $y \in \mathbf{y}$, let

$$t = (x - o_{\mathbf{x}}) - (y - o_{\mathbf{y}}), \quad w = o_{\mathbf{y}} - o_{\mathbf{x}}. \tag{2.9}$$

Then $x - y = -(w - t)$. The separation condition (2.1) leads to $|t| \leq \tau|w| < |w|$. Applying (2.8) to $H_0(\mathsf{k}|x-y|) = H_0(\mathsf{k}|w-t|)$, using (2.7), and noting that $J_{-p}(z) =$

$(-1)^p J_p(z)$ for all $p \in \mathbb{Z}$ and $z \geq 0$, we have the following degenerate expansion:

(2.10)

$$\begin{aligned} H_0(\mathsf{k}|x-y|) &= \sum_{|k|\leq r} H_k(\mathsf{k}|w|)e^{-\mathrm{i}k\theta_{\mathsf{k}w}} J_k(\mathsf{k}|t|)e^{\mathrm{i}k\theta_{\mathsf{k}t}} + E^{H_0}_{r+1}(\mathsf{k}w, \mathsf{k}t) \\ &= \sum_{|k|\leq r} H_k(\mathsf{k}|w|)e^{-\mathrm{i}k\theta_{\mathsf{k}w}} \left(\sum_{|l|\leq r, |k-l|\leq r} (-1)^l g_{k-l}(\mathsf{k}(x-o_{\mathbf{x}}))g_l(\mathsf{k}(y-o_{\mathbf{y}})) \right) \\ &\quad + \varepsilon_r(\mathsf{k}x, \mathsf{k}y) \\ &= \sum_{|k|\leq r} \sum_{|l|\leq r, |k-l|\leq r} \left(H_k(\mathsf{k}|w|)e^{-\mathrm{i}k\theta_{\mathsf{k}w}} (-1)^l g_{k-l}(\mathsf{k}(x-o_{\mathbf{x}}))g_l(\mathsf{k}(y-o_{\mathbf{y}})) \right) + \varepsilon_r(\mathsf{k}x, \mathsf{k}y) \\ &= \sum_{|l|\leq r, |p|\leq r} \left(\sum_{|p+l|\leq r} b_{p,l} g_p(\mathsf{k}(x-o_{\mathbf{x}}))g_l(\mathsf{k}(y-o_{\mathbf{y}})) \right) + \varepsilon_r(\mathsf{k}x, \mathsf{k}y), \end{aligned}$$

where $g_j(\xi) := J_j(|\xi|)e^{\mathrm{i}j\theta_\xi}$ for all $\xi \in \mathbb{C}$,

$$b_{i,j} = \begin{cases} (-1)^j H_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)e^{-\mathrm{i}(i+j)\theta_{\mathsf{k}(o_{\mathbf{y}}-o_{\mathbf{x}})}}, & |i+j| \leq r, \\ 0, & |i+j| > r, \end{cases} \tag{2.11}$$

and the remainder term is

$$\varepsilon_r(\mathsf{k}x, \mathsf{k}y) := E^{H_0}_{r+1}(\mathsf{k}w, \mathsf{k}t) + \sum_{|k|\leq r} H_k(\mathsf{k}|w|)e^{-\mathrm{i}k\theta_{\mathsf{k}w}} E^{J_k}_{r+1}(\mathsf{k}(x-o_{\mathbf{x}}), \mathsf{k}(y-o_{\mathbf{y}})). \tag{2.12}$$

Here, $E^{H_0}_{r+1}$ and $E^{J_k}_{r+1}$ are defined in Lemmas 2.3 and 2.4 respectively. This expansion can then be used to obtain a low-rank approximation to the kernel matrix as follows.

PROPOSITION 2.5. *Suppose $\mathbf{x} = \{x_i\}_{i=1}^n$ and $\mathbf{y} = \{y_j\}_{j=1}^m$ are well-separated with separation ratio τ. Then, the low-rank approximation to the kernel matrix $K_{\mathbf{x},\mathbf{y}}$ in (2.2) with $\kappa(x, y) = H_0(\mathsf{k}|x - y|)$ has the following form:*

$$U := [g_j(\mathsf{k}(x_i - o_{\mathbf{x}}))]_{1\leq i\leq n, -r\leq j\leq r}, \quad V := [g_j(\mathsf{k}(y_i - o_{\mathbf{y}}))]_{1\leq i\leq m, -r\leq j\leq r},$$
$$B := [b_{i,j}]_{-r\leq i,j\leq r},$$

where $g_j(\xi) := J_j(|\xi|)e^{\mathrm{i}j\theta_\xi}$ for all $\xi \in \mathbb{C}$ and $b_{i,j}$ is given in (2.11). Moreover, $\mathcal{E}_r$ in (2.2) satisfies

$$|\mathcal{E}_r| \leq \frac{18\sqrt{2}\tau^{r+1}}{\pi(1-\tau)} C_{r+1}\left(\frac{\mathsf{k}(\delta_{\mathbf{x}} + \delta_{\mathbf{y}})}{\tau}\right), \quad \text{for all} \quad r \geq \max\left(\frac{\mathsf{k}(\delta_{\mathbf{x}} + \delta_{\mathbf{y}})}{\tau}, 2\right), \tag{2.13}$$

with $C_{r+1}(\cdot)$ defined in (2.6). Additionally, the entries of U, V, B satisfy

(2.14)

$$\|U\|_{\max} \leq 1, \quad \|V\|_{\max} \leq 1, \quad \|B\|_{\max} \geq \sqrt{\frac{2}{\pi r}} \left(\frac{2r}{e\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|}\right)^r C_r(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|).$$

Proof. Applying the expansion (2.10) to the entries of $K_{\mathbf{x},\mathbf{y}}$ to immediately get the low-rank factors. Our goal is to bound the remainder term $\varepsilon_r(\mathsf{k}x, \mathsf{k}y)$ in (2.12).

For any $x \in \mathbf{x}$, $y \in \mathbf{y}$, t, w in (2.9) satisfies $|t| \leq \delta_{\mathbf{x}} + \delta_{\mathbf{y}} \leq \tau|w|$. We can set $t_{\max} = \delta_{\mathbf{x}} + \delta_{\mathbf{y}}$ in Lemma 2.4 and obtain

(2.15)

$$\left|E^{H_0}_{r+1}(\mathsf{k}w, \mathsf{k}t)\right| \leq \frac{2\sqrt{2}}{\pi(r+1)} \frac{\tau^{r+1}}{1-\tau} C_{r+1}\left(\frac{\mathsf{k}t_{\max}}{\tau}\right), \quad \text{for all} \quad r \geq \max\left(\frac{\mathsf{k}t_{\max}}{\tau}, 2\right).$$

Since $t_{\max} \geq \max(\delta_{\mathbf{x}}, \delta_{\mathbf{y}}) \geq \max\{|x - o_{\mathbf{x}}|, |y - o_{\mathbf{y}}|\}$, we can utilize a step in the proof of Lemma 2.3 in Appendix A (specifically, by setting $z_{\max} = t_{\max}$ in (A.1)) to get

$$\left| E_{r+1}^{J_k}(\mathsf{k}(x - o_{\mathbf{x}}), \mathsf{k}(y - o_{\mathbf{y}})) \right| \leq 4 \sum_{l=r+1}^{\infty} |J_l(\mathsf{k}t_{\max})| .$$

Recall that, for a fixed nonnegative integer k, $|H_k(\mathsf{k}z)|$ is a strictly decreasing function in $z > 0$ by (2.5). Thus from $|w| \geq (\delta_{\mathbf{x}} + \delta_{\mathbf{y}})/\tau = t_{\max}/\tau$, we have

(2.16)
$$\begin{aligned} \left| H_k(\mathsf{k}|w|) E_{r+1}^{J_k}(\mathsf{k}(x - o_{\mathbf{x}}), \mathsf{k}(y - o_{\mathbf{y}})) \right| &\leq \left| H_k\left(\frac{\mathsf{k}t_{\max}}{\tau}\right) E_{r+1}^{J_k}(\mathsf{k}(x - o_{\mathbf{x}}), \mathsf{k}(y - o_{\mathbf{y}})) \right| \\ &\leq 4 \left| H_k\left(\frac{\mathsf{k}t_{\max}}{\tau}\right) \right| \sum_{l=r+1}^{\infty} |J_l(\mathsf{k}t_{\max})| \leq 4 \sum_{l=r+1}^{\infty} \left| H_l\left(\frac{\mathsf{k}t_{\max}}{\tau}\right) J_l(\mathsf{k}t_{\max}) \right| \\ &\leq \frac{8\sqrt{2}}{\pi(r+1)} C_{r+1}\left(\frac{\mathsf{k}t_{\max}}{\tau}\right)\left(\frac{\tau^{r+1}}{1-\tau}\right), \quad \text{for all } r \geq |k|, \end{aligned}$$

where we have used a similar calculation as in the proof of Lemma 2.4 to arrive at the last inequality. Therefore, the bounds in (2.15) and (2.16) imply

$$\begin{aligned} |E(\mathsf{k}x, \mathsf{k}y)| &\leq \left| E_{r+1}^{H_0}(\mathsf{k}w, \mathsf{k}t) \right| + \sum_{|k| \leq r} \left| H_k(\mathsf{k}|w|) E_{r+1}^{J_k}(\mathsf{k}(x - o_{\mathbf{x}}), \mathsf{k}(y - o_{\mathbf{y}})) \right| \\ &\leq \frac{18\sqrt{2}}{\pi} C_{r+1}\left(\frac{\mathsf{k}t_{\max}}{\tau}\right)\left(\frac{\tau^{r+1}}{1-\tau}\right), \quad \text{for all } r \geq \max\left(\frac{\mathsf{k}t_{\max}}{\tau}, 2\right), \end{aligned}$$

which gives (2.13).

The first and second bounds in (2.14) hold because of the fact that $|J_j(\mathsf{k}z)| \leq 1$ for all $j \in \mathbb{Z}$ and $z \geq 0$ (see [1, 32]). The third bound in (2.14) holds because of (2.6) and the fact that $|H_r(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| \geq |Y_r(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)|$. The proof is completed. ∎

Proposition 2.5 implies that, although the U, V generators have entrywise magnitudes bounded by 1, some entries of B can have extremely large magnitudes if $r > \frac{e\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|}{2}$. From (2.14), we observe that when $o_{\mathbf{x}} - o_{\mathbf{y}}$ is very small, or when a large r is used in (2.10) for high accuracy, we have $\frac{2r}{e\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|} \gg 1$ and thus $\|B\|_{\max}$ grows very fast. This poses a stability risk and may even cause overflow.

It is often preferred to control the norm of B as well [8, 9, 26]. Therefore, we follow a strategy in [8] to scale the low-rank factors U, B, and V. While the strategy in [8] uses some scaling factors based on Stirling's formula, we follow the asymptotic behaviors of Bessel and Hankel functions to design some scaling factors for the Helmholtz kernel. These factors are related to the heuristic ones briefly discussed in [12, 13]. The difference is that our present work rigorously justifies that these scaling factors lead to the simultaneous control of the norms of all the low-rank factors, and such norm control eventually leads to the stability. Our balancing strategy is as follows.

For $p \in \mathbb{Z}$, define the *scaling factors* for the set $\mathbf{x}$ as follows:

$$\lambda_{\mathbf{x},p} := \max\left\{1, |p|!\left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^{|p|}\right\}, \tag{2.17}$$

and define $\lambda_{\mathbf{y},p}$ similarly for the set $\mathbf{y}$.

LEMMA 2.6. *The sequence* $\{\lambda_{\mathbf{x},p}\}_{p \in \mathbb{Z}}$ *satisfies* $\lambda_{\mathbf{x},0} < \lambda_{\mathbf{x},1} = \lambda_{\mathbf{x},-1}$ *and* $\lambda_{\mathbf{x},p} < \lambda_{\mathbf{x},p+\mathrm{sign}(p)}$ *for all* $p \in \mathbb{Z} \setminus \{0\}$, *where* $\mathrm{sign}(\cdot)$ *is the sign function.*

Proof. Without loss of generality, assume $p \geq 0$, and let $a_p = p! \left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^p$. Clearly, $\frac{a_{p+1}}{a_p} = \frac{2(p+1)}{\mathsf{k}\delta_{\mathbf{x}}}$. The sequence $\{a_p\}_{p=0}^{\infty}$ either monotonically increases if $\frac{2(p+1)}{\mathsf{k}\delta_{\mathbf{x}}} > 1$ for all $p \geq 0$, or first monotonically decreases if $\frac{2(p+1)}{\mathsf{k}\delta_{\mathbf{x}}} < 1$ for finitely many p's and afterwards monotonically increases for all other p's. In either case, we have

$$a_p \leq \max\{a_0, a_{p+1}\} = \lambda_{\mathbf{x},p+1}.$$

Therefore, $\lambda_{\mathbf{x},p} = \max\{1, a_p\} \leq \lambda_{\mathbf{x},p+1}$. ∎

Given these scaling factors, we can then modify the expansion in (2.10) to get a stable non-diagonal expansion of the 2D Helmholtz kernel. Accordingly, a stable low-rank approximation to $K_{\mathbf{x},\mathbf{y}}$ can be obtained as follows.

