Fast Multipole Methods for Wave and Charge Source Interactions in Layered Media and Deep Neural Network Algorithms for High-Dimensional PDEs

Wenzhong Zhang
wenzhongz@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_mathematics_etds

Part of the Data Science Commons, Numerical Analysis and Computation Commons, Numerical Analysis and Scientific Computing Commons, and the Special Functions Commons

Recommended Citation
https://scholar.smu.edu/hum_sci_mathematics_etds/15

This Dissertation is brought to you for free and open access by the Mathematics at SMU Scholar. It has been accepted for inclusion in Mathematics Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
FAST MULTIPOLE METHODS FOR WAVE AND CHARGE SOURCE
INTERACTIONS IN LAYERED MEDIA AND DEEP NEURAL
NETWORK ALGORITHMS FOR HIGH-DIMENSIONAL PDES

Approved by:

__________________________
Dr. Wei Cai
Professor of Mathematics

__________________________
Dr. Thomas Hagstrom
Professor of Mathematics

__________________________
Dr. Johannes Tausch
Professor of Mathematics

__________________________
Dr. Jianfeng Lu
Professor of Mathematics
FAST MULTIPOLE METHODS FOR WAVE AND CHARGE SOURCE INTERACTIONS IN LAYERED MEDIA AND DEEP NEURAL NETWORK ALGORITHMS FOR HIGH-DIMENSIONAL PDES

A Dissertation Presented to the Graduate Faculty of the Dedman College: School of Humanities and Sciences

Southern Methodist University

in

Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy

with a

Major in Computational and Applied Mathematics

by

Wenzhong Zhang

B.S., Peking University
M.S., Southern Methodist University

August 4, 2021
ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Professor Wei Cai, for his extensive and professional advice, exceptional passion, considerate encouragement and persistent support during my study in Southern Methodist University as well as in UNC Charlotte. He is really experienced in dividing tough problems into achievable steps, then making practical plans to reach separated goals, with clear instructions for teamwork while still allowing lots of freedom of ideas and thoughts to exchange at any time, which are probably the dominating reasons for the success towards the completion of my thesis.

I would also like to wholeheartedly thank Professor Bo Wang for his enormous and generous help, not only for our collaboration on 6 of the 7 papers included in this thesis, but also for the friendship we nurtured during the three years working together with countless discussions. I’m constantly inspired by his work rate like a marching bulldozer. Without his dedication, work on the FMM in layered media will not reach the current state in this thesis.

I would also like to specially thank Professor Sihong Shao at Peking University for his patience and care during the gap year before I met my advisor, helping me be mentally prepared for independently living as a graduate student.

I would like to thank my thesis committee, Professor Thomas Hagstrom, Professor Jianfeng Lu and Professor Johannes Tausch, for taking the time in a busy schedule for reviewing my work. I also had a great pleasure taking classes from Professor Hagstrom in topics closely related to my research.

Many thanks to the Mathematics Department and the communities of SMU and UNCC for providing a friendly environment in which I could dedicate myself to my research interests. Also thank my friends and colleagues who enriched my life during my Ph.D. studies.

Finally, I want to show my gratitude to my family, for their endless love and trust.
In this thesis, we develop fast algorithms for large scale numerical computations, including the fast multipole method (FMM) in layered media, and the forward-backward stochastic differential equation (FBSDE) based deep neural network (DNN) algorithms for high-dimensional parabolic partial differential equations (PDEs), addressing the issues of real-world challenging computational problems in various computation scenarios.

We develop the FMM in layered media, by first studying analytical and numerical properties of the Green’s functions in layered media for the 2-D and 3-D Helmholtz equation, the linearized Poisson–Boltzmann equation, the Laplace’s equation, and the tensor Green’s functions for the time-harmonic Maxwell’s equations and the elastic wave equation. Then, we propose the far-field expansions in layered media as the natural extension of the Graf’s addition theorems for the free-space problems. We define a modified far-field polarization distance in layered structure according to the exponential convergence behavior of these far-field expansions, and design the FMM framework for layered media based on the modified polarization distance. Numerical tests are conducted for a massive number of particles in layered media, verifying both the efficiency and the accuracy of the developed FMM algorithms.

Numerical algorithms using the deep neural network (DNN) for the purpose of finding a remedy to address the curse of dimensionality are proposed for the solution to high-
dimensional quasi-linear parabolic PDEs based on the Pardoux–Peng theory of the FBSDEs.

The algorithms are shown to generate DNN solutions for a 100-dimensional Black–Scholes–Barenblatt equation that are accurate in a finite region in the solution space, and have a convergence rate similar to that of the Euler–Maruyama discretization scheme used for the FBSDEs.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................. xii

LIST OF TABLES ................................................................. xv

CHAPTER

1. Introduction ............................................................... 1
   1.1. The fast multipole method in the free space............... 1
       1.1.1. The far-field expansions ............................... 1
       1.1.2. Hierarchical structure of the FMM ..................... 3
   1.2. Many-body problems in layered media ....................... 5
       1.2.1. Example in the ion transport model ................... 5
       1.2.2. Challenges of FMM in layered media ................. 7
   1.3. High-dimensional parabolic partial differential equations. 8
   1.4. Aim of thesis .................................................... 8
   1.5. Structure of the layered media ............................... 9

2. FMM for 2-D Helmholtz equation in layered media .............. 11
   2.1. Introduction .................................................... 11
   2.2. The Green’s function for Helmholtz equation in 2-D layered media .......... 12
       2.2.1. Reaction field decomposition of the LMGF in the frequency domain 13
       2.2.2. Sublinear bound of the reflection/transmission coefficients ........ 18
   2.3. Far-field expansions of the reaction field ................. 25
       2.3.1. An identity and some estimates on Bessel functions of the first kind 25
       2.3.2. The multipole expansion of the free-space Green’s function revisited 26
       2.3.3. The far-field expansions and translations ................ 28
           2.3.3.1. Numerical validation of exponential convergence ...... 34
# 2. The FMM framework

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4. The FMM framework</td>
<td>35</td>
</tr>
</tbody>
</table>

# 2.5. Proof of exponential convergence of far-field expansions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1. Convergence of pointwise Bessel-type expansions</td>
<td>37</td>
</tr>
<tr>
<td>2.5.2. Special cases of the integral Bessel-type expansion</td>
<td>38</td>
</tr>
<tr>
<td>2.5.3. Convergence of general integral Bessel-type expansions</td>
<td>44</td>
</tr>
<tr>
<td>2.5.3.1. The Cagniard–de Hoop transform</td>
<td>44</td>
</tr>
<tr>
<td>2.5.3.2. The general Bessel-type expansion</td>
<td>47</td>
</tr>
<tr>
<td>2.5.4. Proof of Theorem 2.12</td>
<td>55</td>
</tr>
</tbody>
</table>

# 2.6. Conclusion

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
</tr>
</tbody>
</table>

# 3. FMM for Helmholtz and linearized PB equation in 3-D layered media

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. Introduction</td>
<td>58</td>
</tr>
<tr>
<td>3.2. The Green’s function in 3-D layered media</td>
<td>59</td>
</tr>
<tr>
<td>3.2.1. Reaction field decomposition of the LMGF in the frequency domain</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2. Recursive formula for the reflection/transmission coefficients</td>
<td>63</td>
</tr>
<tr>
<td>3.3. Far-field expansions of the reaction field</td>
<td>66</td>
</tr>
<tr>
<td>3.3.1. The generalized Funk–Hecke formula</td>
<td>68</td>
</tr>
<tr>
<td>3.3.2. The far-field expansions and translations</td>
<td>73</td>
</tr>
<tr>
<td>3.3.2.1. Numerical validation of exponential convergence</td>
<td>76</td>
</tr>
<tr>
<td>3.4. The FMM framework</td>
<td>78</td>
</tr>
<tr>
<td>3.5. Implementation details for efficiency</td>
<td>79</td>
</tr>
<tr>
<td>3.5.1. Conversion to single integrals</td>
<td>79</td>
</tr>
<tr>
<td>3.5.2. Real arguments for the linearized Poisson–Boltzmann equation</td>
<td>82</td>
</tr>
<tr>
<td>3.5.3. Contraction of associated Legendre functions</td>
<td>83</td>
</tr>
<tr>
<td>3.5.4. Scaling factors</td>
<td>85</td>
</tr>
<tr>
<td>3.6. Numerical results</td>
<td>85</td>
</tr>
<tr>
<td>3.6.1. Test environment</td>
<td>86</td>
</tr>
</tbody>
</table>
3.6.2. Numerical results for FMM for Helmholtz equation in 3-D layered media ........................................ 87
3.6.3. Numerical results for FMM for linearized PB equation in 3-D layered media ................................. 87
3.6.4. Summary ................................................................. 88
3.7. Conclusion ............................................................... 89
4. FMM for Laplace’s equation in 3-D layered media ............................................................... 92
4.1. Introduction ............................................................ 92
4.2. The Green’s function in layered media ................................................................. 93
  4.2.1. Reaction field decomposition of the LMGF in the frequency domain .................................. 94
  4.2.2. Recursive formula for the reflection/transmission coefficients .............. 96
4.3. Far-field expansions of the reaction field and the FMM framework .......... 99
  4.3.1. The Modified Funk–Hecke formula ........................................ 100
  4.3.2. The far-field expansions and translations ........................................ 102
4.4. The FMM framework .................................................. 109
4.5. Implementation details for efficiency ........................................ 109
  4.5.1. Recursion formula for numerical integration .............................................. 111
  4.5.2. Changing contour of integration in the complex plane ......................... 112
4.6. Numerical results ....................................................... 113
4.7. Proof of exponential convergence of far-field expansions ......................... 117
  4.7.1. The Cagniard–de Hoop transform .............................................. 118
  4.7.2. Exponential convergence of single expansions and double expansions 122
  4.7.3. Exponential convergence of far-field expansions .................................... 127
4.8. Conclusion ............................................................... 129
5. A matrix basis formulation for dyadic Green’s functions in layered media ...... 131
  5.1. Introduction ............................................................ 131
5.2. A matrix basis formulation .................................................. 133
  5.2.1. Function fields for the LMDGs ........................................ 133
  5.2.2. The matrix basis ....................................................... 135
  5.2.3. The solution filtering theorem ...................................... 140
5.3. Application to the Maxwell’s equations in layered media ............ 143
  5.3.1. The dyadic Green’s functions of the time harmonic Maxwell’s
equations in the free space............................................... 144
  5.3.2. The dyadic Green’s functions of the time harmonic Maxwell’s
equations in layered media............................................. 145
    5.3.2.1. The equations in the spatial domain ......................... 146
    5.3.2.2. The equations in the frequency domain ..................... 147
    5.3.2.3. Further simplification of the formulation .................. 153
    5.3.2.4. The transverse potential and the Sommerfeld potential.... 157
5.4. Numerical validation of the Maxwell’s LMDG in a 10-layer medium ... 159
5.5. Application to the elastic wave equation in layered media .......... 164
  5.5.1. The dyadic Green’s function of the elastic wave equation in the
free space................................................................. 164
  5.5.2. The elastic wave equation in layered media ...................... 167
    5.5.2.1. The interface conditions of the elastic wave equation ...... 168
  5.5.3. The dyadic Green’s function in layered media with source in solid . 170
    5.5.3.1. Case when there’s a vacuum half space .................... 177
    5.5.3.2. Case when there’s a fluid layer ............................ 177
  5.5.4. The dyadic Green’s function in layered media with source in fluid . 179
    5.5.4.1. The vector basis ............................................... 179
    5.5.4.2. Derivation of the vector basis formulation ............... 181
5.6. A FMM framework using the matrix basis formulation .................. 185
5.7. Conclusion ........................................................................... 187
6. FBSDE based DNN algorithms for high-dimensional quasilinear parabolic PDEs 188

6.1. Introduction ............................................................................................................ 188

6.2. Pardoux–Peng theory on FBSDEs and quasilinear parabolic PDEs........ 190

6.3. FBSDE based neural network algorithms for quasilinear parabolic PDEs... 193

6.3.1. Existing FBSDE based neural network algorithms......................... 194

6.3.1.1. Deep BSDE.................................................................................................. 194

6.3.1.2. FBSNNs, Scheme 1 ................................................................. 195

6.3.2. FBSDE based deep neural network algorithms for PDEs ............. 197

6.3.2.1. FBSDE based algorithms - Scheme 2 ............................. 197

6.3.2.2. FBSDE based algorithms - Scheme 3 ......................... 199

6.4. Application in committor functions of the transition path theory .... 200

6.5. Numerical results .................................................................................................. 202

6.5.1. 100-dimensional Black–Scholes–Barenblatt equation ............... 202

6.5.1.1. Scheme 1 from FBSNNs.................................................. 204

6.5.1.2. Scheme 2 and Scheme 3 .............................................. 204

6.5.1.3. Extrapolation for higher order accuracy in the point solution 206

6.5.1.4. Region of validity of the DNN in a neighborhood of the initial value ............................................................................. 208

6.5.2. MultiscaleDNN for the BSB equation with temporal oscillations... 209

6.5.3. 100-dimensional committor function for a double-well potential 212

6.6. Conclusion ............................................................................................................. 214

7. Conclusion and future work.......................................................... 216

7.1. Overall contributions....................................................................................... 216

7.2. Future work .......................................................................................................... 219