THEOREM 2.7. *Suppose $\mathbf{x} = \{x_i\}_{i=1}^n$ and $\mathbf{y} = \{y_j\}_{j=1}^m$ are well-separated with separation ratio $\tau \leq 2/e$. Then, the low-rank approximation to the kernel matrix $K_{\mathbf{x},\mathbf{y}}$ in (2.2) with $\kappa(x,y) = H_0(\mathsf{k}|x-y|)$ has the following form:*

(2.18)
$$U := [g_j(\mathsf{k}(x_i - o_{\mathbf{x}}))\lambda_{\mathbf{x},j}]_{1\leq i\leq n, -r\leq j\leq r}, \quad V := [g_j(\mathsf{k}(y_i - o_{\mathbf{y}}))\lambda_{\mathbf{y},j}]_{1\leq i\leq m, -r\leq j\leq r},$$
$$B := [b_{i,j}]_{-r\leq i,j\leq r} \quad \textit{with}$$
$$b_{i,j} := \begin{cases} (-1)^j \lambda_{\mathbf{x},i}^{-1}\lambda_{\mathbf{y},j}^{-1} H_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) e^{-\mathrm{i}(i+j)\theta_{\mathsf{k}(o_{\mathbf{y}}-o_{\mathbf{x}})}} & |i+j| \leq r, \\ 0 & |i+j| > r, \end{cases}$$

where $g_j(\xi) := J_j(|\xi|)e^{\mathrm{i}j\theta_\xi}$ for all $\xi \in \mathbb{C}$. Additionally, (2.13) holds. Moreover, the entries of U, V, B satisfy

$$\|U\|_{\max} \leq 1, \quad \|V\|_{\max} \leq 1, \tag{2.19}$$

$$\|B\|_{\max} \leq 8\pi^{-1}\max\{1, K_{\max}\}, \quad \|B\|_{1,1} \leq 72\pi^{-1} r^2 \max\{1, K_{\max}\}, \tag{2.20}$$

where $K_{\max} := \max_{x_i\in\mathbf{x}, y_j\in\mathbf{y}} |\kappa(x_i, y_j)|$.

Proof. With the incorporation of the scaling factors into the degenerate expansion (2.10), the truncation error remains the same and is identical to that in Proposition 2.5. We now analyze how the proposed scaling factors affect the norms of the low-rank factors presented in Proposition 2.5. Recall that, for $z \geq 0$, $|J_j(\mathsf{k}z)| \leq 1$ for all $j \in \mathbb{Z}$, and also $|J_j(z)| \leq (z/2)^j/j!$ for all $j \in \mathbb{N}$ (see [1, 32]). For $1 \leq i \leq n$, $-r \leq j \leq r$, we have

$$|g_j(\mathsf{k}(x_i - o_{\mathbf{x}}))\lambda_{\mathbf{x},j}| = \max\left\{|J_j(\mathsf{k}|x_i - o_{\mathbf{x}}|)|, |J_j(\mathsf{k}|x_i - o_{\mathbf{x}}|)|\,|j|!\left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^{|j|}\right\}$$
$$\leq \max\left\{1, \frac{1}{|j|!}\left(\frac{\mathsf{k}|x_i - o_{\mathbf{x}}|}{2}\right)^{|j|}|j|!\left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^{|j|}\right\} = \max\left\{1, \left(\frac{|x_i - o_{\mathbf{x}}|}{\delta_{\mathbf{x}}}\right)^{|j|}\right\} \leq 1.$$

Therefore, $\|U\|_{\max} \leq 1$. Similarly, we have $\|V\|_{\max} \leq 1$.

Next, we show the upper bounds for norms of B. Since $|H_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| = |H_{|i+j|}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)|$ and $|H_{|i+j|}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| \leq |H_{|i|+|j|}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)|$ for $|i+j| \leq r$ by (2.4), we may assume without loss of generality that $0 \leq i, j \leq r$. Note that by the definition of the scaling factors,

$$\lambda_{\mathbf{x},j}^{-1} = \min\left\{1, \frac{1}{j!}\left(\frac{\mathsf{k}\delta_{\mathbf{x}}}{2}\right)^j\right\}. \tag{2.21}$$

There are three cases to discuss.

(i) Suppose $i = j = 0$. If $|o_{\mathbf{x}} - o_{\mathbf{y}}| \geq |x - y|$, by (2.5), we immediately have $|b_{0,0}| \leq |H_0(\mathsf{k}|x-y|)|$. Now, suppose that $|o_{\mathbf{x}} - o_{\mathbf{y}}| < |x-y|$. Since $\mathbf{x}$ and $\mathbf{y}$ are well-separated, we have

$$\big||x-y| - |o_{\mathbf{x}} - o_{\mathbf{y}}|\big| \leq |x - o_{\mathbf{x}}| + |y - o_{\mathbf{y}}| \leq \delta_{\mathbf{x}} + \delta_{\mathbf{y}} \leq \tau |o_{\mathbf{x}} - o_{\mathbf{y}}|.$$

That is, $(1+\tau)^{-1}|x-y| \leq |o_{\mathbf{x}} - o_{\mathbf{y}}|$. Then,

$$|b_{0,0}| \leq \sqrt{\frac{|x-y|}{|o_{\mathbf{x}} - o_{\mathbf{y}}|}}\, |H_0(\mathsf{k}|x-y|)| \leq \sqrt{1+\tau}\, |H_0(\mathsf{k}|x-y|)|\,.$$

(ii) Suppose $1 \leq i+j \leq \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|$. By item (i) of Lemma 2.1, $|b_{i,j}| \leq |H_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| \leq \sqrt{4/\pi}$.

(iii) Lastly, suppose $1 \leq i+j$ and $\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}| < i+j$. Since $C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) > 0$ and $|H_i(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| \leq \sqrt{2}|Y_i(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)|$, we have

$$\begin{aligned}
|b_{i,j}| &\leq \sqrt{2}\left|\lambda_{\mathbf{x},i}^{-1}\lambda_{\mathbf{y},j}^{-1} Y_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)\right| \\
&\leq \frac{\sqrt{2}}{i!j!}\left(\frac{\mathsf{k}\delta_{\mathbf{x}}}{2}\right)^i \left(\frac{\mathsf{k}\delta_{\mathbf{y}}}{2}\right)^j C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)\sqrt{\frac{2}{\pi(i+j)}}\left(\frac{2(i+j)}{e\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|}\right)^{i+j} \\
&= \frac{2C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)}{i!j!\sqrt{\pi(i+j)}}\left(\frac{\delta_{\mathbf{x}}}{|o_{\mathbf{x}} - o_{\mathbf{y}}|}\right)^i \left(\frac{\delta_{\mathbf{y}}}{|o_{\mathbf{x}} - o_{\mathbf{y}}|}\right)^j \left(\frac{i+j}{e}\right)^{i+j} \\
&\leq \frac{\sqrt{2}C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)}{\pi(i+j)}\tau^{i+j}\frac{(i+j)!}{i!j!}\left(\frac{\delta_{\mathbf{x}}}{\delta_{\mathbf{x}}+\delta_{\mathbf{y}}}\right)^i \left(\frac{\delta_{\mathbf{y}}}{\delta_{\mathbf{x}}+\delta_{\mathbf{y}}}\right)^j \\
&\leq \frac{\sqrt{2}C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)}{\pi(i+j)}\tau^{i+j},
\end{aligned}$$

where we have used Stirling's formula in the fourth line. Furthermore,

$$\begin{aligned}
C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) &\leq \frac{4\sqrt{2}}{e}C_{i+j+1}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) \\
&\leq \frac{4\sqrt{2}}{e}\,|H_{i+j+1}(i+j+1)|\,\sqrt{\frac{\pi(i+j+1)}{2}}\left(\frac{e}{2}\right)^{i+j+1} \\
&\leq \frac{4\sqrt{2}}{e}\sqrt{\frac{4}{\pi}}\sqrt{\frac{\pi(i+j+1)}{2}}\left(\frac{e}{2}\right)^{i+j+1} = 4\sqrt{i+j+1}\left(\frac{e}{2}\right)^{i+j},
\end{aligned}$$

where we have used items (i) and (iii) of Lemma 2.2 in the first and second lines respectively, and item (i) of Lemma 2.1 in the third line. Thus,

$$|b_{i,j}| \leq \frac{\sqrt{2}\tau^{i+j}}{\pi(i+j)}C_{i+j}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) \leq \frac{4\sqrt{2}}{\pi}\frac{\sqrt{i+j+1}}{(i+j)}\left(\frac{e}{2}\tau\right)^{i+j} \leq \frac{8}{\pi}\left(\frac{e}{2}\tau\right)^{2r}.$$

Accordingly, if $\tau \leq 2/e$, then $|b_{i,j}| \leq 8/\pi$.

Combining these three cases, we have the first inequality in (2.20). The second inequality of (2.20) can be deduced from the fact that B is an $(2r+1)\times(2r+1)$ matrix and $\|B\|_{1,1} \leq (2r+1)^2\|B\|_{\max} \leq 9r^2\|B\|_{\max}$. The proof is completed. ∎

This theorem indicates that we can control the norms of B while simultaneously maintaining the maximum-norm bound 1 for the U, V factors. Later in section 5, the norm bounds in Theorem 2.7 will be used to show the stability of the FMM for the

Helmholtz kernel. Additionally, the low-rank factors U, V, B in Theorem 2.7 may be viewed as scaled versions of those in Proposition 2.5. In what follows, U, B, V refer to the low-rank factors in Theorem 2.7, while $\widehat{U}, \widehat{B}, \widehat{V}$ refer to the low-rank factors in Proposition 2.5. We can immediately observe that

$$U = \widehat{U}\Lambda_{\mathbf{x}}, \quad B = \Lambda_{\mathbf{x}}^{-1}\widehat{B}\Lambda_{\mathbf{y}}^{-1}, \quad \text{and} \quad V = \widehat{V}\Lambda_{\mathbf{y}},$$

where $\Lambda_{\mathbf{x}} := \mathrm{diag}(\lambda_{\mathbf{x},-r}, \lambda_{\mathbf{x},-r+1}, \ldots, \lambda_{\mathbf{x},r-1}, \lambda_{\mathbf{x},r})$ with $\{\lambda_{\mathbf{x},p}\}_{p=-r}^{r}$ the scaling factors in (2.17), and $\Lambda_{\mathbf{y}}$ is defined similarly.

2.3. Stable degenerate expansion for the high-frequency regime. In the literature, there is another well-known degenerate expansion of the Hankel function that produces a diagonal matrix for the multipole-to-local expansion [27]. Its derivation and truncation error analysis are further studied in [2, 4, 5]. Because of the nice diagonal property, such an expansion is often called a diagonal form [4] and its computation can be done efficiently by the FFT. Additionally, for sufficiently large arguments of the Hankel function (i.e., in the high-frequency regime), the scheme itself is numerically stable. That is, we do not need to introduce any modifications in this expansion. For small arguments (i.e., in the low-frequency regime), this expansion becomes unstable [16, Section 1.3] due to the entries in the matrix $\tilde{B}$ below (or, to use the classical FMM terminology, the multipole-to-local expansion). This is where we switch to the expansion discussed in subsection 2.2. Even though the expansion in subsection 2.2 can be used in the high-frequency regime, it may not be as efficient as the expansion in this subsection due to the non-diagonal factors. The foregoing explanation describes a key idea of the wideband FMM [12].

Here, we briefly revisit one degenerate expansion of $\kappa(x, y) = H_0(\mathsf{k}|x-y|)$ in the high-frequency regime, present its matrix representation, and give some norm bounds for the low-rank factors. We first recall the corresponding low-rank factors and the truncation error. These results can be found in [2, Theorem 3.1] and [5, Theorem 4.3]. As before, assume $\mathbf{x} = \{x_i\}_{i=1}^{n}$ and $\mathbf{y} = \{y_j\}_{j=1}^{m}$ are well-separated with separation ratio τ. Denote the low-rank factors in (2.2) from this expansion by $\tilde{U}, \tilde{B}, \tilde{V}$, and denote the corresponding error matrix by $\tilde{\mathcal{E}}_r$. Then,

$$\tilde{U} := [f_j(\mathsf{k}(x_i - o_{\mathbf{x}}))]_{1\le i\le n, 1\le j\le 2r+1}, \quad \tilde{V} := [\overline{f_j(\mathsf{k}(y_i - o_{\mathbf{y}}))}]_{1\le i\le m, 1\le j\le 2r+1}, \tag{2.22}$$

$$\tilde{B} := \mathrm{diag}(b_{1,1}, \ldots, b_{2r+1,2r+1}) \quad \text{with} \tag{2.23}$$
$$\tilde{b}_{i,i} := \sum_{k=-r}^{r} \frac{(-\mathrm{i})^k}{2r+1} H_k(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) e^{\mathrm{i}k(\theta_{\mathsf{k}(o_{\mathbf{y}}-o_{\mathbf{x}})} - \frac{2\pi k}{2r+1})},$$

where

$$f_p(\xi) := e^{-\mathrm{i}|\xi|\cos(\frac{2\pi p}{2r+1} - \theta_\xi)}, \quad \text{for} \quad \xi \in \mathbb{C}, p \in \mathbb{N}. \tag{2.24}$$

Moreover, there exists a constant $c_\tau > 0$ that depends only on τ such that, given $0 < \varepsilon < \frac{1}{2}$, we have $|\tilde{\mathcal{E}}_r| < \varepsilon$ for all $r \ge c_\tau(\mathsf{k}(\delta_{\mathbf{x}} + \delta_{\mathbf{y}}) - \log(\varepsilon))$. It is also clear that $\|\tilde{U}\|_{\max} = \|\tilde{V}\|_{\max} = 1$ since $|f_p(\xi)| = |e^{-\mathrm{i}|\xi|\cos(\frac{2\pi p}{2r+1} - \theta_\xi)}| = 1$ for all $\xi \in \mathbb{C}$.

While the previous result suggests that we can theoretically pick r large enough to obtain any desired accuracy, in practice, this may not be feasible. The entries of $\tilde{B}$ in (2.23) rapidly diverge, making them very challenging to compute accurately with double-precision floating-point arithmetic. However, it is possible to find an upper bound for the entries of $\tilde{B}$ if we assume that the rank r is less than a certain quantity. The following lemma gives an upper bound for such entries, which is a direct consequence of item (i) of Lemma 2.1 and will be used in our later stability studies.

LEMMA 2.8. *If $r \le \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|$, then for all $1 \le p \le 2r+1$,*

$$|\tilde{b}_{p,p}| \le \frac{1}{2r+1} \sum_{k=-r}^{r} |H_k(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|)| \le \sqrt{\frac{4}{\pi}}.$$

The previous lemma indicates a tradeoff between ensuring the entries are bounded and achieving a desired level of accuracy. To ensure that the entries are bounded, we need to restrict the rank r, which in turn limits the accuracy of the high-frequency expansion. Finally, we want to point out that there is actually a link between the low-rank factors U, V discussed in subsection 2.2 and the low-rank factors $\tilde{U}, \tilde{V}$ in the current subsection. See [5, Proposition 4.6] for a discussion on the transformation from a Bessel function to a plane wave, and vice versa.

3. Translation relations for the 2D Helmholtz kernel. We then discuss translation relations for the U, V basis matrices. These translation relations are crucial in accelerating the computational speed of the FMM. For the sake of presentation, we rewrite U in (2.18) as $U_{\mathbf{x}}$, where the subscript $\mathbf{x}$ indicates the dependence on the set $\mathbf{x}$. We intend to derive translation relations between $U_{\mathbf{x}}$ and $U_{\mathbf{x}'}$ for $\mathbf{x}' \subset \mathbf{x}$. Suppose $\mathbf{x}'$ has center $o_{\mathbf{x}'}$ and radius $\delta_{\mathbf{x}'}$, while $\mathbf{x}$ has center $o_{\mathbf{x}}$ and radius $\delta_{\mathbf{x}}$ so that the associated disks $\mathcal{D}_{\mathbf{x}'}$ and $\mathcal{D}_{\mathbf{x}}$ satisfy $\mathcal{D}_{\mathbf{x}'} \subset \mathcal{D}_{\mathbf{x}}$, where

$$\mathcal{D}_{\mathbf{x}'} := \{z \in \mathbb{C} : |z - o_{\mathbf{x}'}| \le \delta_{\mathbf{x}'}\} \quad \text{and} \quad \mathcal{D}_{\mathbf{x}} := \{z \in \mathbb{C} : |z - o_{\mathbf{x}}| \le \delta_{\mathbf{x}}\}. \tag{3.1}$$

3.1. Stable translation relation for the low-frequency regime. We start by presenting the stable translation relation corresponding to our stable expansion of the 2D Helmholtz kernel in the low-frequency regime. The original version of this translation relation was first discussed in [28]. In contrast to the original translation relation, ours explicitly incorporates scaling factors.

THEOREM 3.1. *Suppose $\mathbf{x}' \subset \mathbf{x}$ and $\mathcal{D}_{\mathbf{x}'} \subset \mathcal{D}_{\mathbf{x}}$, where $\mathcal{D}_{\mathbf{x}'}$ and $\mathcal{D}_{\mathbf{x}}$ are defined in (3.1) and suppose there exists a constant $\beta \in (0,1)$ such that for all $\mathbf{x}' \subset \mathbf{x}$,*

$$\max\{\delta_{\mathbf{x}'}, |o_{\mathbf{x}} - o_{\mathbf{x}'}|\} \le \beta \delta_{\mathbf{x}}. \tag{3.2}$$

Let $\lambda_{\mathbf{x},p}, \lambda_{\mathbf{x}',p}$ be scaling factors as defined in (2.17) and $g_j(\xi) := J_j(|\xi|)e^{\mathrm{i}j\theta_\xi}$ for all $\xi \in \mathbb{C}$. Let $r \ge \max\{e\mathsf{k}\delta_{\mathbf{x}}/2, 1\}$ such that $(\mathsf{k}\delta_{\mathbf{x}})^{l-r} \le l!/r!$ for all $l \ge r$. Then, the following translation relation holds:

$$U_{\mathbf{x}',\mathbf{x}} = U_{\mathbf{x}'} T_{\mathbf{x}',\mathbf{x}} + \mathcal{F}_r, \tag{3.3}$$

where

$$U_{\mathbf{x}',\mathbf{x}} := [g_j(\mathsf{k}(x - o_{\mathbf{x}}))\lambda_{\mathbf{x},j}]_{x\in\mathbf{x}', -r\le j\le r}, \quad U_{\mathbf{x}'} := [g_j(\mathsf{k}(x - o_{\mathbf{x}'}))\lambda_{\mathbf{x}',j}]_{x\in\mathbf{x}', -r\le j\le r},$$

$$T_{\mathbf{x}',\mathbf{x}} := [t_{i,j}]_{-r\le i,j\le r} \quad \textit{with}$$

$$t_{i,j} := \begin{cases} \lambda_{\mathbf{x},j} g_{i-j}(\mathsf{k}(o_{\mathbf{x}} - o_{\mathbf{x}'}))\lambda_{\mathbf{x}',i}^{-1}, & |i-j| \le r, \\ 0, & |i-j| > r, \end{cases}$$

$$\mathcal{F}_r := [\lambda_{\mathbf{x},j} E_{r+1}^{J_j}(\mathsf{k}|x - o_{\mathbf{x}'}|, \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)]_{x\in\mathbf{x}', -r\le j\le r}.$$

Moreover,

$$\|\mathcal{F}_r\|_{\max} \leq \beta^{r+1}\frac{4\mathsf{k}\delta_{\mathbf{x}}}{(r+1)}, \quad \|T_{\mathbf{x}',\mathbf{x}}\|_{\max} \leq 1, \quad \textit{and} \quad \|T_{\mathbf{x}',\mathbf{x}}\|_1 \leq 2r+1. \tag{3.4}$$

Proof. The relation in (3.3) follows from Graf's addition formula (2.7). We show the bound for $\|\mathcal{F}_r\|_{\max}$. By (3.2), we have $\beta\delta_{\mathbf{x}} \geq \max\{|x - o_{\mathbf{x}'}|, |o_{\mathbf{x}} - o_{\mathbf{x}'}|\}$ for all $x \in \mathbf{x}'$. Letting $z_{\max} = \beta\delta_{\mathbf{x}}$ in Lemma 2.3, we have

$$\left|E^{J_p}_{r+1}(\mathsf{k}|x - o_{\mathbf{x}'}|, \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)\right| \leq \frac{8}{(r+1)!}\left(\frac{\mathsf{k}\beta\delta_{\mathbf{x}}}{2}\right)^{r+1}, \quad x \in \mathbf{x}' \tag{3.5}$$

for all $r \geq \mathsf{k}\delta_{\mathbf{x}}\beta$ such that $(\mathsf{k}\delta_{\mathbf{x}}\beta)^{l-r} \leq l!/r!$ for $l \geq r$. Combining the above inequalities with the monotonicity of the scaling factors and assuming that $r \geq \max\{e\mathsf{k}\delta_{\mathbf{x}}/2, 1\}$, we have

$$\begin{aligned}
\|\mathcal{F}_r\|_{\max} &= \lambda_{\mathbf{x},i}\left|E^{J_i}_{r+1}(\mathsf{k}|x - o_{\mathbf{x}'}|, \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)\right| \leq \lambda_{\mathbf{x},r}\left|E^{J_i}_{r+1}(\mathsf{k}|x - o_{\mathbf{x}'}|, \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)\right| \\
&\leq \max\left\{\left|E^{J_i}_{r+1}(\mathsf{k}|x - o_{\mathbf{x}'}|, \mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)\right|, \frac{8}{(r+1)!}\left(\frac{\mathsf{k}\beta\delta_{\mathbf{x}}}{2}\right)^{r+1} r!\left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^{r}\right\} \\
&\leq \max\left\{\frac{8}{\sqrt{2\pi(r+1)}}\left(\frac{e\mathsf{k}\beta\delta_{\mathbf{x}}}{2(r+1)}\right)^{r+1}, \frac{4\mathsf{k}\beta^{r+1}\delta_{\mathbf{x}}}{(r+1)}\right\} \\
&\leq \beta^{r+1}\max\left\{\frac{8}{\sqrt{2\pi(r+1)}}\left(\frac{e\mathsf{k}\delta_{\mathbf{x}}}{2(r+1)}\right)^{r+1}, \frac{4\mathsf{k}\delta_{\mathbf{x}}}{r+1}\right\} \\
&\leq \beta^{r+1}\frac{4\mathsf{k}\delta_{\mathbf{x}}}{(r+1)}\max\left\{\left(\frac{e\mathsf{k}\delta_{\mathbf{x}}}{2(r+1)}\right)^{r}, 1\right\} \leq \beta^{r+1}\frac{4\mathsf{k}\delta_{\mathbf{x}}}{(r+1)},
\end{aligned}$$

where we applied Stirling's inequality to (3.5) in the third inequality.

Then look at $\|\mathcal{F}_r\|_{\max}$. By Lemma 2.6, we have $\lambda_{\mathbf{x},|i|} \leq \lambda_{\mathbf{x},|i|+|j-i|}$. To find an upper bound for $\|T_{\mathbf{x}',\mathbf{x}}\|_{\max}$, we observe that

$$\begin{aligned}
&\|T_{\mathbf{x}',\mathbf{x}}\|_{\max} = |t_{i,j}| \leq \lambda_{\mathbf{x},|i|+|j-i|}\,|J_{j-i}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)|\,\lambda^{-1}_{\mathbf{x}',|i|} \\
&= \max\left\{\frac{|J_{j-i}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)|}{\lambda_{\mathbf{x}',|i|}}, (|i| + |j-i|)!\left(\frac{2}{\mathsf{k}\delta_{\mathbf{x}}}\right)^{|i|+|j-i|}\frac{|J_{j-i}(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{x}'}|)|}{\lambda_{\mathbf{x}',|i|}}\right\} \\
&\leq \max\left\{\frac{1}{\lambda_{\mathbf{x}',|i|}}, \binom{|j-i| + |i|}{|i|}\left(\frac{|o_{\mathbf{x}} - o_{\mathbf{x}'}|}{\delta_{\mathbf{x}}}\right)^{|j-i|}\left(\frac{\delta_{\mathbf{x}'}}{\delta_{\mathbf{x}}}\right)^{|i|}\right\} \leq 1,
\end{aligned}$$

where we use $|J_p(z)| \leq (z/2)^p/p!$ for all $p \in \mathbb{N} \cup \{0\}$ and $z \geq 0$ to reach the last line. ∎

In this (approximate) translation relation, the norm of the error term decays exponentially with respect to r. Note that the relation (3.2) means that $\delta_{\mathbf{x}'}/\delta_{\mathbf{x}}$ or $|o_{\mathbf{x}} - o_{\mathbf{x}'}|/\delta_{\mathbf{x}}$ is not too close to 1. This is typically the case for situations that are not too extreme so that the FMM can be used to accelerate the computations. Then proper hierarchical paritioning of the point sets can be used to make sure (3.2) holds.

3.2. Translation relation for the high-frequency regime. Next, we state the translation relation corresponding to the expansion of the 2D Helmholtz kernel for the high-frequency regime, which is given in [4, Theorem 3.5] or [5, (4.8)]. As before, we assume that $\mathbf{x}' \subset \mathbf{x}$ and $\mathcal{D}_{\mathbf{x}'} \subset \mathcal{D}_{\mathbf{x}}$, where $\mathcal{D}_{\mathbf{x}'}$ and $\mathcal{D}_{\mathbf{x}}$ are defined in (3.1). This relation is derived from the fact that $f_j(\xi)$ defined in (2.24) can be written as:

$$f_j(\xi) = e^{-\mathrm{i}(\Re(\xi),\Im(\xi))\cdot(\cos\left(\frac{2\pi j}{(2r+1)}\right),\sin\left(\frac{2\pi j}{(2r+1)}\right))}, \quad 1 \le j \le 2r+1.$$

This can be used to immediately deduce that, for all $x \in \mathbf{x}'$,

$$f_j(\mathsf{k}(x - o_{\mathbf{x}})) = f_j(\mathsf{k}(x - o_{\mathbf{x}'}))f_j(\mathsf{k}(o_{\mathbf{x}'} - o_{\mathbf{x}})).$$

For simplicity of presentation, we assume that the expansion order r is fixed for all levels. (The variable expansion order case is discussed in [4, Theorem 3.5] or [5, (4.8)].) Then, the following translation relation holds:

$$\tilde{U}_{\mathbf{x}',\mathbf{x}} = \tilde{U}_{\mathbf{x}'}\tilde{T}_{\mathbf{x}',\mathbf{x}}, \tag{3.6}$$

where

$$\tilde{U}_{\mathbf{x}',\mathbf{x}} := [f_j(\mathsf{k}(x - o_{\mathbf{x}}))]_{x\in\mathbf{x}',1\le j\le 2r+1}, \quad \tilde{U}_{\mathbf{x}'} := [f_j(\mathsf{k}(x - o_{\mathbf{x}'}))]_{x\in\mathbf{x}',1\le j\le 2r+1},$$
$$\tilde{T}_{\mathbf{x}',\mathbf{x}} := \mathrm{diag}(\tilde{t}_{1,1}, \ldots, \tilde{t}_{2r+1,2r+1}) \quad \text{with} \quad \tilde{t}_{j,j} := f_j(\mathsf{k}(o_{\mathbf{x}'} - o_{\mathbf{x}})).$$

Moreover, the entries of $\tilde{T}_{\mathbf{x}',\mathbf{x}}$ obviously satisfy $\|\tilde{T}_{\mathbf{x}',\mathbf{x}}\|_1 = 1$. In addition, if we further assume that $\mathbf{x}'' \subset \mathbf{x}'$ and $\mathcal{D}_{\mathbf{x}''} \subset \mathcal{D}_{\mathbf{x}'}$, where $\mathcal{D}_{\mathbf{x}''}$ is defined similarly as in (2.24), then the following translation relation can be directly deduced: $\tilde{T}_{\mathbf{x}'',\mathbf{x}} = \tilde{T}_{\mathbf{x}'',\mathbf{x}'}\tilde{T}_{\mathbf{x}',\mathbf{x}}$.