BIBLIOGRAPHY......................................................................................................... 220
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Well separation of targets and sources in the multipole expansion (ME).</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>A 3-level quadtree as a demo for the FMM hierarchical structure in 2-D. On the same level, the interaction list of the dark green box consists of the light green boxes.</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Schematic illustration of a hybrid ion channel model [8].</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Sketch of the layer structure for general layered media.</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>The perturbed branch cut in the frequency domain for representing the LMGF of 2-D Helmholtz equation starting from $\pm \tilde{k}<em>\ell$ where $\tilde{k}</em>\ell = k_\ell + \epsilon_\ell \Lambda$, $\ell = 0, \cdots, L$.</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>The far-field distance $\rho_c$ of the ME in various field propagation directions, determined using the polarization sources.</td>
<td>32</td>
</tr>
<tr>
<td>2.3</td>
<td>Relative errors of ME and LE for $P = 3, 4, \cdots, 12$ for three cases, compared to the reference exponential convergence rates indicated by $r_{ME,(k)}, r_{LE,(k)}$, case $(k) = 1, 2, 3$.</td>
<td>36</td>
</tr>
<tr>
<td>2.4</td>
<td>The mapping $\phi : D^- \to D^+$. Regions $D_1$ and $\phi(D_1)$ are shadowed for illustration.</td>
<td>45</td>
</tr>
<tr>
<td>3.1</td>
<td>Relative errors of ME for the reaction field component $u_1^{\dagger\dagger}$ from the LMGF of the Helmholtz equation and the linearized PB equation for $P = 3, 4, \cdots, 12$ in a 3-layer medium for three cases, compared to the reference exponential convergence rates indicated by the polarization distance.</td>
<td>78</td>
</tr>
<tr>
<td>3.2</td>
<td>Polarization sources and boxes in the source tree of FMM.</td>
<td>81</td>
</tr>
<tr>
<td>3.3</td>
<td>Distribution of particles in a 3-layer medium.</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Performance of FMM for Helmholtz equation in 3-D layered media, picture from [76]. The notations $\Phi$ in [76] refers to the field $u$ in this thesis.</td>
<td>88</td>
</tr>
</tbody>
</table>
3.5 Performance of FMM for linearized PB equation in 3-D layered media, picture from [79]. The notations $\Phi$ in [79] refers to the field $u$ in this thesis. ................................................................. 90

4.1 Distribution of particles in a 3-layer medium. Distances between charges and the interfaces are exaggerated. .................................................. 114

4.2 Performance of FMM for Laplace’s equation in a 3-layer medium, picture from [78]. The notations $\Phi$ in [78] refers to the field $u$ in this thesis............. 115

4.3 Performance of FMM for Laplace’s equation in a multi-layer medium, picture from [78]. ............................................................. 115

4.4 The Cagniard–de Hoop transform for $\xi$ from the real axis to $\Gamma_+ \cup \Gamma_-$. ....... 120

5.1 Branch cuts defined by (5.4). .................................................. 134

5.2 Electric fields in a 10-layer medium described in Section 5.4. Fields are computed along $x = 0.2$ for $-5.0 \leq y \leq 5.0$ and $-14.5 \leq z \leq 0.5$. The contouring level is clamped to the range $(-0.05, 0.05)$ for a clearer illustration of wave pattern and avoiding the peak values near the source, which is located in the fourth layer at $r' = (0.0, 0.0, -4.23)$ with an orientation along the direction $\hat{\alpha}' = (1/2, 1/2, 1/\sqrt{2})$. ......................... 162

5.3 Magnetic fields in the 10-layer medium as described in Figure 5.2. ................. 163

6.1 (Non-convergence) relative error of Scheme 1 for $N = 12$ (middle), 48 (bottom) and 192 (top).................................................. 204

6.2 Relative error of Scheme 2 for $N = 12, 48, 192$ and 768. ......................... 205

6.3 Relative error of Scheme 3 for $N = 12, 48, 192$ and 768. ......................... 205

6.4 Prediction of 8 test sample paths from training results of Scheme 2 and Scheme 3, $N = 192$.................................................. 206

6.5 Mean relative error of extrapolation $u_0^{192}(t, X_t)$ for $0 \leq t \leq T$ with Scheme 2. . 208

6.6 Training error verified with initial value $\tilde{x}_0$ from a neighborhood of $x_0$, using Scheme 2, $N = 192$.................................................. 209

6.7 Illustration of a MscaleDNN [52]. ........................................... 211

6.8 Relative training error for fully-connected DNN and MscaleDNN for the model problem with oscillation, using Scheme 2 and $N = 48$. ............ 212

6.9 Prediction of 8 sample paths for problem with oscillation (6.58), using the MscaleDNN with Scheme 2 and $N = 192$. ...................................... 213
6.10 Training result of Scheme 4 from Section 6.4 for the committor function verified on a list of points (6.69) for the model problem in Section 6.5.3. . . . 214


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Comparison of CPU time (in seconds) with multiple CPU cores for FMM for Helmholtz equation in 3-D layered media, the truncation index $P = 4$, data from [76].</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparison of CPU time (in seconds) with multiple CPU cores for FMM for linearized PB equation in 3-D layered media, the truncation index $P = 5$, data from [79].</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of CPU time (in seconds) with multiple CPU cores for FMM for Laplace’s equation in 3-D layered media, the truncation index $P = 5$, data from [78].</td>
</tr>
<tr>
<td>5.1</td>
<td>Maximum absolute error of the continuity of $E_x, E_y, \varepsilon E_z, H_x, H_y$ and $\mu H_z$ across interface planes $z = d_\ell$, $0 \leq \ell \leq 8$ for 10201 $(x, y)$ coordinate pairs in the range of $[-5.0, 5.0] \times [-5.0, 5.0]$. The dipole source locates in layer no. 3.</td>
</tr>
<tr>
<td>6.1</td>
<td>Relative error of $Y_0$ from the network approximation and extrapolation.</td>
</tr>
</tbody>
</table>
To my mother, Xinhui Huang, for her love, courage and determination.
1.1. The fast multipole method in the free space

The fast multipole method (FMM) in the homogeneous free space, proposed by L. Greengard and V. Rokhlin in 1987 [34], has been a revolutionary development in modern computational algorithms for calculating many-body interactions. It has been known as one of the top 10 algorithms in the 20th century [23]. The FMM is a hierarchical algorithm reducing the $O(N^2)$ complexity of computing long-range forces among $N$ particles or sources in the form

$$u(r_i) = \sum_{j=1}^{N} G(r_i, r_j') q_j, \quad i = 1, 2, \cdots, N$$

(1.1)

to $O(N)$ or $O(N \log N)$ both in time and in storage. Such a feature has brought great impact on modern computational biology, astronomy, fluid mechanics, quantum chemistry, and computational electromagnetics, etc. [82, 73, 10, 8, 5, 7].

1.1.1. The far-field expansions

The fundamental mathematical basis of the FMM is the far-field expansions for long-range interactions, including the multipole expansion (ME), the local expansion (LE), and the translation operators between ME and LE, including the multipole-to-local translation (M2L), multipole-to-multipole translation (M2M) and the local-to-local translation (L2L).

Among these far-field expansions, the multipole expansion is the best known, and has also been broadly accepted in other tree-based methods such as in [4, 6, 49]. The ME requires the well separation of target particles and source particles in a manner that sources are gathered
around a source center, while the targets are farther away. The details are described as follows. Given \( M \) target particles at positions \( \{r_i\}_{i=1}^M \) and \( N \) source particles at positions \( \{r'_j\}_{j=1}^N \) with strength \( q_j \) at each source \( r'_j \), the source center is a position \( r_c \) so that there exists a constant number \( c_0 > 1 \) satisfying

\[
\|r_i - r_c\| > c_0\|r'_j - r_c\| \tag{1.2}
\]

for any target \( r_i \) and any source \( r'_j \). When applied to the FMM, \( c_0 \) is usually set to be 2.

The ME of the interaction (1.1) is then proposed as a series expansion

\[
u(r_i) = \sum_{j=1}^N G(r_i, r'_j)q_j \approx \sum_{p=1}^P I_p(r_i, r_c)M_p(r_c), \quad 1 \leq i \leq M, \tag{1.3}
\]

where each \( I_p(r_i, r_c) \) in the series is called a multipole expansion basis function, solely determined by the information of the target \( r_i \) and the source center \( r_c \), while each \( M_p(r_c) \) is the corresponding multipole expansion coefficient which carries the information of the sources. Calculation of all the ME coefficients with \( N \) sources usually has an \( O(PN) \) complexity with a \( P \)-term series, and calculation of the final summation for all the targets has an \( O(MP) \) complexity. When the series expansion converges sufficiently fast, e.g. at an exponential rate, so that a small value of \( P \) is sufficient to reach good accuracy, the overall calculation
has linear complexity.

Similarly, the local expansion is proposed around a target center \( r^l_c \) as

\[
u(r_i) \approx \sum_{s=1}^{S} K_s(r_i, r^l_c)L_s(r^l_c), \quad 1 \leq i \leq M
\]  

(1.4)

provided a similar well-separation condition holds

\[
\|r^l_j - r^l_c\| > c_0 \|r_i - r^l_c\|, \quad 1 \leq i \leq M, \quad 1 \leq j \leq N,
\]  

(1.5)

where each \( K_s(r_i, r^l_c) \) is a LE basis function determined only by the information of the target \( r_i \) and the target center \( r^l_c \), while \( L_s(r^l_c) \) is the corresponding LE coefficient that contains the information of sources.

The multipole-to-local translation operator is interpreted as a series expansion on each LE coefficient

\[
L_s(r^l_c) \approx \sum_{p=1}^{P} A_{sp}(r^l_c, r_c)M_p(r_c), \quad 1 \leq s \leq S.
\]  

(1.6)

The M2L translation coefficients \( A_{sp}(r^l_c, r_c) \) are required to be determined by target centers and source centers, without the dependence on the specific sources.

The multipole-to-multipole and local-to-local translation operators are similarly designed as M2L to shift the centers of MEs and LEs for implementing tree-based algorithms.

1.1.2. Hierarchical structure of the FMM

In practice, targets and sources in the domain are not well separated, hence the far-field expansions can not be straightforwardly applied. To overcome such difficulty, the FMM uses a hierarchical structure so that far-field expansions are used only within the same level for well-separated boxes containing sources and targets, or across adjacent levels. A quadtree or an octree is used in 2-D or in 3-D problems, respectively, to divide the domain into a few levels, see Figure 1.2. The number of levels is chosen so that there are \( O(1) \) particles in each
Figure 1.2: A 3-level quadtree as a demo for the FMM hierarchical structure in 2-D. On the same level, the interaction list of the dark green box consists of the light green boxes.

finest box. For each box not from the top level, its “interaction list” is defined as the close neighboring boxes of its parent excluding the close neighborhood of itself, as shown by the light green boxes in Figure 1.2.

The interaction between each target-source pair is computed in the hierarchical structure by the following rules:

1. On the finest level, if the target and the source locate in the same box, or in the close neighboring boxes, the interaction is calculated by direct computation.

2. Otherwise, we repeatedly shift to parent box centers of the source and of the target, until on the level when the source box belongs to the interaction list of the target box, where the M2L translation is taken on this level, connecting the upward pass of ME/M2M from the source to the current source center, and the downward pass of L2L/LE from the current target center to the target.

A detailed algorithm framework with complexity analysis can be found in [35].

Due to the possible non-uniform distribution of particles, including shortcut operations such as the source-to-local translation (S2L) and the multipole-to-target translation (M2T) may be beneficial. Mathematically, the S2L explicitly computes the LE basis, and the M2T
computes the ME basis. Details of an adaptive implementation regarding these shortcuts with complexity analysis can be found in [54].

1.2. Many-body problems in layered media

While the classical FMM was developed for problems in the free space, a number of important applications also require calculation of pairwise interactions among particles in layered media, such as calculating the capacitance of interconnects (ICs) in very large-scale integrated (VLSI) circuits for microchip designs [59, 89, 84, 68, 66], complex scattering problem in meta-materials [12], electrical potential computation in ion channel simulation [51], etc.

For simulations in VLSI, due to the complex geometric structure of the ICs, the charge potential solution to the Laplace’s equation is usually solved by an integral method with the Green’s function of the layered media (cf. [59, 89]), which results in a huge dense linear algebraic system to be solved by an iterative method such as GMRES [9], etc. Another model of the ion transport will be discussed later. In such applications, because of the scale of the required number of particles for simulation, a numerical algorithm for calculating the pairwise interaction that is efficient in both space and time is demanded.

1.2.1. Example in the ion transport model

We present a computational electrostatic problem arising from the ion transport in ion-channel or nanopores as an example to the many-body problem in layered media [79]. A hybrid model within a three-layer medium of materials with different dielectric constants and inverse Debye–Hückle lengths, corresponding to ionic solvents above and below and a membrane in the middle, is shown in Figure 1.3. A cylinder $\Omega_D$ of height $D$ within the three-layer medium with axis perpendicular to the interfaces of the layered media (see Figure 1.3) is used to represent the ion channel. The finite height cylinder represents a dividing interface in a hybrid solvation model for biomolecule simulations (refer to Section 4.5 in [8] for
more details). The background environment outside the cylinder, composed of a membrane and solvents above and below, are modeled as layered continuum dielectrics described by dielectric constants $\varepsilon_\ell$ and inverse Debye–Hückle screen lengths $\lambda_\ell$ measuring the ionic length in the layer $\Omega_\ell$, respectively, $\ell = 0, 1, 2$.

Suppose there are $M$ charges inside the cylinder locating at $r_k$ with magnitudes $q_k$, $1 \leq k \leq M$, which are the partial charges of the ions and membrane protein molecules inside the explicit atomistic cylinder. In a Monte Carlo or molecular dynamics simulation of the hybrid system, the electrostatic potential $\phi(r)$, thus forces on each charge through the gradient of the potential, will be obtained by solving the following Poisson–Boltzmann (PB) equation

$$\nabla^2 \phi(r) - \lambda^2(r) \phi(r) = -\frac{1}{\varepsilon(r)} \sum_{k=1}^{M} q_k \delta(r - r_k),$$

(1.7)

where the functions $\varepsilon(r)$ and $\lambda(r)$ are the dielectric constants and the inverse Debye–Hückle lengths in each region of the domain, respectively. The solution to (1.7) can be obtained by solving a Lipmann–Schwinger type integral equation using the layered Green’s function $u(r; r')$

$$\phi(r) = -\frac{1}{\varepsilon(r)} \sum_{k=1}^{M} q_k u(r; r_k) + \int_{\Omega_D} \Delta(r') u(r; r') \phi(r') dr',$$

(1.8)

where

$$\Delta(r) = \lambda^2_{\Omega_D}(r) - \lambda^2_\ell, \quad r \in \Omega_\ell, \quad \ell = 0, 1, 2.$$
The volume integral equation (1.8) is then discretized by a Nyström collocation method with an appropriate quadrature formula with \( \{ r_i, \omega_i \}^N_{i=1} \) as the nodes and weights for the domain \( \Omega_D \). More precisely, the discretization at each point \( r_i \in \Omega_D \) is given by

\[
\phi(r_i) = -\frac{1}{\varepsilon(r_i)} \sum_{k=1}^{M} q_k u(r_i; r_k) + \sum_{j=1}^{N} u(r_i; r_j) \cdot \omega_j \Delta(r_j) \phi(r_j), \quad 1 \leq i \leq N, \tag{1.10}
\]

where the second summation is exactly a many-body problem, whose kernel function \( u(r_i; r_j) \) is the layered Green’s function of the PB equation in the 3-layer medium.

1.2.2. Challenges of FMM in layered media

So far, a few approaches have been proposed to handle the wave interaction of sources in layered media using the Green’s functions of the Helmholtz equation, such as image approximations of the layered Green’s function for a half space [21], Taylor expansion-based low-rank representation of Green’s function and FMM method [74, 72], the inhomogeneous plane wave method [41], windowed Green’s function method [18, 8], and cylindrical wave decomposition of the Green’s function in 3-D and 2-D FMMs [18]. In principle, the method of kernel independent compression techniques [83, 80] could also be considered for the layered Green’s functions.

However, these methods either didn’t make full use of the layer structure of the problem, or failed to describe proper far-field relationship for targets and sources in layered media, which is crucial for the acceleration of hierarchical algorithms. The line image approximation [21] might be impossible to be extended to the general multi-layered media due to the complexity of the Green’s functions in general. Similar situations happen to the dielectric media, where the Green’s functions of the Laplace’s equation and the linearized PB equation are studied.

From a mathematical point of view, the FMM in the free space is based on the Graf’s addition theorems for the Green’s functions in the free space, which performs a target-source
separation with an exponential convergence rate for long-range interactions. Unfortunately, for problems in layered media, the corresponding addition theorem hasn’t been found. For this reason, the analytical approach-based FMM work by L. Greengard and V. Rokhlin has not been extended to the layered media.

1.3. High-dimensional parabolic partial differential equations

High dimensional parabolic partial differential equations prevail in various practical applications, such as the Allen–Cahn equations for phase transitions in material sciences, the Schrödinger equations for simulation of the interaction between many particles in quantum mechanics, the Black–Scholes equation in optional pricing, and the Hamilton–Jacobi–Bellman equations in game theories.

To numerically solve PDEs in high dimensions, the major challenge of the traditional numerical methods, such as the finite element method, the finite difference method and spectral methods, is the curse of dimensionality, namely, the number of unknowns in the discretized systems for the PDEs grows exponentially in terms of the dimension of the problem. For problems with high-frequency oscillation, the issue in scale is even worse.

1.4. Aim of thesis

The work that constitutes this thesis is focused on developing fast algorithms addressing the concerns mentioned above, including the fast multipole method in layered media, and the forward-backward stochastic differential equation (FBSDE) based deep neural network (DNN) algorithms for high-dimensional PDEs.

The FMM in layered media is developed in Chapters 2–5, where we study some in-depth properties of the Green’s functions in layered media for the 2-D and 3-D Helmholtz equation, the 3-D linearized PB equation and Laplace’s equation, as well as the tensor Green’s functions for the time-harmonic Maxwell’s equations and the elastic wave equation, both analytically and numerically. With the knowledge of the Green’s functions in layered media, we propose
the far-field expansions in layered media as the natural extension of the Graf’s addition theorems for the free-space problems. The exponential convergence behavior of these far-field expansions suggests a modified definition of the far-field distance in layered structure, based on which the FMM framework for layered media is established. Numerical tests are conducted for a massive number of particles in layered media, verifying both the efficiency and the accuracy.

Numerical algorithms using the deep neural network (DNN) for the purpose of finding a remedy to address the curse of dimensionality are proposed in Chapter 6 for the solution to high-dimensional quasi-linear parabolic PDEs based on the Pardoux–Peng theory of the FBSDEs. The algorithms are shown to generate DNN solutions for a 100-dimensional Black–Scholes–Barenblatt equation that are accurate in a finite region in the solution space, and have a convergence rate similar to that of the Euler–Maruyama discretization scheme used for the FBSDEs.

1.5. Structure of the layered media

Many important physical problems naturally have a layered structure. In this thesis, we suppose in a layered medium, the space is separated into $L + 1$ homogeneous horizontal layers, indexed by $0, \cdots, L$ from top to bottom, respectively. The interface coordinates are at $d_0 > \cdots > d_{L-1}$, as shown in Figure 1.4. (In 2-D problems, we adopt the Cartesian coordinates $(x, y)$, and the interface lines are given with the $y$-coordinates accordingly.)

For the rest of this thesis, there are various piecewisely defined functions, variables and constants that require labeling for layers. When not further specified, we assume the source particle $r'$ locates in the $\ell'$-th layer, and the target particle $r$ locates in the $\ell$-th layer,

$$r = \begin{cases} (x, y) & \text{in 2-D}, \\ (x, y, z) & \text{in 3-D}, \end{cases} \quad r' = \begin{cases} (x', y') & \text{in 2-D}, \\ (x', y', z') & \text{in 3-D}. \end{cases} \quad (1.11)$$

A function, variable or constant that is piecewisely defined in each layer will be assigned the
subscript \((\cdot)_\ell\) to specify the \(\ell\)-th layer. When there’s no need to distinguish different layers in an equation, the subscript may be omitted. For example, suppose layer \(\ell\) has constant density \(\rho_\ell, \ell = 0, 1, \ldots, L\), then the density may be denoted by the letter \(\rho\) in an equation, which means

\[
\rho = \rho_\ell \quad \text{when } r \text{ locates in layer } \ell.
\]  

For functions, variables or constants that depend on both the target index \(\ell\) and the source index \(\ell'\), we use the double subscripts \((\cdot)_{\ell\ell'}\), while sometimes omitted or written with only \(\ell\), in case there’s no worry about confusion. For example, the Green’s function \(u(r, r')\) for the Helmholtz equation is piecewisely denoted both in terms of the target layer and in terms of the source layer, i.e.,

\[
u(r; r') = u_{\ell\ell'}(r; r').
\]
Chapter 2
FMM for 2-D Helmholtz equation in layered media

The content in this chapter is a revision of the following published journal paper in collaboration with Bo Wang and Wei Cai:


In this chapter, we illustrate the theory of FMM for 2-D Helmholtz equation in layered media with the convergence analysis of far-field expansions. The methodology will be applied to the second-order linear partial differential equations in 3-D in the following chapters.

We begin with general theories of the layered media Green’s functions (LMGF) of the 2-D Helmholtz equation in Section 2.2. Then, we propose the far-field expansions of the LMGF in Section 2.3. The exponential convergence properties of the far-field expansions naturally lead to the framework of the FMM in layered media to be elaborated in Section 2.4, while the proof of exponential convergence is given in Section 2.5.

2.1. Introduction

The multipole expansion (ME), local expansion (LE), and multipole-to-local translation (M2L) form the mathematical structure of the FMM for evaluating integral operators associated with the Green’s function of Helmholtz equations in wave scattering [65, 13]. The ME for the Green’s functions in the free space was based on the Graf’s addition theorems for Bessel functions. To extend the FMM for wave scattering in layered media, ME and M2L formulas for Helmholtz equations in a 2-D half-space domain were proposed in [20].
The derivation in [20] for the ME and M2L made use of the image (point and line images) representation of the Green’s function of the half-space domain with an impedance boundary and the MEs, based on the Graf’s addition theorem, for the image charges as well as the original source charges. And, it was shown that the ME coefficients used to compress the far field of the source charges in the free space can also be used to compress the far field of the images, therefore, producing a ME for the Green’s functions of the 2-D half-space domain. It was predicted in [20] that similar results could hold for general layered media by using a frequency domain Sommerfeld representation of the Green’s function [16] expressing the layered Green’s function with plane waves. Also, in the case of the half space with an impedance boundary condition, the image representation of the domain Green’s function justifies the truncation order, thus the exponential convergence, of ME and M2L. A heterogeneous FMM was proposed and implemented in [21], giving an $O(N)$ complexity of evaluating the integral operator of low frequency Helmholtz operators for sources in the half space.

As an image representation of general layered media Green’s function may not exist, in this chapter, an alternative complete derivation for the ME, LE, and M2L operators for the LMGF for 2-D Helmholtz equation by using the generating function of the Bessel functions of the first kind (referred as the Bessel generating function in this thesis) will be presented. Moreover, we will give a rigorous proof of the exponential convergence of the ME, LE, and M2L and L2L translation operators for acoustic wave sources in general 2-D layered media. The convergence analysis reveals a fact that the convergence of ME, LE, L2L, and M2L for the reaction field component of the Green’s function in fact depends on a polarized distance, which is measured between the target and an equivalent polarization source, thus suggesting how the FMM framework should be set for sources and targets in layered media.
2.2. The Green’s function for Helmholtz equation in 2-D layered media

The LMGF for 2-D Helmholtz equation satisfies the differential equation

\[ \nabla^2 u(r; r') + k^2 u(r; r') = -\delta(r - r'), \tag{2.1} \]

where the derivatives of \( \nabla^2 \) are taken on the target coordinates \( r = (x, y) \), \( k \in \mathbb{R}^+ \) usually refers to the wave number for wave propagation in the media, and \( \delta \) is the Dirac delta function on \( \mathbb{R}^2 \). In addition, the upward/downward outgoing radiation conditions [11] must be guaranteed. The LMGF should also satisfy the interface conditions given in the form

\[ [au] = 0, \quad \left[ b \frac{\partial u}{\partial n} \right] = 0, \quad y = d_0, \ldots, d_{L-1}, \tag{2.2} \]

where \( n = (0, 1) \) is the normal vector on the interface, \([.]\) refers to the jump of values in the brackets, \( a, b > 0 \) are piecewisely constant parameters in each layer.

2.2.1. Reaction field decomposition of the LMGF in the frequency domain

To evaluate the LMGF, we first notice that when \( \ell \neq \ell' \), i.e. the target particle \( r \) and the source particle \( r' \) do not locate in the same layer, we always have \(-\delta(r - r') = 0\), resulting in a homogeneous second-order differential equation. Also, when \( \ell = \ell' \), the free-space Green’s function is a solution to the differential equation (2.1), albeit not satisfying the interface conditions. Therefore, we can define the “reaction field” of the LMGF as follows, which always satisfy the homogeneous Helmholtz equation within each layer.

**Definition 2.1 (Reaction field).** Let \( G^f(r; r') \) be the free-space Green’s function determined by the 2-D Helmholtz equation (2.1) with Sommerfeld radiation condition (2.5). The reaction field of the LMGF is defined by

\[ u^r(r; r') = u(r; r') - \delta_{\ell\ell'} G^f(r; r'). \tag{2.3} \]
Note that the free-space Green’s function is given by

\[ G^f(r; r') = \frac{i}{4} H_0^{(1)}(k \rho |r - r'|), \quad (2.4) \]

where \( H_0^{(1)} \) is the Hankel function of the first kind of order zero. \( G^f \) satisfies the Sommerfeld radiation condition

\[ \sqrt{r} \left( \frac{\partial}{\partial r} - ik \right) G^f(r; r') \rightarrow 0 \quad \text{as} \quad r = |r| \rightarrow \infty. \quad (2.5) \]

To solve the “piecewisely homogeneous” differential equation for the reaction field, take the 1-D Fourier transform

\[ f(x) = \int_{-\infty}^{\infty} e^{i\lambda(x-x')} \hat{f}(\lambda) d\lambda. \quad (2.6) \]

In the frequency domain, the reaction field satisfies the piecewisely homogeneous equation

\[ (-\lambda^2 + \partial_{yy}) \hat{u}^r(\lambda, y; r') + k^2 \hat{u}^r(\lambda, y; r') = 0, \quad (2.7) \]

while the interface conditions (2.4) are equivalent to

\[ \begin{bmatrix} a_\ell \hat{u}^r \end{bmatrix} = - \begin{bmatrix} \delta_{\ell,\ell'} a_\ell \hat{G}^f \end{bmatrix}, \quad \begin{bmatrix} b_\ell \partial_y \hat{u}^r \end{bmatrix} = \begin{bmatrix} \delta_{\ell,\ell'} b_\ell \partial_y \hat{G}^f \end{bmatrix} \quad \text{at} \quad y = d_\ell, \quad 0 \leq \ell \leq L - 1, \quad (2.8) \]

where

\[ \hat{G}^f(\lambda, y; r') = \frac{e^{-\sqrt{\lambda^2 - k^2_\ell}|y-y'|}}{4\pi \sqrt{\lambda^2 - k^2_\ell}} \quad (2.9) \]

and we have used the fact that \( \partial_y G^f = -\partial_y G^f \).