4. Stable matrix version of the wideband FMM.

Based on results in the previous sections, we describe a stable matrix version of the wideband FMM for the 2D Helmholtz kernel. We also discuss some recurrence relations for quickly and stably computing certain generators. A detailed discussion of the matrix version for the basic 2D FMM can be found [26, Section 4]. The matrix version of the wideband FMM for the 2D Helmholtz kernel is slightly more involved, since we may use different expansions depending on the levels of the hierarchical partitioning of the point sets.

4.1. Wideband FMM matrix.

In the following discussion, we follow the notation in [26, Section 4]. Without loss of generality, suppose the point sets $\mathbf{X}, \mathbf{Y}$ (at the beginning of section 1) are located within a square domain which is hierarchically partitioned into L levels following a quadtree so that the leaf-level subdomains contain $O(1)$ points. Suppose the root level is at level $l = 0$ and the leaves are at level $l = L$. As discussed earlier, we may set a *switching level* $\tilde{L}$ such that for all levels $l \ge \tilde{L}$ we use the degenerate expansion in subsection 2.2 and the translation relation in subsection 3.1. Otherwise, we use those in subsections 2.3 and 3.2. One way to determine the switching level is by setting a condition on the product of the wavenumber k and the size of a partition box at a given level [11, 12, 14]. The numerical experiments in [11, Table 1] and the discussion in [12, Section 3] suggest that this switching level also depends on the target accuracy and our desire to control the entrywise magnitudes of $\tilde{B}$ (see the discussion related to Lemma 2.8).

In the hierarchical partitioning, subsets that satisfy (2.1) correspond to far-field blocks in the kernel matrix K in (1.1). The other blocks of K are categorized as near-field blocks and are stored in dense forms. Far-field blocks can be approximated by low-rank forms following the two types of degenerate expansions of $H_0(\mathsf{k}|x - y|)$ discussed before. Following the terminology in the matrix version FMM in [26], each subset $\mathbf{x} \subset \mathbf{X}$ at level $l > \tilde{L}$ *contributes* a basis matrix U. The U basis matrices at the leaf level are directly obtained as in Theorem 2.7, and those at levels $L-1, \ldots, \tilde{L}+1$ are implicitly available through the translation operators T as in Theorem 3.1. Similarly, we can obtain the V factors and their translation operators. Theorem 2.7 also gives the factors B. These factors can be used to approximate any far-field block corresponding to the low-frequency regime and are called *generators* of the FMM matrix.

For the high-frequency regime, although the generators $\tilde{U}, \tilde{V}$ in (2.22) are only first used at level $\tilde{L}$, their row sizes may be large for them to be efficient. Thus, we form the basis contribution $\tilde{U}$ in (2.22) at level L and use the translation operators $\tilde{T}$ in (3.6) to implicitly produce upper level basis generators $\tilde{U}$. Similarly, produce the leaf-level $\tilde{V}$ generators and associated translation operators. Also, use (2.23) to obtain the $\tilde{B}$ generators (only for levels $\tilde{L}, \ldots, 2$). (Note that at level $l = 1$ the subsets are all near-field neighbors.)

With all these generators, K in (1.1) can be decomposed as

$$K = K^{(0)} + \sum_{l=2}^{\tilde{L}-1} \tilde{U}^{(l)} \tilde{B}^{(l)} (\tilde{V}^{(l)})^{\mathsf{T}} + \sum_{l=\tilde{L}}^{L} U^{(l)} B^{(l)} (V^{(l)})^{\mathsf{T}} + \mathcal{E}, \tag{4.1}$$

where

- $K^{(0)}$ corresponds to the near-field blocks,
- $\tilde{U}^{(l)}$, $\tilde{B}^{(l)}$, $\tilde{V}^{(l)}$ are block diagonal matrices containing low-rank factors that make up each far-field block at level l using the expansion for the high-frequency regime discussed in subsection 2.3,
- $U^{(l)}$, $B^{(l)}$, $V^{(l)}$ are block diagonal matrices containing low-rank factors that make up each far-field block at level l using the expansion for the low-frequency regime discussed in subsection 2.2, and
- $\mathcal{E} := \sum_{l=2}^{\tilde{L}-1} \tilde{\mathcal{E}}^{(l)} + \sum_{l=\tilde{L}}^{L} \mathcal{E}^{(l)}$ with $\tilde{\mathcal{E}}^{(l)}$ and $\mathcal{E}^{(l)}$ error matrices from the low-rank approximations and translation relations.

Note that the basis matrices $\tilde{U}^{(l)}$, $\tilde{V}^{(l)}$, $U^{(l)}$, and $V^{(l)}$ satisfy the following nested relations because of the translation relations:

$$\tilde{U}^{(l)} = \tilde{U}^{(l+1)} \tilde{R}^{(l)}, \quad \tilde{V}^{(l)} = \tilde{V}^{(l+1)} \tilde{R}^{(l)}, \quad U^{(l)} = U^{(l+1)} R^{(l)}, \quad V^{(l)} = V^{(l+1)} R^{(l)},$$

where $R^{(l)}, \tilde{R}^{(l)}$ are block diagonal matrices such that each diagonal block contains a translation generator associated with a node at level l. Similar to [26, (4.7)], we can use these relations to rewrite (4.1) as $K = \tilde{K} + \mathcal{E}$, where

(4.2)
$$\begin{aligned}\tilde{K} = K^{(0)} &+ \tilde{U}^{(L)} \tilde{R}^{(L-1)} \cdots \tilde{R}^{(\tilde{L}-1)} \Big(\tilde{R}^{(\tilde{L}-2)} \Big(\cdots \big(\tilde{R}^{(2)} \tilde{B}^{(2)} (\tilde{R}^{(2)})^{\mathsf{T}} + \tilde{B}^{(3)} \big) \cdots \Big) \\ &\cdot (\tilde{R}^{(\tilde{L}-2)})^{\mathsf{T}} + \tilde{B}^{(\tilde{L}-1)} \Big) (\tilde{R}^{(\tilde{L}-1)})^{\mathsf{T}} \cdots (\tilde{R}^{(L-1)})^{\mathsf{T}} (\tilde{V}^{(L)})^{\mathsf{T}} \\ &+ U^{(L)} \Big(R^{(L-1)} \Big(\cdots \big(R^{(\tilde{L})} B^{(\tilde{L})} (R^{(\tilde{L})})^{\mathsf{T}} + B^{(\tilde{L}+1)} \big) \cdots \Big) (R^{(L-1)})^{\mathsf{T}} + B^{(L)} \Big) (V^{(L)})^{\mathsf{T}}.\end{aligned}$$

$\tilde{K}$ is the *wideband FMM matrix*, which is the approximation to K as produced by the wideband FMM. Like in [26], here the $U, V, \tilde{U}, \tilde{V}$ generators can be simply understood as the basis contributions from relevant point subsets, the $B, \tilde{B}$ generators reflect translations between basis contributions of far-field subsets, and $R, \tilde{R}$ generators reflect translations between parent and child basis contributions. This avoids the need to distinguish multipole and local expansions in the usual FMM and makes it convenient to understand the wideband FMM.

We can determine the total error of the wideband FMM matrix by using a similar argument used in the backward stability analysis (Theorem 5.2) and taking into account of the errors coming from the low-rank approximations in both frequency regimes and the translation relations in the low-frequency regime.

The wideband FMM matrix $\tilde{K}$ can be multiplied with a vector q quickly with $O(r(M+N))$ complexity when we choose $\tilde{r} = r$ and set the number of points in each finest-level subset to be $O(r)$. It simply follows the structured matrix-vector multiplications that have been well studied for various kinds of hierarchical structured matrices [7, 10, 20, 21, 35]. For convenience, we show the multiplication algorithm in Algorithm 4.1 so as to facilitate our later stability analysis.

Algorithm 4.1 Wideband FMM matrix-vector multiplication $\phi(=Kq) \approx \tilde{K}q$

$\tilde{v}^{(L)} \leftarrow (\tilde{V}^{(L)})^\mathsf{T} q$
for level $l = L-1, \ldots, \tilde{L}-1$ **do**
  $\tilde{v}^{(l)} \leftarrow (\tilde{R}^{(l)})^\mathsf{T} \tilde{v}^{(l+1)}$
end for
$\tilde{t}^{(\tilde{L}-1)} \leftarrow \tilde{B}^{(\tilde{L}-1)} \tilde{v}^{(\tilde{L}-1)}$
for level $l = \tilde{L}-2, \ldots, 2$ **do** ▷ *Bottom-up traversal using subsections 2.3 and 3.2*
  $\tilde{v}^{(l)} \leftarrow (\tilde{R}^{(l)})^\mathsf{T} \tilde{v}^{(l+1)}$
  $\tilde{t}^{(l)} \leftarrow \tilde{B}^{(l)} \tilde{v}^{(l)}$
end for
$\tilde{u}^{(2)} \leftarrow \tilde{t}^{(2)}$
for level $l = 3, \ldots, \tilde{L}-1$ **do** ▷ *Top-down traversal using subsections 2.3 and 3.2*
  $\tilde{u}^{(l)} \leftarrow \tilde{R}^{(l-1)} \tilde{u}^{(l-1)} + \tilde{t}^{(l)}$
end for
for level $l = \tilde{L}, \ldots, L$ **do**
  $\tilde{u}^{(l)} \leftarrow \tilde{R}^{(l-1)} \tilde{u}^{(l-1)}$
end for
$\phi_1 \leftarrow \tilde{U}^{(L)} \tilde{u}^{(L)}$
$v^{(L)} \leftarrow (V^{(L)})^\mathsf{T} q$
$t^{(L)} \leftarrow B^{(L)} v^{(L)}$
for level $l = L-1, \ldots, \tilde{L}$ **do** ▷ *Bottom-up traversal using subsections 2.2 and 3.1*
  $v^{(l)} \leftarrow \left(R^{(l)}\right)^\mathsf{T} v^{(l+1)}$
  $t^{(l)} \leftarrow B^{(l)} v^{(l)}$
end for
$u^{(\tilde{L})} \leftarrow t^{(\tilde{L})}$
for level $l = \tilde{L}+1, \ldots, L$ **do** ▷ *Top-down traversal using subsections 2.2 and 3.1*
  $u^{(l)} \leftarrow R^{(l-1)} u^{(l-1)} + t^{(l)}$
end for
$\phi_2 \leftarrow U^{(L)} u^{(L)} + K^{(0)} q$
$\phi \leftarrow \phi_1 + \phi_2$ ▷ *Evaluation*

4.2. Recurrence relations for computing generators. We provide some details on how to quickly and stably compute the entries of low-rank factors and translation relations associated with our stable expansion for the low-frequency regime. For the high-frequency regime, the entries of $\tilde{U}$ and $\tilde{V}$ are plane waves with a modulus of one, so there is no risk of running into overflow issues. As discussed earlier, the choice of the switching level may depend on the accuracy, which is implicitly influenced by how well we can compute the entries of $\tilde{B}$. In the wideband FMM algorithm, we intentionally set the switching level to ensure the entries of $\tilde{B}$ are computed accurately, avoiding overflow or stability issues that cannot be corrected by our stabilization strategy. Thus, we focus on the generators from the low-frequency regime.

Theorems 2.7 and 3.1 state that the matrices U, B, V, and the translation factors T satisfy some norm bounds. The entries of these factors can often be computed accurately and in a stable manner by direct calculations. However, certain situations may arise, where direct calculations become problematic. For example, when we have to multiply an extremely large number with an extremely small number, even though the result is relatively moderate in magnitude. For that reason, we present the following relations that can compute the entries of U, B, V in stably and efficiently.

Recall that the following recurrence relation holds for[1, (9.1.27)]:

$$Z_{p+1} + Z_{p-1} = \frac{2p}{z} Z_p, \quad p \in \mathbb{N} \cup \{0\}, \quad z > 0, \quad Z_p \in \{J_p, Y_p, H_p\}$$

The recurrence relation for J_p is backward stable and can be computed by the Miller's recurrence algorithm (e.g., see [19] and references therein). In the forward direction, the recurrence relation for J_p is numerically unstable. On the other hand, the recurrence relation for Y_p is forward stable. Given these considerations, the following relations can be used to quickly and stably compute the entries of U, B, V, T.