The branch cut of the square root is specified as follows: for any \( z = re^{i\theta} \in \mathbb{C} \) with \( r \geq 0 \), \( \theta \in [-\pi, \pi) \), define

\[ \sqrt{z} = \sqrt{r} e^{i\frac{\theta}{2}}. \quad (2.10) \]

In addition, for each square root \( \sqrt{\lambda^2 - k^2_\ell} \), the corresponding branch cut in the \( \lambda \)-plane is
The union of the imaginary axis and the real interval \([-\kappa_\ell, \kappa_\ell]\). This should be understood in the sense that in a realistic physical case where the medium in layer \(\ell\) is lossy with a perturbed wave number
\[
\tilde{\kappa_\ell} = \kappa_\ell + \epsilon_\ell i, \quad 0 < \epsilon_\ell \ll 1.
\] (2.11)

The perturbed branch cut is then shown in Figure (2.1). The branch cut of \(\sqrt{\lambda^2 - \kappa_\ell^2}\) is the limit of the perturbed one as \(\epsilon_\ell \to 0^+\).

By the separation of \(y\) and \(y'\) variables, the reaction field is decomposed into up to 4 components regarding different field propagation directions
\[
\hat{u}^r(\lambda; y; r) = \sum_{\ast, \ast' \in \{\uparrow, \downarrow\}} \hat{u}^{r, \ast\ast'}_{\ell\ell'}(\lambda; x') e^{-\tau\ast \sqrt{\lambda^2 - \kappa_{\ell'}^2} y} e^{-\tau\ast' \sqrt{\lambda^2 - \kappa_{\ell'}^2} y'},
\] (2.12)

where
\[
\tau^{\uparrow} = +1, \quad \tau^{\downarrow} = -1
\] (2.13)

are corresponding to the upward and the downward directions of the field propagation, respectively. Due to the upward/downward outgoing radiation conditions [11], on the top layer, the incoming wave from \(y = +\infty\) is prohibited. So is the incoming wave from \(y = -\infty\) on the bottom layer. Thus,
\[
\hat{u}^{r, \downarrow\ast}_{0e'} = \hat{u}^{r, \uparrow\ast}_{L'e'} = 0, \quad \ast \in \{\uparrow, \downarrow\}.
\] (2.14)

For the source \(r' = (x', y')\), when \(r'\) locates in the top layer, the factor \(e^{\sqrt{\lambda^2 - \kappa_{\ell'}^2} y'}\) won’t appear
in the interface equations (2.8), so this term has no contribution to the decomposition (2.12).

The same applies to $e^{-\sqrt{\lambda^2-k'^2_{\ell}}y'}$ when $r'$ locates in the bottom layer. Thus,

$$\hat{u}_{t0}^{r,*\downarrow} = \hat{u}_{tL}^{r,*\uparrow} = 0, \quad * \in \{\uparrow, \downarrow\}. \quad (2.15)$$

To better study the mathematical properties of the reaction field, we introduce the local coordinates to polish the representation used in (2.12). Define

$$d_{\ell}^\uparrow = d_{\ell}, \quad 0 \leq \ell \leq L - 1, \quad (2.16)$$

$$d_{\ell}^\downarrow = d_{\ell-1}, \quad 1 \leq \ell \leq L, \quad (2.17)$$

which are the lower and the upper interface coordinates of the $\ell$-th layer, respectively, except the top-most and the bottom-most ones. For the sake of convenience, let

$$d_0^\downarrow = +\infty, \quad d_L^\uparrow = -\infty. \quad (2.18)$$

**Definition 2.2 (Reaction field decomposition).** The reaction field decomposition of $\hat{u}^r$ in the frequency domain is defined by

$$\hat{u}^r(\lambda, y; r') = \sum_{*,* \in \{\uparrow, \downarrow\}} \sigma_{\ell*}^{**}(\lambda; x') e^{-\tau* \sqrt{\lambda^2-k'^2_{\ell}(y-d_{\ell}^* \tau)}} e^{-\tau* \sqrt{\lambda^2-k'^2_{\ell}y'}(y'-d_{\ell}^* \tau)}, \quad (2.19)$$

where $\sigma_{\ell*}^{**}(\lambda; x')$ are the reflection/transmission coefficients of the layered media in the frequency domain. Corresponding to the expansion in the spatial domain, the reaction field decomposition can be written by

$$u^r(r; r') = \int_{-\infty}^{\infty} e^{i\lambda(x-x')} \sum_{*,* \in \{\uparrow, \downarrow\}} \sigma_{\ell*}^{**}(\lambda; x') e^{-\tau* \sqrt{\lambda^2-k'^2_{\ell}(y-d_{\ell}^* \tau)}} e^{-\tau* \sqrt{\lambda^2-k'^2_{\ell}y'}(y'-d_{\ell}^* \tau)} d\lambda$$

$$= \sum_{*,* \in \{\uparrow, \downarrow\}} u_{\ell*}^{**}(r; r'), \quad (2.20)$$
where each reaction field component

\[ u_{\ell'\ell}^*(r; r') = \int_{-\infty}^{\infty} E_{\ell'\ell}^*(r, r', \lambda) \sigma_{\ell'\ell}^*(\lambda; x') d\lambda, \]  

(2.21)

where

\[ E_{\ell'\ell}^*(r, r', \lambda) = e^{i\lambda(x-x')-r'\sqrt{\lambda^2-k_t^2(y-d_{\ell'})^2}}. \]  

(2.22)

The reflection/coefficients \( \sigma_{\ell'\ell}^*(\lambda; x') \) in fact do not depend on \( x' \), which will be a direct corollary of Theorem 2.5, where the linear system whose unknowns consist of \( \sigma_{\ell'\ell}^*(\lambda; x') \) does not have any coefficient depending on \( x' \). Thus, we can simply denote

\[ \sigma_{\ell'\ell}^*(\lambda) = \sigma_{\ell'\ell}^*(\lambda; x') \]  

(2.23)

in the above reaction field decomposition.

Note that when using local coordinates \( d_{\ell}^* \) and \( d_{\ell'}^* \), each coefficient \( \sigma_{\ell'\ell}^*(\lambda) \to 0 \) as \( \lambda \to \infty \), which will be explained in details in the next section. This guarantees that each reaction field component \( u_{\ell'\ell}^*(r; r') \) is well-defined in the above integral form.

Also, there are simply coordinate shifts from (2.12) to the reaction field decomposition, so

\[ \sigma_{0\ell'}^\downarrow = \sigma_{L\ell'}^\uparrow = \sigma_{\ell,0}^\downarrow = \sigma_{\ell L}^\uparrow = 0, \quad u_{0\ell'}^\downarrow = u_{L\ell'}^\uparrow = u_{\ell,0}^\downarrow = u_{\ell L}^\uparrow = 0, \quad *,* \in \{\uparrow, \downarrow\}. \]  

(2.24)

Remark 2.3. The integrand of (2.20) may have real poles, whose integration should be treated as the limiting case of the field in lossy physical media. To understand the real poles, we adopt the perturbed wave numbers (2.11) with the perturbed branch cut described in Figure 2.1.

Let \( \lambda_v \) be a real pole of \( \sigma_{\ell'\ell}^*(\lambda) \) in the integrand of (2.20), which is known as a surface wave pole [42, 70]. Integration across the surface wave pole is understood as the limiting case of the perturbed system with lossy media as mentioned above. For simplicity, suppose
\( \sigma^{**}_{\ell' \ell}(\lambda) = \sigma(\lambda; k_1, \cdots, k_L) \) is the limit of the perturbed field \( \sigma(\lambda; \tilde{k}_1, \cdots, \tilde{k}_L) \) with pole \( \tilde{\lambda}_\nu \), and \( \tilde{\lambda}_\nu \to \lambda_\nu \in (x_0, x_1) \) as all the \( \epsilon_\ell \to 0^+ \). Let \( \sigma_\nu = \lim_{\lambda \to \lambda_\nu} \sigma(\lambda, \lambda - \lambda_\nu) \). Given any smooth function \( h(\lambda) \), the limiting integral \( \int_{x_0}^{x_1} h(\lambda) \sigma(\lambda) d\lambda \) is evaluated using the formula

\[
\int_{x_0}^{x_1} h(\lambda) \sigma(\lambda; \tilde{k}_1, \cdots, \tilde{k}_L) d\lambda \to \int_{x_0}^{x_1} \left( h(\lambda) \sigma(\lambda) - \frac{h(\lambda_\nu) \sigma_\nu}{\lambda - \lambda_\nu} \right) d\lambda + p.v. \int_{x_0}^{x_1} \frac{h(\lambda_\nu) \sigma_\nu}{\lambda - \lambda_\nu} d\lambda \pm i\pi h(\lambda_\nu) \sigma_\nu.
\]

(2.25)

The \( \pm \) sign is positive (or negative) when the perturbed pole \( \tilde{\lambda}_\nu \to \lambda_\nu \) from the upper (or the lower) half of the complex plane, and the principal value part vanishes if \( (x_0, x_1) = (-\infty, +\infty) \).

In a well-posed problem, the poles will be at most of order one, and \( \tilde{\lambda}_\nu \) should remain in one side of the half planes as all the perturbation parameters \( \epsilon_\ell \) are sufficiently small, otherwise, the limit of the integral does not exist and the field is not well-defined. Also, 0 can not be a surface wave pole, otherwise the surface wave does not propagate [42, 70].

Remark 2.4. Modes of the layered system are classified as the radiation modes, the guided modes (the real poles) and the leaky modes (the other complex poles) [42].

2.2.2. Sublinear bound of the reflection/transmission coefficients

With the reaction field decomposition (2.19), we can rewrite the interface equations (2.8) in terms of the reflection/transmission coefficients \( \sigma^{**}_{\ell' \ell}(\lambda) \) as

\[
-a_\ell \sigma^{\uparrow \ast}_{\ell' \ell} - a_\ell e_\ell \sigma^{\downarrow \ast}_{\ell' \ell} + a_{\ell+1} e_{\ell+1} \sigma^{\uparrow \ast}_{\ell+1, \ell'} + a_{\ell+1} \sigma^{\downarrow \ast}_{\ell+1, \ell'} = v^{\ast}_{\ell, \ell'},
\]

\[
b_\ell h_\ell \sigma^{\uparrow \ast}_{\ell' \ell} - b_\ell h_\ell e_\ell \sigma^{\downarrow \ast}_{\ell' \ell} - b_{\ell+1} h_{\ell+1} e_{\ell+1} \sigma^{\uparrow \ast}_{\ell+1, \ell'} + b_{\ell+1} h_{\ell+1} \sigma^{\downarrow \ast}_{\ell+1, \ell'} = w^{\ast}_{\ell, \ell'},
\]

(2.26)

for \( \ell = 0, \cdots, L - 1, \ast \in \{\uparrow, \downarrow\} \), where

\[
v^{\uparrow}_{\ell, \ell'} = \delta_{\ell, \ell'} \frac{a_\ell}{4\pi h_\ell}, \quad v^{\downarrow}_{\ell, \ell'} = -\delta_{\ell+1, \ell'} \frac{a_{\ell+1}}{4\pi h_{\ell+1}}, \quad w^{\uparrow}_{\ell, \ell'} = \delta_{\ell, \ell'} \frac{b_\ell}{4\pi}, \quad w^{\downarrow}_{\ell, \ell'} = \delta_{\ell+1, \ell'} \frac{b_{\ell+1}}{4\pi}.
\]

(2.27)
and the coefficients

\[ h_\ell = \sqrt{\lambda^2 - k_\ell^2}, \quad 0 \leq \ell \leq L, \quad (2.28) \]
\[ e_\ell = e^{-h_\ell(d_\ell-1-d_\ell)}, \quad 1 \leq \ell \leq L - 1, \quad e_0 = e_L = 0, \quad (2.29) \]

As \( e_0 \) and \( e_L \) vanish in (2.26), the corresponding terms will be ignored from the equations. All the interface equations (2.26) altogether form a pair of linear systems of unknowns \( \sigma^\uparrow_{\ell'} \) consisting of components \( \sigma^\uparrow_{\ell\ell}' \), and unknowns \( \sigma^\downarrow_{\ell'} \) consisting of components \( \sigma^\downarrow_{\ell\ell}' \), in the form

\[ A(\lambda)\sigma^\uparrow_{\ell'}(\lambda) = b^\uparrow_{\ell'}, \quad A(\lambda)\sigma^\downarrow_{\ell'}(\lambda) = b^\downarrow_{\ell'}, \quad (2.30) \]

where \( A \) does not depend on the source layer \( \ell' \) or the source-related direction \( \star \).

Let \( k_{\text{max}} = \max_{0 \leq \ell \leq L} k_\ell \) be the maximum of the wave numbers in the layered medium. The following theorem claims that (2.30) is solvable when \( \lambda \in \mathbb{R}^+ \) is sufficiently large, so the coefficient matrix \( A \) of the linear systems in (2.30) is of full rank.

**Theorem 2.5.** For any \( \lambda \in (k_{\text{max}}, +\infty) \), the linear systems (2.30) have unique solutions.

**Proof.** We will establish a recursive solver for \( \sigma^\uparrow_{\ell'} \). A recursive solver can be found for \( \sigma^\downarrow_{\ell'} \) with the same method. The idea has been first proposed by Bo Wang in [78] and [77].

Since \( \lambda > k_\ell \) for any \( \ell = 0, \cdots, L \), we have each \( h_\ell = \sqrt{\lambda^2 - k_\ell^2} \in \mathbb{R}^+ \), and each \( e_\ell \in [0, 1) \).

When treating \( \sigma_{\ell',\ell}^\uparrow \) and \( \sigma_{\ell',\ell}^\downarrow \) as unknowns in each equation of (2.26), the equations can be solved with a transition matrix

\[ \begin{bmatrix} \sigma_{\ell',\ell}^\uparrow \\ \sigma_{\ell',\ell}^\downarrow \end{bmatrix} = T_{\ell,\ell+1} \begin{bmatrix} \sigma_{\ell+1,\ell'}^\uparrow \\ \sigma_{\ell+1,\ell'}^\downarrow \end{bmatrix} + \frac{1}{2} \begin{bmatrix} (a_\ell)^{-1} & (b_\ell h_\ell)^{-1} \\ -(a_\ell e_\ell)^{-1} & -(b_\ell h_\ell e_\ell)^{-1} \end{bmatrix} \begin{bmatrix} -v_\ell^\uparrow \\ w_\ell^\uparrow \end{bmatrix}, \quad 1 \leq \ell \leq L - 1, \quad (2.31) \]
where

\[
T^{\ell,\ell+1} = \frac{1}{2e^\ell} \begin{bmatrix} e^\ell & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma^+_{\ell+1} & \gamma^-_{\ell+1} \\ \gamma^-_{\ell} & \gamma^+_{\ell} \end{bmatrix} \begin{bmatrix} e^{\ell+1} & 0 \\ 0 & 1 \end{bmatrix}, \quad \gamma^\pm_{\ell} = \frac{a_{\ell+1}}{a^\ell} \pm \frac{b_{\ell+1} h_{\ell+1}}{b^\ell h^\ell}, \quad 1 \leq \ell \leq L-1. \tag{2.32}
\]

We can extend the definition of \( T^{\ell,\ell+1} \) and \( \gamma^\pm_{\ell} \) for \( \ell = 0 \). It’s straightforward to verify that

\[
\begin{bmatrix}
\sigma_{00'}^{\uparrow\uparrow} \\
\sigma_{00'}^{\downarrow\uparrow}
\end{bmatrix} = T^{0,1} \begin{bmatrix} \sigma_{11'}^{\uparrow\uparrow} \\
\sigma_{11'}^{\downarrow\uparrow}
\end{bmatrix} + \frac{1}{2} \begin{bmatrix} (a_0)^{-1} & (b_0 h_0)^{-1} \\ (a_0)^{-1} & -(b_0 h_0)^{-1} \end{bmatrix} \begin{bmatrix} -v^\uparrow_0 \\
w^\uparrow_0
\end{bmatrix}. \tag{2.33}
\]

Recursively define matrices

\[ A^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^{(\ell+1)} = A^{(\ell)} T^{\ell,\ell+1}, \quad 0 \leq \ell \leq L-1. \tag{2.34}\]

By induction, there exist variables \( s_1 \) and \( s_2 \) independent from the unknowns such that

\[
\begin{bmatrix}
\sigma_{00'}^{\uparrow\uparrow} \\
\sigma_{00'}^{\downarrow\uparrow}
\end{bmatrix} = A^{(L)} \begin{bmatrix} \sigma_{11'}^{\uparrow\uparrow} \\
\sigma_{11'}^{\downarrow\uparrow}
\end{bmatrix} + \begin{bmatrix} s_1 \\
s_2
\end{bmatrix}. \tag{2.35}
\]

Since \( \sigma_{00'}^{\uparrow\uparrow} = \sigma_{11'}^{\uparrow\uparrow} = 0 \), the second row of this equation is reduced to

\[
0 = A_{22}^{(L)} \sigma_{11'}^{\uparrow\uparrow} + s_2. \tag{2.36}
\]

We claim that

\[ A_{22}^{(L)} \neq 0, \tag{2.37}\]

so that \( \sigma_{11'}^{\uparrow\uparrow} \) is uniquely solved, which follows that all the \( \sigma_{11'}^{\uparrow\uparrow} \) will be uniquely decided by the recursion formula (2.31) and (2.33). Indeed this can be proven by induction on a stronger
statement that $|A_{22}^{(ℓ)}| > |A_{21}^{(ℓ)}|$. For $ℓ = 0$ it is obvious. Now suppose the statement holds for index $ℓ$. Direct calculation suggests that

$$A_{21}^{(ℓ+1)} = e_ℓ \left( A_{21}^{(ℓ)} e_ℓ \gamma_ℓ^+ + A_{22}^{(ℓ)} \gamma_ℓ^- \right),$$

(2.38)

$$A_{22}^{(ℓ+1)} = A_{21}^{(ℓ)} e_ℓ \gamma_ℓ^- + A_{22}^{(ℓ)} \gamma_ℓ^+.$$  

(2.39)

Since

$$|A_{22}^{(ℓ+1)}|^2 - |A_{21}^{(ℓ)} e_ℓ \gamma_ℓ^+ + A_{22}^{(ℓ)} \gamma_ℓ^-|^2 = \left( |A_{22}^{(ℓ)}|^2 - |A_{21}^{(ℓ)}|^2 |e_ℓ|^2 \right) \left( |γ_ℓ^+|^2 - |γ_ℓ^-|^2 \right) > 0,$$

(2.40)

it immediately follows that $|A_{22}^{(ℓ+1)}|^2 > |A_{21}^{(ℓ+1)}|^2$, which completes the proof.

Remark 2.6. The same result holds if $λ ∈ (0, \min_{0 ≤ ℓ ≤ L} k_ℓ)$.

The Cramer’s rule can be applied to solve the linear systems (2.30), writing each $σ_{ℓℓ}^{**}$ in terms of the ratio of two determinants of matrices, one of which is det $A$. The asymptotic behavior of $σ_{ℓℓ}^{**}(λ)$ can be studied via this formulation. Let

$$\mathcal{P} = \{ p(e_1, \cdots, e_{L-1}, h_0, \cdots, h_L) : p \text{ is a } 2L\text{-variable polynomial} \}. $$

(2.41)

Obviously, each $σ_{ℓℓ}^{**}$ can be written in terms of the ratio of two functions in $\mathcal{P}$.

Lemma 2.7 (Asymptotic behavior of functions in $\mathcal{P}$). Take any $T ∈ \mathbb{R}^+$, define the open set

$$Ω_T = \{ v + wi : v > 0, -vT < w < vT \} \setminus \{ 0, k_{\text{max}} \}$$

(2.42)

in the complex plane. Then, for any function $f(λ) ∈ \mathcal{P}$ that is not identical to zero, there exist constants $A ∈ \mathbb{R}$, $m ∈ \mathbb{Z}$ and $B ∈ \mathbb{C}$, $B ≠ 0$ such that

$$f(λ) \sim e^{Aλ} λ^m B$$

(2.43)
as \( \Re \lambda \to \infty, \lambda \in \Omega_T \).

**Proof.** Let \( \mathcal{G} \) be the collection of holomorphic functions in \( \Omega_T \) such that the number of nonzero terms with positive exponent is finite in the Laurent series at \( \infty \), namely,

\[
\mathcal{G} = \left\{ g(\lambda) = \sum_{n=0}^{\infty} c_n \lambda^{m-n} : m \in \mathbb{Z}, c_n \in \mathbb{C}, c_0 \neq 0, g(\lambda) \text{ is holomorphic in } \Omega_T \right\}. \tag{2.44}
\]

One can verify that each \( h_\ell \in \mathcal{G} \), because it is holomorphic in \( \Omega_T \), and

\[
h_\ell = \sqrt{\lambda^2 - k_\ell^2} = \sum_{n=0}^{\infty} \frac{\sqrt{\pi}(-k_\ell^2)^n}{2\Gamma(n+1)\Gamma(-n+\frac{3}{2})} \lambda^{1-2n}. \tag{2.45}
\]

Then, let \( \mathcal{S} \) be the collection of all holomorphic functions \( s(\lambda) \) on \( \Omega_T \) in the form

\[
\mathcal{S} = \left\{ s(\lambda) = \sum_{q=1}^{Q} e^{A_q \lambda} g_q(\lambda) : Q \geq 0, A_1 > \cdots > A_Q, \text{ each } g_q \in \mathcal{G} \right\}. \tag{2.46}
\]

We claim that each \( e_\ell \in \mathcal{S} \). To show this, notice that \( e_\ell = e^{-\sqrt{\lambda^2-k_\ell^2}(d_{\ell-1}-d_\ell)} \) is holomorphic in \( \Omega_T \), and that \( e_\ell = e^{-\lambda(d_{\ell-1}-d_\ell)} e^{-\sqrt{\lambda^2-k_\ell^2}(d_{\ell-1}-d_\ell)} \). For the second factor, let \( \mu = \lambda^{-1} \), we have

\[
e^{-\sqrt{\lambda^2-k_\ell^2}(d_{\ell-1}-d_\ell)} = \exp \left(-\left(d_{\ell-1} - d_\ell\right) \sum_{n=0}^{\infty} \frac{\sqrt{\pi}(-k_\ell^2)^{n+1}}{2\Gamma(n+2)\Gamma(-n+\frac{1}{2})} \mu^{2n+1} \right), \tag{2.47}
\]

which is regular in a neighborhood of \( \mu = 0 \). Therefore, the Laurent series at 0 in the \( \mu \)-plane has zero principle part, which implies \( e^{-\sqrt{\lambda^2-k_\ell^2}(d_{\ell-1}-d_\ell)} \in \mathcal{G} \) and \( e_\ell \in \mathcal{S} \).

It’s obvious that \( \mathcal{G} \subset \mathcal{S} \), and that \( \mathcal{S} \) is a ring with function addition and multiplication, so by induction on the degree of polynomials, we have

\[
\mathcal{P} \subset \mathcal{S}. \tag{2.48}
\]
For any function \( f(\lambda) \in \mathcal{P} \subset \mathcal{S} \) which is not identical to zero, write

\[
 f(\lambda) = e^{A\lambda}g(\lambda) + \sum_{q=2}^{Q} e^{A_q\lambda}g_q(\lambda), \quad g(\lambda) = B\lambda^{m} + \sum_{n=1}^{\infty} c_n\lambda^{m-n}, \quad A > A_2 > \cdots > A_Q \tag{2.49}
\]
as in the definitions of \( \mathcal{S} \) and \( \mathcal{G} \), respectively. Since in \( \Omega_T \) we have \( \Re \lambda \leq |\lambda| \leq \sqrt{1 + T^2|\Re \lambda|} \), the limit as \( |\lambda| \to \infty \) and the limit as \( \Re \lambda \to \infty \) happen together. On can see that as \( |\lambda| \to \infty \), the leading term of each \( g_q(\lambda) \) dominates the Laurent series, and when \( \Re \lambda \to \infty \), \( e^{A\lambda}\lambda^{m}B \) dominates the expansion of \( f(\lambda) \) compared to other \( e^{A_q\lambda}g_q(\lambda) \) terms. In total,

\[
 f(\lambda) \sim e^{A\lambda}\lambda^{m}B \tag{2.50}
\]
as \( \Re \lambda \to \infty \).

Remark 2.8. Given any two nonzero asymptotic orders \( e^{A_1\lambda}\lambda^{m_1}B_1 \) and \( e^{A_2\lambda}\lambda^{m_2}B_2 \) where \( B_1, B_2 \neq 0 \), we can always compare their orders, i.e. the limit of their ratio in \( \Omega_T \) as \( \Re \lambda \to \infty \) is either infinity or a complex number.

Theorem 2.9 (Sublinear bound). With the conditions introduced in Lemma 2.7, for \( \lambda \in \Omega_T \), as \( \Re \lambda \to \infty \), all the coefficients \( \sigma_{\ell\ell}^*(\lambda) \) have asymptotic order up to \( |\lambda|^{-1} \).