We first discuss the recurrence relations for the entries of U. For $1 \le i \le m$, let

$$\begin{aligned}
\tilde{\psi}_{i,r+1} &= 0, \quad \tilde{\psi}_{i,r} = 1, \\
\tilde{\psi}_{i,j} &= \tfrac{\lambda_{\mathbf{x},j}}{\lambda_{\mathbf{x},j+1}} \left(\tfrac{2(j+1)}{\mathsf{k}|x_i - o_{\mathbf{x}}|} \tilde{\psi}_{i,j+1} - \tfrac{\lambda_{\mathbf{x},j+1}}{\lambda_{\mathbf{x},j+2}} \tilde{\psi}_{i,j+2} \right), \quad 0 \le j \le r-1, \\
\psi_{i,j} &= J_0(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) \tilde{\psi}_{i,0}^{-1} \tilde{\psi}_{i,j}, \\
\psi_{i,-j} &= (-1)^j \psi_{i,j}, \quad 1 \le j \le r.
\end{aligned}$$

Then, $U = [u_{i,j}]_{1 \le i \le m, -r \le j \le r} = [e^{\mathrm{i} j \theta_{\mathsf{k}(x_i - o_{\mathbf{x}})}} \psi_{i,j}]_{1 \le i \le m, -r \le j \le r}$. Note that we have used the underlying symmetry of the entries in the last line. The entries of V can be computed similarly.

Next, we discuss the recurrence relations for the entries of B. Let

$$\begin{aligned}
\nu_{0,0} &= Y_0(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|), \quad \nu_{1,0} = \lambda_1^{-1} Y_1(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|), \\
\nu_{i,0} &= \tfrac{\lambda_{\mathbf{x},i-1}}{\lambda_{\mathbf{x},i}} \left(\tfrac{2i}{\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|} \nu_{i-1,0} - \tfrac{\lambda_{\mathbf{x},i-2}}{\lambda_{\mathbf{x},i-1}} \nu_{i-2,0} \right), \quad 2 \le i \le r, \\
\nu_{-i,0} &= (-1)^i \nu_{i,0}, \quad 1 \le i \le r, \\
\nu_{i-1,j+1} &= \tfrac{\lambda_{\mathbf{x},i}}{\lambda_{\mathbf{x},i-1}} \tfrac{\lambda_{\mathbf{y},j}}{\lambda_{\mathbf{y},j+1}} \nu_{i,j}, \quad 0 \le j \le r-1, \quad |i+j| \le r, \\
\nu_{-i,-j} &= (-1)^{i+j} \nu_{i,j}, \quad 0 \le j \le r-1, \quad |i+j| \le r,
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\mu}_{r+1,0} &= 0, \quad \tilde{\mu}_{r,0} = 1, \\
\tilde{\mu}_{i,0} &= \tfrac{\lambda_{\mathbf{x},i+1}}{\lambda_{\mathbf{x},i}} \left(\tfrac{2(i+1)}{\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|} \tilde{\mu}_{i+1,0} - \tfrac{\lambda_{\mathbf{x},i+2}}{\lambda_{\mathbf{x},i+1}} \tilde{\mu}_{i+2,0} \right), \quad 0 \le i \le r-1, \\
\mu_{i,0} &= J_0(\mathsf{k}|o_{\mathbf{x}} - o_{\mathbf{y}}|) \tilde{\mu}_{0,0}^{-1} \tilde{\mu}_{i,0}, \\
\mu_{-i,0} &= (-1)^i \mu_{i,0}, \quad 1 \le i \le r, \\
\mu_{i-1,j+1} &= \tfrac{\lambda_{\mathbf{x},i}}{\lambda_{\mathbf{x},i-1}} \tfrac{\lambda_{\mathbf{y},j}}{\lambda_{\mathbf{y},j+1}} \mu_{i,j}, \quad 0 \le i \le r-1, \quad |i+j| \le r, \\
\mu_{-i,-j} &= (-1)^{i+j} \mu_{i,j}, \quad 0 \le i \le r-1, \quad |i+j| \le r.
\end{aligned} \tag{4.3}$$

Then $B = [b_{i,j}]_{-r\le i,j\le r} = [(-1)^j e^{-\mathrm{i}(i+j)\theta_{\mathrm{k}(o_{\mathbf{y}}-o_{\mathbf{x}})}}(\mu_{i,j}+\mathrm{i}\nu_{i,j})]_{-r\le i,j\le r}$. Once again, we note that we have used the underlying symmetry of the entries to obtain $\nu_{-i,-j}$ and $\mu_{-i,-j}$ for all $0 \le i \le r-1$ and $|i+j| \le r$.

Finally, the entries of T can be computed the same way as (4.3) by changing the line for $\mu_{i-1,j+1}$ to

$$\mu_{i-1,j+1} = \tfrac{\lambda_{\mathbf{x},i}}{\lambda_{\mathbf{x},i-1}}\tfrac{\lambda_{\mathbf{y},j+1}}{\lambda_{\mathbf{y},j}}\mu_{i,j}, \quad 0 \le i \le r-1, \quad |i+j| \le r.$$

Then, $T = [t_{i,j}]_{-r\le i,j\le r} = [e^{-\mathrm{i}(i-j)\theta_{\mathrm{k}(o_{\mathbf{y}}-o_{\mathbf{x}})}}\mu_{i,-j}]_{-r\le i,j\le r}$.

5. Backward stability analysis. Now, we study the backward stability of the wideband FMM matrix-vector multiplication Algorithm 4.1 based on the norm bounds presented in subsections 2.2, 2.3, 3.1, and 3.2.

Assumption 5.1. For convenience, the backward stability analysis utilizes the following notation and assumptions.

- For simplicity, suppose $|\mathbf{X}| = |\mathbf{Y}| = N$ for $\mathbf{X}, \mathbf{Y}$ in (1.1).
- The FMM tree $\mathcal{T}$ is a full quadtree with L levels such that there are 4^l nodes at level l for $0 \le l \le L$.
- Each node $\mathbf{i} \in \mathcal{T}$ is associated with some generators of the FMM matrix in (4.2). For convenience, use $U_{\mathbf{i}}$, $V_{\mathbf{i}}$, $R_{\mathbf{i}}$ etc. to denote the generators associated with $\mathbf{i}$ at level l, where $R_{\mathbf{i}}$ is the translation generator and is a diagonal block of $R^{(l)}$ in (4.2). The symbols $\tilde{U}_{\mathbf{i}}$, $\tilde{V}_{\mathbf{i}}$, $\tilde{R}_{\mathbf{i}}$ etc. can be similarly understood.
- For each node $\mathbf{i} \in \mathcal{T}$, denote its level in $\mathcal{T}$ by $\mathrm{lv}(\mathbf{i})$. If $\tilde{L} \le \mathrm{lv}(\mathbf{i}) \le L$, $U_{\mathbf{i}}$ and $V_{\mathbf{i}}$ have column sizes $2r+1$ (see (2.18)). Similarly, $\tilde{U}_{\mathbf{i}}$ and $\tilde{V}_{\mathbf{i}}$ for $2 \le \mathrm{lv}(\mathbf{i}) \le \tilde{L}-1$ are assumed to have column sizes $2\tilde{r}+1$ (see (2.22)). Each $R_{\mathbf{i}}$ is $(2r+1)\times(2r+1)$ and each $\tilde{R}_{\mathbf{i}}$ is $(2\tilde{r}+1)\times(2\tilde{r}+1)$. If $\mathrm{lv}(\mathbf{i}) = L$, $U_{\mathbf{i}}$ and $V_{\mathbf{i}}$ have row sizes $|x_{\mathbf{i}}| = |y_{\mathbf{i}}| = N_0 = O(1)$.
- The sparse block matrix $K^{(0)}$ has a maximum of w columns in any of nonzero near-field blocks. w is related to the maximum number of points in the leaf nodes at level L.
- For each level where the expansion in subsection 2.2 is used, the assumptions in Lemma 2.8 hold and $\tilde{r}$ is chosen accordingly. This assumption is related to the switching level $\tilde{L}$ used in the algorithm. For the other levels, the assumptions in Theorem 3.1 hold and r is chosen accordingly.
- The inequality $3\max\{\gamma_{2r+1}, \gamma_{2\tilde{r}+1}, \gamma_{N_0}\}\log_2(N) \le 1/2$ holds, where
$$\gamma_n := \frac{n\epsilon_{\mathrm{mach}}}{1-n\epsilon_{\mathrm{mach}}}, \quad (\epsilon_{\mathrm{mach}}\text{: machine epsilon}).$$
- The error matrix $\mathcal{E}$ in (4.1) satisfies $|\mathcal{E}| \le \varepsilon$, where we assume $\varepsilon \ge \epsilon_{\mathrm{mach}}$ for convenience.

In addition, use $\mathrm{fl}(\cdot)$ to represent the result from a floating point operation. To simplify the presentation, the stability analysis also uses the standard multi-index notation like in [26]:

$$\boldsymbol{\alpha} := (\alpha_1, \dots, \alpha_n), \quad \alpha_i \in \{0,1\}, \quad 1 \le i \le n, \quad n \in \mathbb{N}.$$

Also, let $|\boldsymbol{\alpha}| := \alpha_1 + \cdots + \alpha_n$. For $k \ge j$, define

$$\left(\prod_{i=k}^{\ge j} A_i\right)^{\boldsymbol{\alpha}} := A_k^{\alpha_1}\dots A_j^{\alpha_{k-j+1}} \quad \text{and} \quad \Delta^{\boldsymbol{\alpha}}\left(\prod_{i=k}^{\ge j} A_i\right) := (\Delta^{\alpha_1}A_k)\dots(\Delta^{\alpha_{k-j+1}}A_j),$$

where $\Delta^0 A_i := A_i$ and $\Delta^1 A_i := \Delta A_i$ for $j \le i \le k$ and terms like ΔA_i appear in the perturbation analysis below.

THEOREM 5.2. *With Assumption 5.1, the wideband FMM (Algorithm 4.1) is backward stable and the multiplication of $\tilde{K}$ in (4.2) with a vector q satisfies*

$$\mathrm{fl}(\tilde{K}q) = (K + \Delta K)q \quad \textit{with}$$
$$\|\Delta K\|_{\max} \le \frac{54 \cdot 2^{11}}{\pi} \max\{\gamma_{2r+1}, \gamma_{2\tilde{r}+1}, \gamma_{N_0}\} \log_2(N) \mathsf{p}(r, \tilde{r}) \max\{1, \|K\|_{\max}\} + 7\gamma_w \|K\|_{\max} + 2\varepsilon,$$

where $\mathsf{p}(r, \tilde{r}) := (2\tilde{r}+1)^2 + (2r+1)^{2(L-\tilde{L}+2)}$.

Proof. The proof of the theorem is similar to the proof of [26, Theorem 5.8], except that the norm bounds of some translation matrices may be larger than 1, which leads to potential error growth that needs to be quantified. We follow the notation in [26] and apply a similar procedure twice (once for each expansion used in the FMM matrix). For each expansion, we carefully track the error propagation of the bottom-up and top-down traversals. Below, we give the main steps of the proof and leave out lengthy technical details of certain steps that can be referred to [26].

Lines 1–17 of Algorithm 4.1: We first consider the high-frequency expansion of the FMM algorithm. The floating operations in lines 1–4 of Algorithm 4.1 yield

$$(\tilde{R}^{(L-1)} + \Delta \tilde{R}^{(L-1)})^{\mathsf{T}} \dots (\tilde{R}^{(\tilde{L}-1)} + \Delta \tilde{R}^{(\tilde{L}-1)})^{\mathsf{T}} (\tilde{V}^{(L)} + \Delta \tilde{V}^{(L)})^{\mathsf{T}} q = (\tilde{V}^{(\tilde{L}-1)} + \Delta \tilde{V}^{(\tilde{L}-1)})^{\mathsf{T}} q,$$

where

$$\Delta \tilde{V}^{(\tilde{L}-1)} := (\Delta \tilde{V}^{(L)}) \left(\tilde{R}^{(L-1,\tilde{L}-1)}\right) + \sum_{|\boldsymbol{\alpha}|=1}^{L-\tilde{L}+1} (\tilde{V}^{(L)})(\Delta^{\boldsymbol{\alpha}}(\tilde{R}^{(L-1,\tilde{L}-1)})) + \sum_{|\boldsymbol{\alpha}|=1}^{L-\tilde{L}+1} (\Delta \tilde{V}^{(L)})(\Delta^{\boldsymbol{\alpha}}(\tilde{R}^{(L-1,\tilde{L}-1)})). \tag{5.1}$$

From subsection 3.2, we have $\|\tilde{T}_{\mathbf{x}',\mathbf{x}}\|_1 = \|\tilde{T}_{\mathbf{x}',\mathbf{x}}\|_\infty = 1$ for $\mathbf{x}' \subset \mathbf{x}$. Let $\mathbf{i}, \mathbf{j}$ be two leaf nodes at level L. The subscript below stands for a node at level L if it is attached to $\tilde{V}^{(\tilde{L}}$ or $\tilde{U}^{(\tilde{L})}$. If it is attached to $R^{(L-1,j)}$, then it stands for a product of $L-j$ matrices of size $(2r+1)\times(2r+1)$ (corresponding to the node indicated by the subscript), where the nth matrix of this product, $R_{\mathbf{i}}^{(L-n)}$, is a submatrix of $R^{(L-n)}$. We have

$$\|(\Delta \tilde{V}_{\mathbf{j}}^{(L)}) \left(\tilde{R}_{\mathbf{j}}^{(L-1,\tilde{L}-1)}\right)\|_\infty \le \gamma_{2\tilde{r}+1}(2\tilde{r}+1),$$
$$\left\|(\tilde{V}_{\mathbf{j}}^{(L)}) \sum_{|\boldsymbol{\alpha}|=1}^{L-\tilde{L}+1} (\Delta^{\boldsymbol{\alpha}}(\tilde{R}_{\mathbf{j}}^{(L-1,\tilde{L}-1)}))\right\|_\infty \le (2\tilde{r}+1)2\gamma_{2\tilde{r}+1} \log_2 N,$$
$$\left\|(\Delta \tilde{V}_{\mathbf{j}}^{(L)}) \sum_{|\boldsymbol{\alpha}|=1}^{L-\tilde{L}+1} (\Delta^{\boldsymbol{\alpha}}(\tilde{R}_{\mathbf{j}}^{(L-1,\tilde{L}-1)}))\right\|_\infty \le 2\gamma_{2\tilde{r}+1}^2(2\tilde{r}+1) \log_2 N,$$

where we used [26, (5.12)] for the last two estimates. It follows that

$$\|(\Delta \tilde{V}_{\mathbf{j}}^{(\tilde{L}-1)})^{\mathsf{T}}\|_1 = \|\Delta \tilde{V}_{\mathbf{j}}^{(\tilde{L}-1)}\|_\infty \le 6\gamma_{2\tilde{r}+1}(2\tilde{r}+1) \log_2 N.$$