Proof. First, we claim each coefficient \( \sigma_{\ell\ell}^*(\lambda) \) has an asymptotic order in the form \( e^{A\lambda}\lambda^{m}B \). Write \( \sigma_{\ell\ell}^*(\lambda) \) in terms of two determinants which belong to \( \mathcal{P} \), i.e.

\[
 \sigma_{\ell\ell}^*(\lambda) = \frac{p_1(e_1, \cdots, e_{L-1}, h_0, \cdots, h_L)}{p_2(e_1, \cdots, e_{L-1}, h_0, \cdots, h_L)}, \tag{2.51}
\]
where \( p_1, p_2 \in \mathcal{P} \) and \( p_2 = \det A \). By Theorem 2.5, \( p_2 \neq 0 \), so by Lemma 2.7, there exist \( A_2 \in \mathbb{R}, m_2 \in \mathbb{Z} \) and \( B_2 \in \mathbb{C}, B_2 \neq 0 \) such that

\[
 p_2 = \det A \sim e^{A_2\lambda}\lambda^{m_2}B_2 \tag{2.52}
\]
as $\Re \lambda \to \infty$. For the numerator, if $p_1 \equiv 0$, then $\sigma_{\ell\ell'}^*(\lambda) \equiv 0$, otherwise there exist $A_1 \in \mathbb{R}$, $m_1 \in \mathbb{Z}$ and $B_1 \in \mathbb{C}$, $B_1 \neq 0$ such that

$$p_1 \sim e^{A_1 \lambda} \lambda^{m_1} B_1,$$

(2.53)

so

$$\sigma_{\ell\ell'}^*(\lambda) \sim e^{(A_1 - A_2) \lambda} \lambda^{m_1 - m_2} \frac{B_1}{B_2}$$

(2.54)

as $\Re \lambda \to \infty$.

If every $\sigma_{\ell\ell'}^*(\lambda) \equiv 0$, the proof is trivial. Otherwise, without loss of generality, suppose among all the reflection/transmission coefficients, $\sigma_{\ell\ell'}^+$ has the highest asymptotic order as $\Re \lambda \to \infty$. Suppose for contradiction that

$$\lim_{\Re \lambda \to \infty} \frac{1}{\lambda \sigma_{\ell\ell'}^+(\lambda)} = 0.$$

(2.55)

In the interface equations (2.26), by dividing the first equation by $\sigma_{\ell\ell'}^+$, we have

$$-a_\ell - a_\ell e_\ell \frac{\sigma_{\ell+1,\ell'}^+}{\sigma_{\ell\ell'}^+} + a_{\ell+1} e_{\ell+1} \frac{\sigma_{\ell+1,\ell'}^+}{\sigma_{\ell\ell'}^+} + a_{\ell+1} \frac{\sigma_{\ell+1,\ell'}^+}{\sigma_{\ell\ell'}^+} = \frac{v_{\ell\ell'}^+}{\sigma_{\ell\ell'}^+}.$$  

(2.56)

Since as $\Re \lambda \to \infty$, $e_\ell \to 0$, $e_{\ell+1} \to 0$, and $h_\ell, h_{\ell+1} \sim \lambda$, by taking the limit, we get

$$\lim_{\Re \lambda \to \infty} \sigma_{\ell+1,\ell'}^+ = \frac{a_\ell}{a_{\ell+1}} > 0.$$  

(2.57)

By dividing the second equation of (2.26) by $\sigma_{\ell\ell'}^+$ and taking the limit, similarly, we get

$$\lim_{\Re \lambda \to \infty} \frac{\sigma_{\ell+1,\ell'}^+}{\sigma_{\ell\ell'}^+} = -\frac{b_\ell}{b_{\ell+1}} < 0,$$  

(2.58)

which contradicts the previous inequality.  \quad \square
Remark 2.10. The asymptotic estimate suggests that each \( \sigma_{\ell\ell}^*(\lambda) \) has a finite number of poles in the region \( \Omega_T \).

2.3. Far-field expansions of the reaction field

In this section, we begin with some properties of the Bessel functions of the first kind, which inspires an alternative derivation of the ME of the free-space Green’s function. These properties will be the key to deriving various far-field expansions for the reaction field \( u^r(r; r') \) of the LMGF.

2.3.1. An identity and some estimates on Bessel functions of the first kind

Recall the Bessel generating function \( g(z, \omega) = \exp\left(\frac{z}{2}(\omega - \omega^{-1})\right) = \sum_{p=-\infty}^{\infty} J_p(z)\omega^p. \) (2.59)

The identity (2.59) expresses a plane wave function in terms of cylindrical functions, in contrast to the Sommerfeld integral representation of the Green’s function, which expresses cylindrical functions in terms of plane waves (2.65). This duality facilitates the derivation of the far-field expansions in this chapter.

The above series converges absolutely, which is a corollary of the following lemma.

Lemma 2.11 (An estimate on Bessel functions of the first kind). Let \( p \in \mathbb{Z}, z \in \mathbb{C}, p \) and \( z \) are not both zero. Then

\[
|J_p(z)| \leq \frac{1}{|p|!} \left(\frac{|z|}{2}\right)^{|p|} e^{3|z|}.
\]

Proof. When \( p \geq 0 > -\frac{1}{2} \), the inequality is given by [1, (9.1.62)]. Then, the identity

\[
J_p(z) = (-1)^p J_{-p}(z)
\]
covers the case $p < 0$.

In particular, for $z \in \mathbb{R}$ and $z \geq 0$, the inequality

$$|J_p(z)| \leq \frac{1}{|p|!} \left( \frac{z}{2} \right)^{|p|}$$

(2.60)

(with the convention $0^0 = 1$) will be used to derive the exponential convergence estimates for far-field expansions in this section.

2.3.2. The multipole expansion of the free-space Green’s function revisited

We present an alternative derivation of the well-known ME of the free-space Green’s function. Consider $N$ sources with strength $q_j$ placed at locations $r_j = (x_j, y_j)$, $j = 1, 2, \ldots, N$ within a circle centered at $r_c = (x_c, y_c)$ with a radius $r$ in the free space $\mathbb{R}^2$, then, the field located at $r$ due to all sources is given by

$$u_f(r) = \sum_{j=1}^{N} q_j G_f(r; r_j),$$

(2.61)

where

$$G_f(r; r_j) = \frac{i}{4} H_0^{(1)}(k|r - r_j|)$$

(2.62)

is the free-space Green’s function, $k$ is the wave number, and $H_0^{(1)}$ is the zeroth order Hankel function of the first kind. A target $r$ is well-separated from the sources if

$$|r - r_c| \geq 2r.$$  

(2.63)

By Graf’s addition theorem [1], the free-space Green’s function for the well-separated sources $r_j$ and the target $r$ can be compressed with a multipole expansion

$$G_f(r; r_j) = \frac{i}{4} \sum_{p=-\infty}^{\infty} \alpha_p H_p^{(1)}(k|r - r_c|) e^{i p \theta_c} \approx \frac{i}{4} \sum_{|p| < P} \alpha_p H_p^{(1)}(k p_c) e^{i p \theta_c}$$

(2.64)
where \( \alpha_p = \sum_{j=1}^{N} q_j J_p(k \rho_j) e^{-ip \theta_j} \), \((\rho_c, \theta_c)\) are the polar coordinates of \( r - r_c \), \((\rho_j, \theta_j)\) are the polar coordinates of \( r_j - r_c \), and the truncation index \( P \) is a constant independent of the number of the sources \( N \) [65].

On the other hand, the multipole expansion can also be derived in the frequency domain using (2.59) as follows. Consider one source \( r_j \) and suppose \( y - y_j > 0, \ y - y_c > 0 \) for simplicity. The interaction between \( r \) and \( r_j \) can be represented by a Sommerfeld integral of plane waves [21],

\[
G^f(r; r_j) = \frac{i}{4} H_0^{(1)}(k |r - r_j|) = \frac{i}{4} \frac{1}{i \pi} \int_{-\infty}^{\infty} \frac{e^{-\sqrt{\lambda^2 - k^2} (y - y_j)}}{\sqrt{\lambda^2 - k^2}} e^{i \lambda (x - x_j)} d\lambda, \quad (2.65)
\]

while each term \( H_p^{(1)}(k \rho_c) e^{ip \theta_c} \) in (2.64) has a similar representation [21]

\[
H_p^{(1)}(k \rho_c) e^{ip \theta_c} = \frac{1}{i \pi} \int_{-\infty}^{\infty} \frac{e^{-\sqrt{\lambda^2 - k^2} (y - y_c)}}{\sqrt{\lambda^2 - k^2}} e^{i \lambda (x - x_c)} (-i)^p \left( \frac{\lambda - \sqrt{\lambda^2 - k^2}}{k} \right)^p d\lambda. \quad (2.66)
\]

These integral forms give an alternative derivation for the multipole expansion of \( \frac{i}{4} H_0^{(1)}(k |r - r_j|) = \frac{i}{4} H_0^{(1)}(k (r - r_c) + (r_c - r_j)) \) with separable plane wave factors in the integrands.
involving \((r - r_c)\) and \((r_c - r_j)\),

\[
\frac{i}{4} H_0^{(1)}(k|r - r_j|) = \frac{i}{4} \frac{1}{i\pi} \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2}(y-y_j)} e^{i\lambda(x-x_j)} d\lambda \\
= \frac{i}{4} \frac{1}{i\pi} \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2}(y-y_c)} e^{i\lambda(x-x_c)} e^{-\sqrt{\lambda^2 - k^2}(y_c-y_j)+i\lambda(x_c-x_j)} d\lambda \\
= \frac{i}{4} \frac{1}{i\pi} \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2}(y-y_c)} e^{i\lambda(x-x_c)} \cdot g(k \rho_j, -ie^{-i\theta_j} w(\lambda)) d\lambda \\
= \frac{i}{4} \frac{1}{i\pi} \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2}(y-y_c)} e^{i\lambda(x-x_c)} \cdot \sum_{p=-\infty}^{\infty} J_p(k \rho_j)e^{-ip\theta_j} (-iw(\lambda))^p d\lambda \\
= \frac{i}{4} \sum_{p=-\infty}^{\infty} J_p(k \rho_j)e^{-ip\theta_j} \cdot H_p^{(1)}(k \rho_c)e^{i\theta_c} \\
\]

where

\[
w(\lambda) = \frac{\lambda - \sqrt{\lambda^2 - k^2}}{k}. \quad (2.67)
\]

The interchangeability of the summation and the integration is verified by the validity of the identity itself, i.e. the Graf’s addition theorem.

2.3.3. The far-field expansions and translations

For the sake of convenience, we focus on the interaction between one source and one target unit charge. We will derive far-field expansions for each reaction field component \(u_{\ell\ell}'(r; r')\) in a natural generalization of the free-space case, then show their exponential convergence. The derivation makes use of the following two types of series expansions.

Suppose \((\rho_0, \theta_0)\) are the polar coordinates of \((x_0, y_0)\). Denote

\[
w_\ell(\lambda) = \frac{\lambda - \sqrt{\lambda^2 - k^2}}{k_\ell}, \quad 0 \leq \ell \leq L. \quad (2.68)
\]
By using the Bessel generating function (2.59), we have

\[ e^{-\sqrt{\lambda^2 - k^2} y_0 - i\lambda x_0} = g \left( k_{e} \rho_0, -ie^{i\tau_0} w_{e}(\lambda) \right) \]

\[ = \sum_{p=-\infty}^{\infty} J_p(k_{e} \rho_0) e^{i p r^* \theta_0} \cdot (-i w_{e}(\lambda))^p, \quad (2.69) \]

\[ e^{-\sqrt{\lambda^2 - k^2} y_0 + i\lambda x_0} = g \left( k_{e} \rho_0, ie^{i\tau_0} w_{e}(\lambda)^{-1} \right) \]

\[ = \sum_{m=-\infty}^{\infty} J_m(k_{e} \rho_0) e^{i m r^* \theta_0} \cdot (i w_{e}(\lambda)^{-1})^m. \quad (2.70) \]

For the ME, we split the difference \( r - r' = (r - r_c) + (r_c - r') \), namely, we shift the source \( r' \) to a common source center \( r_c = (x_c, y_c) \), which is assumed to be on the same side of the interface \( y = d^* \), i.e. \( y_c = d^* \) and \( y' = d^* \) have the same sign. Let \((\rho_c', \theta_c')\) be the polar coordinates of \( r' - r_c \). Using (2.69) with \((\rho_0, \theta_0) = (\rho_c', \theta_c')\) and the separability of the plane wave factor \( E_{\ell \ell'}^{*}(r, r') \) (2.22), we get an approximation

\[ u_{\ell \ell'}^{**}(r; r') = \int_{-\infty}^{\infty} E_{\ell \ell'}^{**}(r, r', \lambda) \sigma_{\ell \ell'}^{**}(\lambda) d\lambda \]

\[ = \int_{-\infty}^{\infty} E_{\ell \ell'}^{**}(r, r_c, \lambda) e^{-\sqrt{\lambda^2 - k^2} (y' - y_c) + i\lambda (x_c - x')} \sigma_{\ell \ell'}^{**}(\lambda) d\lambda \]

\[ = \int_{-\infty}^{\infty} E_{\ell \ell'}^{**}(r, r_c, \lambda) \sigma_{\ell \ell'}^{**}(\lambda) \sum_{p=-\infty}^{\infty} J_p(k_{e} \rho_0') e^{i p r^* \theta_c'} (-i w_{e}(\lambda))^p d\lambda \]

\[ \approx \sum_{|p| < P} I_{p}^{**}(r, r_c) M_{p}^{*}(r', r_c) \]

where the expansion function

\[ I_{p}^{**}(r, r_c) = \int_{-\infty}^{\infty} E_{\ell \ell'}^{**}(r, r_c, \lambda) \sigma_{\ell \ell'}^{**}(\lambda) (-i w_{e}(\lambda))^p d\lambda, \quad (2.72) \]

and the ME coefficient

\[ M_{p}^{*}(r', r_c) = J_p(k_{e} \rho_0') e^{i p r^* \theta_c'}. \quad (2.73) \]
For the LE, we split the difference $r - r' = (r - r^l_c) + (r^l_c - r')$, namely, we shift the target $r$ to a common target (local) center $r^l_c = (x^l_c, y^l_c)$, which is assumed to be on the same side of the interface $y = d^e_x$. Let $(\rho^l, \theta^l)$ be the polar coordinates of $r - r^l_c$. Using (2.70) with $(\rho_0, \theta_0) = (\rho^l, \theta^l)$ and the separability of the plane wave factor $E_{\ell\ell}^*(r, r')$ (2.22), we get an approximation

$$u_{\ell\ell}^*(r; r') = \int_{-\infty}^{\infty} \mathcal{E}_{\ell\ell}^*(r^l_c, \lambda) \sigma_{\ell\ell}^*(\lambda) \sum_{m=-\infty}^{\infty} J_m(k_\ell \rho^l) e^{im_r \theta^l} \cdot (iw_\ell(\lambda)^{-1})^m d\lambda \approx \sum_{|m| < M} L_{m}^*(r^l_c, r') K_{m}^*(r, r^l_c)$$

(2.74)

where the expansion function

$$K_{m}^*(r, r^l_c) = J_m(k_\ell \rho^l) e^{im_r \theta^l}, \quad (2.75)$$

and the LE coefficient

$$L_{m}^*(r^l_c, r') = \int_{-\infty}^{\infty} \mathcal{E}_{\ell\ell}^*(r^l_c, \lambda) \sigma_{\ell\ell}^*(\lambda) (iw_\ell(\lambda)^{-1})^m d\lambda. \quad (2.76)$$

Now, the M2L can be derived directly by using the splitting $r^l_c - r' = (r^l_c - r_c) + (r_c - r')$ in $L_{m}^*(r^l_c, r')$, i.e.

$$L_{m}^*(r^l_c, r') = \int_{-\infty}^{\infty} \mathcal{E}_{\ell\ell}^*(r^l_c, \lambda) \sigma_{\ell\ell}^*(\lambda) (iw_\ell(\lambda)^{-1})^m \sum_{p=-\infty}^{\infty} J_p(k_\ell \rho^l) e^{ip_r \theta^l} \cdot (iw_\ell(\lambda))^p d\lambda \approx \sum_{|p| < P} A_{mp}^*(r^l_c, r_c) M_{p}^*(r', r_c). \quad (2.77)$$

where the translation coefficients $A_{mp}^*(r^l_c, r_c)$ are given by

$$A_{mp}^*(r^l_c, r_c) = \int_{-\infty}^{\infty} \mathcal{E}_{\ell\ell}^*(r^l_c, \lambda) \sigma_{\ell\ell}^*(\lambda) (-iw_\ell(\lambda))^p (iw_\ell(\lambda)^{-1})^m d\lambda. \quad (2.78)$$
The L2L shifts the local center \( r_c^l \) in each integral \( L_m^*(r_c^l, r') \) to a new local center \( \tilde{r}_c^l = (\tilde{x}_c^l, \tilde{y}_c^l) \). Let \((\tilde{\rho}, \tilde{\theta})\) be the polar coordinates of \( \tilde{r}_c^l - r_c^l \). Using (2.70) with \((\rho_0, \theta_0) = (\tilde{\rho}, \tilde{\theta})\),

\[
L_m^*(\tilde{r}_c^l, r') = \int_{-\infty}^{\infty} \mathcal{E}_{l\ell}^{**}(r_c^l, r', \lambda) \sigma_{l\ell}^{**}(\lambda) \left(iw_\ell(\lambda)^{-1}\right)^m \sum_{p=-\infty}^{\infty} J_p(k_\ell \tilde{\rho}) e^{ipr_1 \tilde{\rho}} \cdot (iw_\ell(\lambda)^{-1})^p \, d\lambda
\approx \sum_{|p+m|<P} J_p(k_\ell \tilde{\rho}) e^{ipr_1 \tilde{\rho}} \int_{-\infty}^{\infty} \mathcal{E}_{l\ell}^{**}(r_c^l, r', \lambda) \sigma_{l\ell}^{**}(\lambda) \left(iw_\ell(\lambda)^{-1}\right)^m (iw_\ell(\lambda)^{-1})^p \, d\lambda
= \sum_{|p|<P} L_p^*(r_c^l, r') K_{p-m}(\tilde{r}_c^l, r').
\]

(2.79)

The M2M is the same as the free-space version with the wave number \( k_{\ell'} \), see [65].

Before we present the main result on the convergence of the series expansions above, we introduce the concept of “polarization distance” unique to the interaction in layered media. Given the target layer index \( \ell \), the source layer index \( \ell' \), and direction marks \(*, \ast \in \{\uparrow, \downarrow\}\), for a target \( r_1 = (x_1, y_1) \) and a source \( r_2 = (x_2, y_2) \), the polarization distance is defined as

\[
D_{l\ell}(r_1, r_2) = \sqrt{(x_1 - x_2)^2 + (\tau^*(y_1 - d_1^*) + \tau^*(y_2 - d_2^*))^2}, \tag{2.80}
\]

provided both \( \tau^*(y_1 - d_1^*) > 0 \) and \( \tau^*(y_2 - d_2^*) > 0 \). (Note that the polarization distance is not symmetric with respect to \( r_1 \) and \( r_2 \). This distance is in fact the distance between the target \( r_1 \) and an equivalent polarization source for the source point \( r_2 \) (see (2.92) for its definition and Figure 2.2 for an illustration of the locations of the polarization sources for different reaction components).

**Theorem 2.12 (Exponential convergence of far-field expansions of LMGF).** Let the integral \( u_{l\ell}^{**}(r; r') \) be a reaction field component derived from a well-posed Helmholtz problem in layered media as in (2.20). Then, we have the truncation error of ME (2.71)

\[
\left| u_{l\ell}^{**}(r; r') - \sum_{|p|<P} I_p^*(r, r_c^l) M_p^*(r', r_c^l) \right| \leq c_{ME}(P) \left( \frac{|r' - r_c^l|}{D_{l\ell}^{**}(r, r_c^l)} \right)^P, \tag{2.81}
\]

31
Figure 2.2: The far-field distance $\rho_c$ of the ME in various field propagation directions, determined using the polarization sources.
the truncation error of LE (2.74)

\[
|u^{**}_{\ell\ell}(r; r') - \sum_{|m| \leq M} L^{**}_{m}(r_c, r')K^{*}_{m}(r, r_c)| \leq c^{LE}(M) \left( \frac{|r - r_c|}{D^{**}_{\ell\ell}(r_c, r')} \right)^{M},
\]

(2.82)

the truncation error of M2L (2.77) for each LE coefficient

\[
|L^{**}_{m}(r_c, r') - \sum_{|p| < P} A^{**}_{mp}(r_c, r_c)M^{*}_{p}(r', r_c)| \leq c^{M2L}(P) \left( \frac{|r' - r_c|}{D^{**}_{\ell\ell}(r_c, r')} \right)^{P},
\]

(2.83)

and the truncation error of L2L (2.79) for each LE coefficient

\[
|L^{**}_{m}(\tilde{r}_c, r') - \sum_{|p| < P} L^{**}_{p}(r_c, r_c)K^{*}_{p-m}(\tilde{r}_c, r_c)| \leq c^{L2L}(P) \left( \frac{|\tilde{r}_c - r_c|}{D^{**}_{\ell\ell}(r_c, r')} \right)^{P},
\]

(2.84)

for some functions \( c^{ME}(\cdot) \), \( c^{LE}(\cdot) \), \( c^{M2L}(\cdot) \) and \( c^{L2L}(\cdot) \) having polynomial growth rates, provided that for some given \( \mu > 1 \), the far-field conditions measured with the polarization distances,

\[
D^{**}_{\ell\ell}(r, r_c) \geq \mu |r' - r_c|, \quad D^{**}_{\ell\ell}(r_c, r') \geq \mu |r - r_c|,
\]

(2.85)

\[
D^{**}_{\ell\ell}(r_c, r) \geq \mu |r' - r_c|, \quad D^{**}_{\ell\ell}(r', r_c) \geq \mu |\tilde{r}_c - r_c|
\]

hold, respectively. If all the sources, targets and the centers involved above are bounded by a given box, the distances from every center to its nearby interface have a given nonzero lower bound, and there exist \( 0 < \rho_m \leq \rho_M \) such that

\[
\rho_m \leq D^{**}_{\ell\ell}(r, r_c), D^{**}_{\ell\ell}(r_c, r'), D^{**}_{\ell\ell}(r_c, r_c), D^{**}_{\ell\ell}(r', r_c) \leq \rho_M,
\]

then the functions \( c^{ME}(\cdot) \), \( c^{LE}(\cdot) \), \( c^{M2L}(\cdot) \) and \( c^{L2L}(\cdot) \) can be chosen to be determined by these bounds, without dependence on the actual positions of the source locations.

The proof will be given in Section 2.5.
2.3.3.1. Numerical validation of exponential convergence

Here we present some numerical examples showing the exponential convergence rates of MEs and LEs. Consider a 3-layer medium with a source $r'$ and a target $r$ both in the middle layer with

$$d_0 = -0.5, \quad d_1 = 0.5, \quad k_0 = 2.0, \quad k_1 = 3.0, \quad k_2 = 4.7,$$

(2.86)

and the parameters for interface conditions specified by

$$a_0 = a_1 = a_2 = 1, \quad b_0 = 1.0, \quad b_1 = 3.0, \quad b_2 = 5.0.$$

(2.87)

We will take a test on the reaction field component

$$u_{11}(r; r') = \int_{-\infty}^{\infty} e^{i\lambda(x-x')} - \sqrt{\lambda^2 - k_1^2(d_0-y) - \sqrt{\lambda^2 - k_1^2(d_0-y')}} \sigma_{11}(\lambda) d\lambda,$$

(2.88)

where the closed form of $\sigma_{11}(\lambda)$ is given by

$$\sigma_{11}(\lambda) = \frac{1}{4\pi h_1 (b_0 h_0 + b_1 h_1)(b_2 h_2 + b_1 h_1) - e^2 (b_0 h_0 - b_1 h_1)(b_2 h_2 - b_1 h_1)}. $$

(2.89)

$\sigma_{11}(\lambda)$ has a pair of real poles at $\lambda = \pm k_1$. If we consider the perturbed wave numbers in each layer $\tilde{k}_l = k_l + \epsilon_l i$, $0 < \epsilon_l \ll 1$, then the perturbed real poles are $\pm (k_1 + \epsilon_1 i)$ with positive and negative imaginary part, respectively. Hence (2.25) will be used to evaluate the integrals.

Suppose the source center $r_c$ and the target center $r_c'$ are in the middle layer, and the far-field conditions $D_{11}(r, r_c) > |r' - r_c|$ and $D_{11}(r_c, r') > |r - r_c'|$ are met. By Theorem 2.12, for $u_{11}(r; r')$, the relative error of the ME at source center $r_c$, and that of the LE at
target center $r^i_c$ are defined for a given truncation index $P$

$$
e^ME_P = \left| u_{11}^{\uparrow\downarrow}(r; r') - \sum_{|\rho|<P} I^\downarrow_P(r, r_c) M^\downarrow_P(r', r_c) \right| / \left| u_{11}^{\uparrow\downarrow}(r; r') \right|,$$

$$
e^LE_P = \left| u_{11}^{\uparrow\downarrow}(r; r') - \sum_{|\rho|<P} I^\downarrow_P(r^i_c, r') K^\uparrow_P(r, r^i_c) \right| / \left| u_{11}^{\uparrow\downarrow}(r; r') \right|. \tag{2.90}$$

For comparison, we define the reference exponential convergence ratio

$$r_{ME} = \frac{|r' - r_c|}{D_{11}^{\uparrow\downarrow}(r, r_c)}, \quad r_{LE} = \frac{|r - r^i_c|}{D_{11}^{\uparrow\downarrow}(r^i_c, r')} \tag{2.91}$$

Fix the source center $r_c = (0.0, 0.0)$ and the target center $r^i_c = (0.6, 0.2)$. We select three target-source pairs for numerical testing: (1) $r = (0.5, 0.3), r' = (0.3, 0.4)$; (2) $r = (0.5, 0.4), r' = (-0.1, -0.3)$; (3) $r = (0.0, 0.2), r' = (-0.1, 0.2)$. For each pair we compute and plot the relative errors of ME and LE of $u_{11}^{\uparrow\downarrow}(r, r')$ for $P = 3, 4, \cdots, 12$ in Figure 2.3. Then we compare the results with the reference exponential convergence rates indicated by the corresponding colored dashed lines with slopes $\log_{10} r_{ME}$ and $\log_{10} r_{LE}$, respectively, which shows that the relative errors decay at the expected exponential rates, determined by the polarization distance.

2.4. The FMM framework

In the far-field conditions (2.85), the polarization distances $D_{\uparrow\downarrow}^{\uparrow\downarrow}$ play the role of the far-field distances as in the free-space cases for the FMM implementation. To make use of this fact for the setup of FMM, we define the polarization source as follows.

**Definition 2.13 (Polarization source).** Let $r_2 = (x_2, y_2)$ be any source point satisfying $\tau^*(y_2 - d^*_r) > 0$. Let

$$P_{\uparrow\downarrow}^{\uparrow\downarrow} : r_2 \mapsto \tilde{r}_2 = (x_2, d^*_r - \tau^* \tau^*(y_2 - d^*_r)) \tag{2.92}$$
be a bijective mapping. $P_{ll'}^{**}(r_2)$ is called the polarization source of $r_2$.