For a given $\tilde{u}^{(\tilde{L}-1)}$ at the end of line (13) in Algorithm 4.1, line (14)-(17) yields

$$\begin{aligned}(\tilde{U}^{(L)} + \Delta\tilde{U}^{(L)}) &\left(\tilde{R}^{(L-1,\tilde{L}-1)} + \sum_{|\boldsymbol{\alpha}|=1}^{L-\tilde{L}+1} \Delta^{\boldsymbol{\alpha}}(\tilde{R}^{(L-1,\tilde{L}-1)})\right) \mathrm{fl}(\tilde{u}^{(\tilde{L}-1)}) \\ &= (\tilde{U}^{(\tilde{L}-1)} + \Delta\tilde{U}^{(\tilde{L}-1)})\mathrm{fl}(\tilde{u}^{(\tilde{L}-1)}),\end{aligned}$$

where $\Delta\tilde{U}^{(\tilde{L}-1)}$ is defined as in (5.1) with $\tilde{V}$ replaced by $\tilde{U}$. Moreover,

$$\|\Delta\tilde{U}_{\mathbf{i}}^{(\tilde{L}-1)}\|_\infty \le 6\gamma_{2\tilde{r}+1}(2\tilde{r}+1)\log_2 N.$$

Thus, we have $\mathrm{fl}(\phi_1) = (K_{\mathrm{high}} + \Delta K_{\mathrm{high}})$ and by the proof of [26, Lemma 5.7],

$$\begin{aligned}K_{\mathrm{high}} &:= \sum_{k=2}^{\tilde{L}-1} \tilde{U}^{(\tilde{L}-1)}\tilde{R}^{(\tilde{L}-2,k)}\tilde{B}^{(k)}(\tilde{R}^{(\tilde{L}-2,k)})^{\mathsf{T}}(\tilde{V}^{(\tilde{L}-1)})^{\mathsf{T}}, \\ \Delta K_{\mathrm{high}} &:= \sum_{k=2}^{\tilde{L}-1}\sum_{|\boldsymbol{\beta}|=1}^{6} (\Delta^{\beta_1}\tilde{U}^{(\tilde{L}-1)})(\Delta^{\beta_2}\tilde{M}^{(k)})(\Delta^{\beta_3}Z^{(k)})(\Delta^{\beta_4}\tilde{B}^{(k)})(\Delta^{\beta_5}\tilde{P}^{(k)})^{\mathsf{T}} \\ &\qquad \cdot (\Delta^{\beta_6}\tilde{V}^{(\tilde{L}-1)})^{\mathsf{T}}\end{aligned}$$

with $|\Delta Z^{(k)}| \le \epsilon_{\mathrm{mach}} I$, $\tilde{M}^{(k)} := \widehat{M}^{(\tilde{L}-2,k)}$, $\widehat{M}^{(j)} := \tilde{R}^{(j)}$, $\tilde{P}^{(k)} := \tilde{R}^{(\tilde{L}-2,k)}$,

$$\begin{aligned}\Delta\tilde{M}^{(k)} &:= \sum_{|\boldsymbol{\alpha}|=1}^{\tilde{L}-1-k} \Delta^{\boldsymbol{\alpha}}\left(\widehat{M}^{(\tilde{L}-2,k)}\right), \quad \Delta\tilde{P}^{(k)} := \sum_{|\boldsymbol{\alpha}|=1}^{\tilde{L}-1-k} \Delta^{\boldsymbol{\alpha}}\left(\tilde{R}^{(\tilde{L}-2,k)}\right), \\ \Delta\widehat{M}^{(j)} &:= \Delta\tilde{R}^{(j)} + \Delta Z^{(j+1)}\tilde{R}^{(j)} + \Delta Z^{(j+1)}\Delta\tilde{R}^{(j)}.\end{aligned}$$

Moreover, $\tilde{M}^{(k)}$, $\Delta\tilde{M}^{(k)}$, $\tilde{P}^{(k)}$, $\Delta\tilde{P}^{(k)}$ satisfy the following bounds by repeating the calculation done in [26, Lemma 5.6]:

$$\begin{aligned}&\|\tilde{M}_{\mathbf{i}}^{(k)}\|_1 \le 1, \quad \|\tilde{P}_{\mathbf{j}}^{(k)}\|_1 \le 1, \\ &\|\Delta\tilde{M}_{\mathbf{i}}^{(k)}\|_1 \le \sum_{|\boldsymbol{\alpha}|=1}^{\tilde{L}-1-k} \binom{\tilde{L}-1-k}{|\boldsymbol{\alpha}|}(3\gamma_{2\tilde{r}+1})^{|\boldsymbol{\alpha}|} \le 6\gamma_{2\tilde{r}+1}\log_2(N), \\ &\|(\Delta\tilde{P}_{\mathbf{j}}^{(k)})^{\mathsf{T}}\|_{\max} = \|\Delta\tilde{P}_{\mathbf{j}}^{(k)}\|_{\max} = \|\Delta\tilde{P}_{\mathbf{j}}^{(k)}\|_1 \le 2\gamma_{2\tilde{r}+1}\log_2(N).\end{aligned}$$

Using a similar argument as in [26, (5.19)-(5.20)], we have

$$|\Delta\tilde{K}_{\mathrm{high}}| \le \frac{2^7 \cdot 6}{\sqrt{\pi}}\gamma_{2\tilde{r}+1}\log_2(N)(2\tilde{r}+1)^2.$$

Lines 18–28 of Algorithm 4.1: Now, we deal with the low-frequency part which is coupled with the near-field evaluation. Following the proof of [26, Lemma 5.7], we have $\mathrm{fl}(\phi_2) = (\tilde{K}_{\mathrm{low}} + \Delta\tilde{K}_{\mathrm{low}})q$, where

$$\tilde{K}_{\mathrm{low}} := \sum_{k=\tilde{L}}^{L} U^{(L)}R^{(L-1,k)}B^{(k)}(R^{(L-1,k)})^{\mathsf{T}}(V^{(L)})^{\mathsf{T}} + K^{(0)},$$

$$\Delta\tilde{K}_{\text{low}} := \sum_{k=\tilde{L}}^{L}\sum_{|\boldsymbol{\beta}|=1}^{7}(\Delta^{\beta_1}H)(\Delta^{\beta_2}U^{(L)})(\Delta^{\beta_3}M^{(k)})(\Delta^{\beta_4}Z^{(k)})(\Delta^{\beta_5}B^{(k)})(\Delta^{\beta_6}P^{(k)})^{\mathsf{T}}$$
$$\cdot(\Delta^{\beta_7}V^{(L)})^{\mathsf{T}} + (\Delta K^{(0)} + \Delta H K^{(0)} + \Delta H\Delta K^{(0)}),$$

with $|\Delta H| \le \epsilon_{\text{mach}} I$, $M^{(k)} := \breve{M}^{(L-1,k)}$, $\breve{M}^{(j)} := R^{(j)}$, $P^{(k)} := R^{(L-1,k)}$,

$$\Delta M^{(k)} := \sum_{|\boldsymbol{\alpha}|=1}^{L-k}\Delta^{\boldsymbol{\alpha}}\left(\breve{M}^{(L-1,k)}\right), \quad \Delta P^{(k)} := \sum_{|\boldsymbol{\alpha}|=1}^{L-k}\Delta^{\boldsymbol{\alpha}}\left(R^{(L-1,k)}\right),$$
$$\Delta\breve{M}^{(j)} := \Delta R^{(j)} + \Delta Z^{(j+1)}R^{(j)} + \Delta Z^{(j+1)}\Delta R^{(j)}.$$

Let $\mathbf{i}, \mathbf{j}$ be two leaf nodes at level L, and let the subscripts $\mathbf{i}, \mathbf{j}$ carry the same meaning as before if they are attached to U, V, and R. In what follows, if the subscript is attached to $B^{(k)}$, then it stands for the multipole-to-local expansion at level k for the points in the nodes $\mathbf{i}$ and $\mathbf{j}$. The above matrices satisfy the following estimates (which can be obtained by using the same procedure as in [26, Lemmas 5.5-5.7] and taking into account of the norm bounds in subsection 3.1)

$$\|\Delta^{\beta_1}H_{\mathbf{i}}\|_\infty \le \epsilon_{\text{mach}}^{\beta_1}, \quad \|\Delta^{\beta_2}U_{\mathbf{i}}^{(L)}\|_\infty \le \gamma_{2r+1}^{\beta_2}(2r+1),$$
$$\|\Delta^{\beta_3}M_{\mathbf{i}}^{(k)}\|_1 \le (6\gamma_{2r+1}\log_2 N)^{\beta_3}(2r+1)^{L-k},$$
$$\|\Delta^{\beta_4}Z_{\mathbf{i}}^{(k)}\|_1 \le \epsilon_{\text{mach}}^{\beta_4}, \quad \|\Delta^{\beta_5}B_{\mathbf{i},\mathbf{j}}^{(k)}\|_{1,1} \le \gamma_{2r+1}^{\beta_5}\left(\frac{72}{\pi}r^2\max\{1, \|K\|_{\max}\}\right),$$
$$\|\Delta^{\beta_6}P_{\mathbf{j}}^{(k)}\|_1 \le (2\gamma_{2r+1}\log_2 N)^{\beta_6}(2r+1)^{L-k}, \quad \|(\Delta^{\beta_7}V_{\mathbf{j}}^{(L)})^{\mathsf{T}}\|_1 \le \gamma_{N_0}^{\beta_7}(2r+1),$$

where $\Delta H_{\mathbf{i}}$ is a submatrix of ΔH. A similar argument as in [26, (5.19)-(5.20)] yields

$$|\Delta\tilde{K}_{\text{low}}| \le \tfrac{54\cdot 2^{10}}{\pi}(2r+1)^{2(L-\tilde{L}+2)}\max\{1, \|K\|_{\max}\}\max\{\gamma_{2r+1}, \gamma_{N_0}\}\log_2 N + 3\gamma_w|K|,$$

Line 29 of Algorithm 4.1: Finally, the floating operation for line (29) in Algorithm 4.1 yields

$$\text{fl}(\phi) = (I + \Delta G)(\tilde{K}_{\text{high}} + \Delta\tilde{K}_{\text{high}} + \tilde{K}_{\text{low}} + \Delta\tilde{K}_{\text{low}})q = (K + \Delta K)q,$$

where $\tilde{K} = \tilde{K}_{\text{high}} + \tilde{K}_{\text{low}}$, $|\Delta G| \le \epsilon_{\text{mach}} I$, and

$$\Delta K := -E + \Delta\tilde{K}_{\text{high}} + \Delta\tilde{K}_{\text{low}} + \Delta G(K - E) + \Delta G\Delta\tilde{K}_{\text{high}} + \Delta G\Delta\tilde{K}_{\text{low}}.$$

It follows that

$$|\Delta K| \le |E| + (1+\epsilon_{\text{mach}})|\Delta\tilde{K}_{\text{high}}| + (1+\epsilon_{\text{mach}})|\Delta\tilde{K}_{\text{low}}| + |\Delta G||K| + |\Delta G||E|$$
$$\le \frac{54\cdot 2^{11}}{\pi}\max\{1, \|K\|_{\max}\}\max\{\gamma_{2r+1}, \gamma_{2\tilde{r}+1}, \gamma_{N_0}\}\log_2(N)\mathsf{p}(r, \tilde{r}) + 7\gamma_w|K| + 2\varepsilon,$$

where $\mathsf{p}(r, \tilde{r}) := (2\tilde{r}+1)^2 + (2r+1)^{2(L-\tilde{L}+2)}$. The proof is completed. ∎

This theorem indicates that the wideband FMM is backward stable. In the bound for $|\Delta K|$, the term that depends on N is $\log_2(N)\mathsf{p}(r, \tilde{r})$, where the main growth results from the factor $(2r+1)^{2(L-\tilde{L}+2)}$. Since typically $L = O(\log N)$, this factor is a polynomial term in N. Thus, in the worst case, the backward error has about polynomial growth. In comparison, it is known that some other rank-structured matrix-vector multiplications may be unstable [6, 9], even if the corresponding dense matrix-vector multiplication is stable. Theorem 5.2 confirms the stability of the our wideband FMM with balancing for the low-frequency regime.

The bound for $|\Delta K|$ in Theorem 5.2 may be conservative in terms of the ranks $\tilde{r}$ and r. To achieve a sharper dependence on the ranks, it is necessary to find sharper bounds for $\|B\|_{1,1}$ in Theorem 2.7 and $\|T_{\mathbf{x}',\mathbf{x}}\|_1$ in Theorem 3.1. However, finding these sharper bounds is challenging, as they involve more sophisticated properties/bounds of Bessel and Hankel functions. We would also like to mention that the backward error in Theorem 2.7 is not exactly in the form of a relative error. This is mainly due to the bound for $\|B\|_{1,1}$ in Theorem 2.7. Now, if $\|K\|_{\max} \geq 1$, then we can obtain a relative backward error bound as follows

$$\|\Delta K\|_{\max} \leq \left(\frac{6 \cdot 2^{11}}{\pi} \max\{\gamma_{2r+1}, \gamma_{2\tilde{r}+1}, \gamma_{N_0}\} \log_2(N)\mathsf{p}(r,\tilde{r}) + 9\max\{\gamma_w, \varepsilon\}\right) \|K\|_{\max}.$$

6. Numerical experiments. In this section, we show some numerical tests to evaluate the performance of the stable wideband FMM. As mentioned earlier, a source of instability/overflow stems from the low-frequency regime. Thus, the following numerical experiments mostly aim to evaluate the performance of our stable expansion of the 2D Helmholtz kernel for the low-frequency regime in subsection 2.2.

In the tests, we compare two versions of the FMM:

- Stable ($\tilde{L}$): our stable wideband FMM with balancing in the degenerate expansion (Theorem 2.7) and the translation relation (Theorem 3.1) for the low-frequency regime;
- Regular ($\tilde{L}$): a regular wideband FMM without any balancing for the low-frequency regime (and in particular, with the unscaled expansion in Proposition 2.5 and with unscaled translation relations).

For the high-frequency regime, both versions use the degenerate expansion in subsection 2.3 and the translation relation in subsection 3.2. Following the hierarchical partitioning of the domain, the high-frequency regime is involved when $\tilde{L} \geq 4$.

To carry out the experiments, we replicate the setting of [26, Section 6.1] and utilize its sets $\mathbf{X}$ and $\mathbf{Y}$. The points in the sets $\mathbf{X}$ and $\mathbf{Y}$ are randomly sampled from the standard normal distribution, shifted, and scaled so that $\Re(z), \Im(z) \in [0, 400]$ for each $z \in \mathbf{X} \cup \mathbf{Y}$. See [26, Figure 6.1] for a visualization of the scaled sets. In all of the following examples, we further scale both sets to observe how the magnitudes of the B generators behave and how the matrix-vector multiplication results are affected.

Since the wideband FMM matrix has all the U, V generators and translation matrices T with max-norms bounded by 1, we mainly inspect the impact of the norms of the B generators. Each table lists the expansion order r, the maximum value of $\|B\|_{\max}$ for all the B generators (denoted $\mathcal{B}$), and the accuracy of the FMM matrix-vector multiplication. The accuracy is measured by $\|\tilde{\phi} - \phi\|_2/\|\phi\|_2$, where ϕ is the result of the direct matrix-vector multiplication in (1.1), while $\tilde{\phi}$ is the result of the FMM matrix-vector multiplication via Algorithm 4.1, and q is a random vector generated from the standard normal distribution. The total number of levels is $L = 8$. Each partition at the leaf level contains at most $N_0 = 32$ points. We use the separation ratio $\tau = 0.6$ and $r = \tilde{r}$. All the computations are done in Matlab. We run several tests with different wavenumbers k and scaling of $\mathbf{X}$ and $\mathbf{Y}$.

First, Table 6.1 shows some test results with $\tilde{L} = 0$, i.e., using only expansions and translation relations based on the low-frequency regime. We observe that $\mathcal{B}$ from the regular FMM grows rapidly as we increase the number of expansion terms r.

This raises stability risks. Meanwhile, $\mathcal{B}$ from the stable FMM is bounded, which is consistent with our theoretical findings. When r increases beyond a certain number, it is not surprising that the regular FMM encounters overflow (so some of the entries under Regular appear as Inf and NaN in Matlab). Even though the regular FMM is quite accurate for certain r, the method is susceptible to stability issues, as pointed out in [8, 26]. For example, the accuracy may quickly deteriorate if we recompress or reorthonormalize some generators in the regular version.

TABLE 6.1
Comparison of the stable and the regular versions of the wideband FMM, where $\mathsf{k} = 1$ *and the sets* $10^{-10}\mathbf{X}$ *and* $10^{-10}\mathbf{Y}$ *are used (i.e., we multiply the sets* $\mathbf{X}$ *and* $\mathbf{Y}$ *by* 10^{-10}*).*

	Stable ($\tilde{L} = 0$)		Regular ($\tilde{L} = 0$)	
r	$\mathcal{B}$	$\frac{\Vert\tilde{\phi}-\phi\Vert_2}{\Vert\phi\Vert_2}$	$\mathcal{B}$	$\frac{\Vert\tilde{\phi}-\phi\Vert_2}{\Vert\phi\Vert_2}$
10	1.3376E+1	3.9548E-07	3.6792E+98	3.9548E-07
15	1.3376E+1	1.0273E-08	4.9886E+150	1.0273E-08
20	1.3376E+1	3.3858E-10	3.9286E+203	3.3858E-10
25	1.3376E+1	1.4195E-11	1.1309E+257	1.4195E-11
30	1.3376E+1	1.0863E-12	Inf	NaN
35	1.3376E+1	8.3062E-14	Inf	NaN
40	1.3376E+1	1.0250E-13	Inf	NaN
45	1.3376E+1	1.0399E-13	Inf	NaN
50	1.3376E+1	1.0404E-13	Inf	NaN

Next, in Table 6.2, we compare the two versions of the wideband FMM with various switching levels. One direct observation is that $\mathcal{B}$ from the stable wideband FMM is typically several magnitudes lower than that from the regular wideband FMM. Additionally, the regular FMM that relies solely on the high-frequency expansion gives highly inaccurate results. For the stable wideband FMM, the magnitudes of $\mathcal{B}$ are much smaller. If we take into account of the condition of Lemma 2.8 in setting the switching level $\tilde{L}$, then the choice of $\tilde{r}(= r)$ in the tests cannot be as high as we wish. For example, with the consideration of all pairwise distance of centers multiplied by the wavenumber at level 4, Lemma 2.8 dictates that we need to choose $r = 10$ to ensure the entrywise magnitudes of the B generators are bounded by $\sqrt{4/\pi}$. This may limit the accuracy. On the other hand, if $\mathcal{B}$ is allowed to be larger so that larger r is used, then higher accuracy can be achieved.

We want to point out that the condition in Lemma 2.8 is very conservative, and leads us to use stable low-frequency expansions at more levels if we want to both achieve high accuracy and have bounded $\mathcal{B}$. The conservative condition in Lemma 2.8 primarily comes from the challenge in quantifying the behavior of the Hankel function when its degree is larger than the argument, since it can grow very fast. The authors of [14] also made a similar remark on the growth of the Hankel function.

When a higher frequency is used, similar results can be observed. See Table 6.3.

7. Conclusions. In this work, we gave a stable matrix version of the wideband FMM for the 2D Helmholtz kernel. We used some simple properties of Bessel functions to derive a degenerate expansion of the 2D Helmholtz kernel for the low-frequency regime and showed a balancing strategy to mitigate the stability risk. We proved the norm bounds of the low-rank factors. A stable translation relation with balancing was derived for the low-frequency regime, along with the norm analysis for the translation operator. An efficient implementation of such a strategy was also discussed.

TABLE 6.2
Comparison of the stable and the regular versions of the wideband FMM with different $\tilde{L}$, where $\mathsf{k} = 10$ *and the sets* $10^{-2}\mathbf{X}$ *and* $10^{-2}\mathbf{Y}$ *are used (i.e., we multiply the sets* $\mathbf{X}$ *and* $\mathbf{Y}$ *by* 10^{-2}*).*

	Stable ($\tilde{L} = 0$)		Regular ($\tilde{L} = 0$)		Regular ($\tilde{L} = L$)	
r	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$
10	8.10E-1	3.94E-3	3.76E+08	3.94E-3	3.58E+7	4.88E-3
15	8.10E-1	2.22E-4	5.06E+15	2.22E-4	3.2E+14	8.37E-1
20	8.10E-1	2.22E-6	3.97E+23	2.22E-6	1.94E+22	4.16E+7
25	8.10E-1	4.85E-9	1.14E+32	4.85E-9	4.47E+30	9.47E+15
30	8.10E-1	1.80E-11	9.16E+40	1.80E-11	3.00E+39	1.82E+25
35	8.10E-1	4.99E-13	1.72E+50	4.99E-13	4.86E+48	1.75E+34
40	8.10E-1	2.15E-14	6.72E+59	2.15E-14	1.66E+58	3.25E+44
45	8.10E-1	5.32E-15	4.94E+69	5.32E-15	1.09E+68	9.48E+53
50	8.10E-1	5.28E-15	6.37E+79	5.28E-15	1.26E+78	1.02E+64

r	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$
	Stable ($\tilde{L} = 4$)		Regular ($\tilde{L} = 4$)	
10	8.10E-1	4.89E-3	3.76E+8	4.89E-3
15	8.10E-1	1.93E-4	5.06E+15	1.93E-4
20	1.98	2.21E-6	3.97E+23	2.21E-6
25	4.16E+2	5.89E-9	1.14E+32	5.89E-9
	Stable ($\tilde{L} = 5$)		Regular ($\tilde{L} = 5$)	
10	8.10E-1	4.88E-3	3.76E+8	4.88E-3
15	5.93E+1	1.93E-4	5.06E+15	1.93E-4
	Stable ($\tilde{L} = 6$)		Regular ($\tilde{L} = 6$)	
10	2.75E+1	4.88E-3	3.76E+8	4.88E-3

The stable wideband FMM is developed based on a combination of these techniques for the low-frequency regime and some expansion and translation strategies for the high-frequency regime. The matrix form is presented in terms of some FMM generators, including those reflecting basis contributions and translations between basis contributions. This simple interpretation makes it convenient to understand the wideband FMM. We further showed that this wideband FMM and is backward stable in comprehensive stability analysis.

As one possible future direction, we may consider the possibility of obtaining sharper bounds for the matrix norms discussed in section 5 and for the entries in Lemma 2.8. This may involve more sophisticated properties of Bessel and Hankel functions. The stabilization idea in this work may also potentially be extended to the FMM for the 3D Helmholtz kernel.

REFERENCES

[1] M. Abramowitz, I. A. Stegun, and R. H. Romer, Handbook of mathematical functions with formulas, graphs, and mathematical tables, Dover, 1974.

[2] S. Amini and A. T. J. Profit, *Analysis of a diagonal form of the fast multipole algorithm for scattering theory*, BIT Numer. Math., 39 (1999), no. 4, pp. 585–602.

[3] S. Amini and A. T. J. Profit, *Analysis of the truncation errors in the fast multipole method for scattering problems*, J. Comput. Appl. Math., 115 (2000), pp. 23–33.

[4] S. Amini and A. T. J. Profit, *Multi-level fast multipole solution of the scattering problem*, Eng. Anal. Bound. Elem., 27 (2003), pp. 547–564.

TABLE 6.3
Comparison of the stable and the regular versions of the wideband FMM with different $\tilde{L}$, where $\mathsf{k} = 100$ *and the sets* $10^{-2}\mathbf{X}$ *and* $10^{-2}\mathbf{Y}$ *are used (i.e., we multiply* $\mathbf{X}$ *and* $\mathbf{Y}$ *by* 10^{-2}*).*

	Stable ($\tilde{L}=0$)		Regular ($\tilde{L}=0$)		Regular ($\tilde{L}=L$)	
r	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$
100	2.70E-1	2.05E-3	3.91E+90	2.05E-3	3.89E+88	2.73E+75
120	2.70E-1	6.43E-4	2.29E+118	6.43E-4	1.90E+116	1.43E+103
140	2.70E-1	3.94E-5	3.91E+147	3.94E-5	2.78E+145	2.63E+132
160	2.70E-1	4.34E-9	1.19E+178	4.34E-9	7.44E+175	7.20E+162
180	2.70E-1	1.52E-14	4.52e+209	1.52E-14	2.51E+207	6.44E+194

	Stable ($\tilde{L}=4$)		Regular ($\tilde{L}=4$)	
r	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$	$\mathcal{B}$	$\frac{\|\tilde{\phi}-\phi\|_2}{\|\phi\|_2}$
100	2.70E-1	2.19E-3	3.91E+90	2.05E-3
120	2.70E-1	7.18E-4	2.29E+118	7.18E-4
140	2.70E-1	4.44E-5	3.91E+147	4.44E-5
160	4.34E+3	5.35E-9	1.19E+178	5.35E-9

[5] L. Banjai and W. Hackbusch, *Hierarchical matrix techniques for low- and high- frequency Helmholtz problems*, IMA J. Numer. Anal., 28 (2008), pp. 46–79.

[6] T. Bella, V. Olshevsky, and M. Stewart, *Nested product decomposition of quasiseparable matrices*, SIAM. J. Matrix Anal. Appl., 34 (2013), pp. 1520–1555.

[7] S. Börm, L. Grasedyck, and W. Hackbusch, *Introduction to hierarchical matrices with applications*, Eng. Anal. Bound. Elem., 27 (2003), pp. 405–422.

[8] D. Cai and J. Xia, *A stable matrix version of the fast multipole method: stabilization strategies and examples*, Electron. Trans. Numer. Anal., 54 (2021), pp. 581–609.

[9] C. Cao and J. Xia, *Kernel matrix approximations by sums of exponentials and stability of fast structured transforms*, preprint (2024).

[10] S. Chandrasekaran, P. Dewilde, M. Gu, and T. Pals, *A fast ULV decomposition solver for hierarchically semiseparable representations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 603–622.

[11] H. Cheng, W. Crutchfield, Z. Gimbutas, L. Greengard, J. Ethridge, J. Huang, V. Rokhlin, N. Yarvin, J. Zhao, *A wideband fast multipole method for the Helmholtz equation in three dimensions*, J. Comput. Phys., 216 (2006), pp. 300-325.

[12] H. Cheng, W. Crutchfield, Z. Gimbutas, L. Greengard, J. Huang, V. Rokhlin, N. Yarvin, J. Zhao, Remarks on the implementation of the wideband FMM for the Helmholtz equation in two dimensions. *Inverse problems, multi-scale analysis and effective medium theory*, 99-110, Contemp. Math., 408, *Amer. Math. Soc., Providence, RI*, 2006.

[13] W. C. Chew, J. M. Jin, E. Michielssen, and J. Song, Fast and efficient algorithms in computational electromagnetics, Artec House, Boston, 2001.

[14] M. H. Cho and W. Cai, *A wideband fast multipole method for the two-dimensional complex Helmholtz equation.*, Comput. Phys. Commun., 181 (2010), pp. 2086–2090.

[15] M. H. Cho, J. Huang, D. Chen, and W. Cai, *A heterogeneous FMM for layered media Helmholtz equation I: Two layers in* $\mathbb{R}^2$, J. Comput. Phys., 369 (2018), pp. 237–251.

[16] E. Darve and P. Havé, *Efficient fast multipole method for low-frequency scattering*, J. Comput. Phys., 197 (2004), pp. 341-363.

[17] B. Engquist and L. Ying, *A fast directional algorithm for high-frequency acoustic scattering in two dimensions*, Commun. Math. Sci., 7 (2009), no. 2, pp. 327–345.

[18] B. Engquist and H. Zhao, *Approximate separability of the Green's function of the Helmholtz equation in the high-frequency limit*, Commun. Pure Appl. Math., 71 (2018), no. 11, pp. 2220–2274.

[19] W. Gautschi, *Computational aspects of three-term recurrence relations*, SIAM Rev., 9 (1967), no. 1, pp. 24–82.

[20] L. Grasedyck, R. Kriemann, and S. Le Borne, *Domain decomposition based $\mathcal{H}$-LU preconditioning*, Numer. Math., 112 (2009), pp. 565–600.

[21] L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys., 73

(1987), no. 607, pp. 325–348.
[22] L. Greengard and V. Rokhlin, *On the efficient implementation of the fast multipole algorithm*, Technical Report RR-602, Department of Computer Science, Yale University, New Haven, (1988).
[23] W. Meng, *Explicit error bound of the fast multipole method for scattering problems in 2-D*, Calcolo, 60 (2023), no. 14, pp. 1–23.
[24] W. Meng and L. Wang, *Bounds for truncation errors of Graf's and Neumann's addition theorems*, Numer. Algorithms, 72 (2016), pp. 91-106.
[25] S. Ohnuki and W. C. Chew, *Truncation error analysis of multipole expansion*, SIAM J. Sci. Comput., 25 (2003), no. 4, pp. 1293–1306.
[26] X. Ou, M. Michelle, and J. Xia, *A stable matrix version of the 2D fast multipole method*, SIAM J. Matrix Anal. Appl., to appear, (2024). https://www.math.purdue.edu/~xiaj/work/fmm2d.pdf
[27] V. Rokhlin, *Rapid solution of integral equations of classical potential theory*, J. Comput. Phys., 60 (1985), no. 2, pp. 187–207.
[28] V. Rokhlin, *Rapid solution of integral equations of scattering theory in two dimensions*, J. Comput. Phys., 86 (1990), pp. 414–439.
[29] X. Sun and N. P. Pitsianis, *A matrix version of the fast multipole method*, SIAM Rev., 43 (2001), pp. 289–300.
[30] B. Wang, W. Zhang, and W. Cai, *Fast multipole method for 3-D Helmholtz equations in layered media*, SIAM J. Sci. Comput., 41 (2019), no. 6, pp. A3954–3981.
[31] B. Wang, D. Chen, B. Zhang, W. Zhang, M. H. Cho, and W. Cai, *Taylor expansion based fast multipole method for 3-D Helmholtz equations in layered media*, J. Comput. Phys., 401 (2020), 109008.
[32] G. N. Watson, A treatise on the theory of Bessel functions, Cambridge university press, 1995.
[33] Y. Xi and J. Xia, *On the stability of some hierarchical rank structured matrix algorithms*, SIAM J. Matrix Anal. Appl., 37 (2016), pp. 1279–1303.
[34] Y. Xi, J. Xia, S. Cauley, and V. Balakrishnan, *Superfast and stable structured solvers for Toeplitz least squares via randomized sampling*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 44–72.
[35] J. Xia, S. Chandrasekaran, M. Gu, and X. S. Li, *Fast algorithms for hierarchically semiseparable matrices*, Numer. Linear Algebra Appl., 17 (2010), pp. 953–976.
[36] W. Zhang, B. Wang, and W. Cai, *Exponential convergence for multipole and local expansions and their translations for sources in layered media: two-dimensional acoustic wave*, SIAM J. Numer. Anal., 58 (2020), no. 3, pp. 1440–1468.

A. Appendix: proofs of Lemmas 2.1 to 2.4.

Proof of Lemma 2.1. We first prove item (i). Suppose $z \geq \frac{1}{2}$ and $0 \leq p \leq z$. Since $|H_p(z)| < |H_z(z)|$ for a fixed $z > 0$ by (2.4), it suffices to show $|H_z(z)| < \sqrt{\frac{4}{\pi}}$ for $z \geq \frac{1}{2}$. By Nicholson's formula [32, Section 13.73 (pg 444), Section 6.15 (pg 172)],

$$|H_p(z)|^2 = \frac{8}{\pi^2}\int_0^\infty \int_0^\infty e^{-2z\sinh(t)\cosh(v)}\cosh(2pt)\mathrm{d}t\mathrm{d}v.$$

Let $G(z)$ be the first derivative of $|H_z(z)|^2$ with respect to z. Then,

$$G(z) = \frac{16}{\pi^2}\int_0^\infty \int_0^\infty e^{-2z\sinh(t)\cosh(v)}(t\sinh(2zt) - \sinh(t)\cosh(2zt)\cosh(v))\mathrm{d}t\mathrm{d}v.$$

Since for $z, v, t > 0$, $t < \sinh(t)$, $\sinh(2zt) < \cosh(2zt)$, $1 < \cosh(v)$, and $t\sinh(2zt) < \sinh(t)\cosh(2zt)\cosh(v)$, we have $G(z) < 0$. Therefore, for $z \geq \frac{1}{2}$, $|H_z(z)|^2 < \left|H_{1/2}(1/2)\right|^2 = 4/\pi$.

We then prove item (ii) of Lemma 2.1. Since $\sinh(t)$ is convex for all $t \in [0, \infty)$ and $\sinh(0) = 0$, we have $\lambda\sinh(t) \geq \sinh(\lambda t)$ for all $\lambda \in (0, 1)$ and all $t \in [0, \infty)$. Therefore,

$$|H_p(\lambda z)|^2 = \frac{8}{\pi^2}\int_0^\infty \int_0^\infty e^{-2\lambda z\sinh(t)\cosh(v)}\cosh(2pt)\mathrm{d}t\mathrm{d}v$$

$$\leq \frac{8}{\pi^2}\int_0^\infty\int_0^\infty e^{-2z\sinh(\lambda t)\cosh(v)}\cosh(2pt)\mathrm{d}t\mathrm{d}v$$
$$= \frac{8}{\pi^2}\int_0^\infty\int_0^\infty \frac{1}{\lambda}e^{-2z\sinh(t)\cosh(v)}\cosh\left(2\tfrac{p}{\lambda}t\right)\mathrm{d}t\mathrm{d}v = \lambda^{-1}|H_{p/\lambda}(z)|^2.$$

The proof is completed. ∎

Proof of Lemma 2.2. We first prove item (i). According to [3, Lemma 4], we have $\frac{Y_p(z)}{Y_{p+1}(z)} < \frac{z}{p}$ for $p \geq 2$ and $0 < z \leq p$. Thus, by (2.6), we have

$$\frac{C_p(z)}{C_{p+1}(z)} = \frac{Y_p(z)\sqrt{\frac{\pi p}{2}}\left(\frac{ez}{2p}\right)^p}{Y_{p+1}(z)\sqrt{\frac{\pi(p+1)}{2}}\left(\frac{ez}{2(p+1)}\right)^{p+1}} < \frac{2}{e}\left(1+\frac{1}{p}\right)^{p+\frac{1}{2}} \leq \frac{4\sqrt{2}}{e}.$$

Next, we prove items (ii) and (iii). Since $C_p'(z) = -\sqrt{\frac{\pi p}{2}}\left(\frac{ez}{2p}\right)^p\left(Y_p'(z)+\frac{p}{z}Y_p(z)\right)$, and the following recurrence relations hold for Bessel functions (see [32, 1]):

$$Y_{p-1}(z)+Y_{p+1}(z) = \tfrac{2p}{z}Y_p(z) \quad \text{and} \quad Y_{p-1}(z)-Y_{p+1}(z) = 2Y_p'(z),$$

we have $C_p'(z) = -\sqrt{\frac{\pi p}{2}}\left(\frac{ez}{2p}\right)^p Y_{p-1}(z)$. So, $C_p(z)$ is a strictly increasing function of z. According to [1, Section 9.5] or [3, Lemma 1], $Y_p(z) < 0$ for $0 < z \leq p$. It follows that $C_p(z) \leq C_p(p) = -Y_p(p)\sqrt{\frac{\pi p}{2}}\left(\frac{e}{2}\right)^p \leq |H_p(p)|\sqrt{\frac{\pi p}{2}}\left(\frac{e}{2}\right)^p$. This completes the proof. ∎

Proof of Lemma 2.3. Recall that $|J_k(\mathsf{k}z)|$ is an increasing function in z if $0 \leq \mathsf{k}z \leq |k|$ (see [1, Section 9.5], [32], or [3, Lemma 1]). Thus,

$$|E_r^{J_p}(\mathsf{k}z_1,\mathsf{k}z_2)| \leq \sum_{|l|\geq r \text{ or } |p+l|\geq r}|J_{p+l}(\mathsf{k}z_{\max})J_l(\mathsf{k}z_{\max})|$$
$$\leq \sum_{|p+l|\geq r}|J_{p+l}(\mathsf{k}z_{\max})| + \sum_{|l|\geq r}|J_l(\mathsf{k}z_{\max})| = 4\sum_{l=r}^\infty|J_l(\mathsf{k}z_{\max})|. \tag{A.1}$$

Since $|J_p(z)| \leq (z/2)^p/p!$ for all $p \in \mathbb{N}\cup\{0\}$ and $z \geq 0$ (see [32, 1]), we have

$$\sum_{l=r}^\infty|J_l(\mathsf{k}z_{\max})| \leq \sum_{l=r}^\infty\frac{1}{l!}\left(\frac{\mathsf{k}z_{\max}}{2}\right)^l \leq \sum_{l=r}^\infty\frac{1}{r!}\left(\frac{\mathsf{k}z_{\max}}{2}\right)^r\frac{1}{2^{l-r}} = \frac{2}{r!}\left(\frac{\mathsf{k}z_{\max}}{2}\right)^r$$

for all $r \geq \mathsf{k}z_{\max}$ and $(\mathsf{k}z_{\max})^{l-r} \leq l!/r!$. The proof is completed. ∎

Proof of Lemma 2.4. Note that $|J_p(z)| \leq |Y_p(z)|$ and $|H_p(z)| \leq \sqrt{2}|Y_p(z)|$ for all $0 < z \leq p$. Additionally, $|J_p(z)| \leq (z/2)^p/p! \leq (2\pi p)^{-1/2}(ez)^p(2p)^{-p}$ for all $p \in \mathbb{N}\cup\{0\}$ and $z \geq 0$ (see [32, 1]). Thus, we have

$$|E_r^{H_0}(\mathsf{k}w,\mathsf{k}t)| \leq 2\sum_{l=r}^\infty\left|H_l\left(\tfrac{\mathsf{k}t_{\max}}{\tau}\right)J_l(\mathsf{k}t_{\max})\right| \leq 2\sqrt{2}\sum_{l=r}^\infty\left|Y_l\left(\tfrac{\mathsf{k}t_{\max}}{\tau}\right)J_l(\mathsf{k}t_{\max})\right|$$
$$\leq 2\sqrt{2}\sum_{l=r}^\infty C_l\left(\tfrac{\mathsf{k}t_{\max}}{\tau}\right)\sqrt{\tfrac{2}{\pi l}}\left(\tfrac{2l\tau}{e\mathsf{k}t_{\max}}\right)^l\tfrac{1}{\sqrt{2\pi l}}\left(\tfrac{e\mathsf{k}t_{\max}}{2l}\right)^l$$
$$\leq \tfrac{2\sqrt{2}}{\pi r}\sum_{l=r}^\infty C_l\left(\tfrac{\mathsf{k}t_{\max}}{\tau}\right)\tau^l \leq \tfrac{2\sqrt{2}}{\pi r}\tfrac{\tau^r}{1-\tau}C_r\left(\tfrac{\mathsf{k}t_{\max}}{\tau}\right),$$

where we used (2.6) in the third line and the fact that $C_p(z) > C_{p+1}(z)$ for all $p \geq 2$ and $p \geq z > 0$ (see the remark before Lemma 2.2). ∎