Clearly,

$$D_{ll'}^{**}(r_1, r_2) = \| r_1 - P_{ll'}^{**}(r_2) \|,$$

where $\| \cdot \|$ is the Euclidean distance. Figure 2.2 shows how $P_{ll'}^{**}$ maps the sources to their equivalent polarization sources.

The FMM for layered media can then be set up to evaluate each reaction field component $u_{ll'}^{**}$ as follows: $P_{ll'}^{**}$ maps the source layer $l'$ to a neighboring layer (below or above) of the target layer $l$, where all the far-field distances become Euclidean as in (2.93). Then, to calculate the interaction due to any of the reaction component $u_{ll'}^{**}$, we simply move the source charges to the locations of their corresponding equivalent polarization sources. Implementations for the Helmholtz equation, the Laplace’s equation and the linearized Poisson–Boltzmann equation in 3-D layered media based on this approach are given in the following chapters, see also [76, 78, 79].
2.5. Proof of exponential convergence of far-field expansions

In this section we give the convergence estimates on general Bessel-type expansions, of which Theorem 2.12 will be a special case. The Bessel-type expansions are defined as follows. Let \( k > 0, (\rho, \theta), (\rho', \theta') \) be the polar coordinates of \( \mathbf{r} = (x, y) \) and \( \mathbf{r}' = (x', y') \), respectively. Suppose \( y > 0, y + y' > 0 \) and \( \rho > \rho' \geq 0 \). For simplicity, define

\[
\Psi(\lambda) \equiv \Psi(\mathbf{r}, \lambda) = e^{-\sqrt{x^2 - k^2}y + i\lambda x}, \\
\Psi'(\lambda) \equiv \Psi'(\mathbf{r}', \lambda) = e^{-\sqrt{x'^2 - k^2}y' - i\lambda x'}. \tag{2.94}
\]

Then, we claim the pointwise Bessel-type expansion for a given \( \lambda_{\nu} \in \mathbb{C} \),

\[
e^{-\sqrt{x^2 - k^2}(y+y') + i\lambda_{\nu}(x-x')} = \sum_{p=-\infty}^{\infty} J_p(k\rho')e^{ip\theta'}\Psi(\mathbf{r}, \lambda_{\nu})(-iw(\lambda_{\nu}))^p \tag{2.95}
\]

and the integral Bessel-type expansion for the integration over \( \lambda \in (a, b), -\infty \leq a < b \leq +\infty \),

\[
\int_{a}^{b} e^{-\sqrt{x^2 - k^2}(y+y') + i\lambda(x-x')} f(\lambda) d\lambda = \sum_{p=-\infty}^{\infty} J_p(k\rho')e^{ip\theta'} F_p(x, y), \tag{2.96}
\]

where \( f(\lambda) \) is a complex function defined on \((a, b)\) satisfying certain conditions to be specified later, and \( F_p(x, y) \) is the expansion function

\[
F_p(x, y) = \int_{a}^{b} \Psi(\mathbf{r}, \lambda) (-i w(\lambda))^p f(\lambda) d\lambda. \tag{2.97}
\]

2.5.1. Convergence of pointwise Bessel-type expansions

We first present the convergence of (2.95).

**Lemma 2.14.** Let \( \mu > 1, k > 0 \). Suppose \((\rho', \theta')\) are the polar coordinates of \((x', y')\). Suppose \( x \in \mathbb{R}, y \in \mathbb{R}^+ \) satisfying \( \rho = \sqrt{x^2 + y^2} > \mu \rho' \geq 0 \) and \( x \cdot \Im \lambda_{\nu} \geq 0 \). Then, the
Bessel-type expansion (2.95) holds with truncation error estimate

\[ \left| \sum_{|p| \geq P} J_p(k \rho') e^{i \rho' \theta'} \Psi(r, \lambda) (-i w(\lambda)) \right|^p \leq \frac{2 \mu}{\mu - 1} \left( \frac{\rho'}{\rho} \right)^P \]  

(2.98)

for any

\[ P \geq e(|\lambda| + k/2) \rho. \]  

(2.99)

**Proof.** The equality of (2.95) is given by the Bessel generating function (2.59)

\[ e^{-\sqrt{\lambda^2 - k^2 y + i \lambda x}} = \Psi(\lambda, x, y) \]  

\[ = \sum_{p=-\infty}^{\infty} J_p(k \rho') e^{i \rho' \theta} \Psi(\lambda, x, y) \]  

(2.100)

With the given conditions, \(|\exp(-\sqrt{\lambda^2 - k^2 y + i \lambda x})| \leq 1, |w(\lambda)| \leq (2|\lambda| + k)/k.\) Hence

for each \(p,\) using Lemma 2.11,

\[ |J_p(k \rho') e^{i \rho' \theta} e^{-\sqrt{\lambda^2 - k^2 y + i \lambda x}} (-i w(\lambda))|^p \leq \frac{1}{|p|!} \left( \frac{k \rho'}{2} \right)^{|p|} \left( \frac{2|\lambda| + k}{k} \right)^{|p|}. \]

For \(|p| \geq e(|\lambda| + k/2) \rho,\) using the Stirling’s formula [64],

\[ |p|! \geq \left( \frac{|p|}{e} \right)^{|p|} \geq \left( \left| \lambda \right| + \frac{k}{2} \right)^{|p|}, \]

we have

\[ |J_p(k \rho') e^{i \rho' \theta} e^{-\sqrt{\lambda^2 - k^2 y + i \lambda x}} (-i w(\lambda))|^p \leq \left( \frac{\rho'}{\rho} \right)^{|p|} \]

which will give the estimate of the truncation error after summing over \(|p| \geq P.\)

\[ \square \]

2.5.2. Special cases of the integral Bessel-type expansion

First, we will prove (2.96) for a more general setting when the integral is defined on a bounded interval \([-k', k']\).
Lemma 2.15. Let $\mu > 1$, $k > 0$. Let $(\rho, \theta)$ and $(\rho', \theta')$ be the polar coordinates of $r = (x, y)$ and $r' = (x', y')$, respectively. Suppose $y > 0$, $\rho > \mu \rho' \geq 0$. Let $\kappa \subset \mathbb{C}$ be a complex contour which is parameterized as

$$\kappa: \lambda = \lambda(s) = a(s) + b(s)i, \quad 0 \leq s \leq 1,$$

(2.101)

where $a(s)$ and $b(s)$ are real differentiable functions. Suppose $x \cdot b(s) \geq 0$ for any $s \in [0, 1]$. Let $f(\lambda)$ be a complex function on $\kappa$ satisfying a convergence condition

$$\int_0^1 \left| f(\lambda(s)) \sqrt{a'(s)^2 + b'(s)^2} \right| ds = S < \infty.$$  

(2.102)

Then, the following series expansion

$$E_\kappa = \int_\kappa \Psi(\lambda) \Psi'(\lambda) f(\lambda) d\lambda = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\theta'} \int_\kappa \Psi(\lambda) (-iw(\lambda))^p f(\lambda) d\lambda$$

(2.103)

holds with a truncation error estimate

$$\left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\theta'} \int_\kappa \Psi(\lambda) (-iw(\lambda))^p f(\lambda) d\lambda \right| \leq \frac{2\mu S}{\mu - 1} \left( \frac{\rho'}{\rho} \right)^P$$

(2.104)

for any $P \geq e(\lambda_M + k/2) \rho$ where $\lambda_M = \max_{\lambda \in \kappa} |\lambda|$.

Proof. Using the results from the proof of Lemma 2.14, for $\lambda \in \kappa$,

$$\left| e^{-\sqrt{\lambda^2 - k^2} y + i\lambda x} \right| \leq 1, \quad |w(\lambda)|^{\pm 1} \leq \frac{2\lambda_M + k}{k},$$

so for each $p$, using Lemma 2.11,

$$\int_0^1 \left| J_p(k\rho') e^{ip\theta'} \Psi(\lambda) (-iw(\lambda))^p f(\lambda) (a'(s) + b'(s)i) \right| ds$$

$$\leq \frac{1}{|p|!} \left( \frac{k\rho'}{2} \right)^{|p|} \cdot 1 \cdot \left( \frac{2\lambda_M + k}{k} \right)^{|p|} \cdot S.$$
Hence, using the Bessel generating function (2.59) and the Fubini’s theorem,

\[
E_{\kappa} = \int_{0}^{1} \Psi(\lambda) \Psi'(\lambda) f(\lambda) (a'(s) + b'(s)i) \, ds \\
= \sum_{p=-\infty}^{\infty} \int_{0}^{1} J_{p}(k\rho') e^{ip\theta'} \Psi(\lambda) (-i\omega(\lambda))^{p} f(\lambda) (a'(s) + b'(s)i) \, ds \\
= \sum_{p=-\infty}^{\infty} \int_{\kappa}^{\infty} J_{p}(k\rho') e^{ip\theta'} \Psi(\lambda) (-i\omega(\lambda))^{p} f(\lambda) d\lambda,
\]

we obtain the equality of (2.103). When \(|p| \geq e(\lambda_{M} + k/2)\rho\), using the Stirling’s formula [64], \(|p|! \geq (|p|/e)^{|p|}\), we can show that each integral

\[
\left| \int_{\kappa}^{\infty} J_{p}(k\rho') e^{ip\theta'} \Psi(\lambda) (-i\omega(\lambda))^{p} f(\lambda) d\lambda \right| \leq \frac{1}{|p|!} \left( \frac{k\rho'}{2} \cdot \frac{2\lambda_{M} + k}{k} \right)^{|p|} S \leq S \left( \frac{\rho'}{\rho} \right)^{|p|}.
\]

By adding up the bounds for \(|p| \geq P\) we get a truncation error estimate with the following bound

\[
\left| \sum_{|p| \geq P} J_{p}(k\rho') e^{ip\theta'} \int_{\kappa}^{\infty} \Psi(\lambda) (-i\omega(\lambda))^{p} f(\lambda) d\lambda \right| \leq \sum_{|p| \geq P} S \left( \frac{\rho'}{\rho} \right)^{|p|} \leq \frac{2\mu S}{\mu - 1} \left( \frac{\rho'}{\rho} \right)^{P}.
\]

A similar result on a bounded real interval follows immediately.

**Lemma 2.16.** Let \(\mu > 1\), \(k' \geq k > 0\). Let \((\rho, \theta)\) and \((\rho', \theta')\) be the polar coordinates of \(r = (x, y)\) and \(r' = (x', y')\), respectively. Suppose \(y > 0\), \(\rho > \mu \rho' \geq 0\), and the function \(f(\lambda)\) on \([-k', k']\) satisfies \(\int_{-k'}^{k'} |f(\lambda)| d\lambda = S < +\infty\), then the integral Bessel-type expansion (2.96) holds on \([-k', k']\) with truncation error estimate

\[
\left| \sum_{|p| \geq P} J_{p}(k\rho') e^{ip\theta'} F_{p} \right| \leq \frac{2\mu S}{\mu - 1} \left( \frac{\rho'}{\rho} \right)^{P} \tag{2.105}
\]

for any \(P \geq ek'\rho\).
Proof. The same proof of Lemma 2.15 can be applied by using the following estimate instead,

\[ |w(\lambda)|^{\pm 1} = \left| \frac{\lambda - \sqrt{\lambda^2 - k^2}}{k} \right|^{\pm 1} \leq \frac{2k'}{k} \]

for any \( \lambda \in [-k', k'] \), which gives the necessary lower bound of \( P \).

Next, we consider the special case \((x, y) = (0, \rho)\) in the Bessel-type expansion (2.96) over an infinite interval.

**Lemma 2.17.** Let \( \mu > 1, k' \geq k > 0, x', y' \in \mathbb{R}, \rho > \mu \rho' \geq 0 \). Suppose \( f(\lambda) \) is a continuous function on \([k', \infty)\) such that \( |f(\lambda)| \leq C\lambda^K \) for some given positive constant \( C \) and nonnegative integer \( K \). For the integral

\[ E_p^+ = \int_{k'}^\infty e^{-\sqrt{\lambda^2 - k'^2} \rho} (-i w(\lambda))^p f(\lambda) d\lambda, \quad p \in \mathbb{Z}, \quad (2.106) \]

we have the estimate

\[ |E_p^+| \leq \int_{k'}^\infty e^{-\sqrt{\lambda^2 - k'^2} \rho} w(\lambda)^p |f(\lambda)| d\lambda \leq 3C (|p| + K)! \left( \frac{2}{\rho} \right)^{K+1} \left( \frac{k\rho}{2} \right)^{-|p|} \quad (2.107) \]

for any \( |p| \geq (k\rho)^2/4 + 1 - K \). In addition, the Bessel-type expansion (2.96) holds with \((x, y) = (0, \rho)\) on the interval \((k', \infty)\), and the truncation error is given by

\[ \left| \int_{k'}^\infty e^{-\sqrt{\lambda^2 - k'^2} (\rho + y')} + i \lambda (-x') f(\lambda) d\lambda - \sum_{|p| < P} J_p(k\rho') e^{ip\theta'} E_p^+ \right| \leq c(P, \rho) \left( \frac{P'}{\rho} \right)^P \quad (2.108) \]

for all \( P \geq (k\rho)^2/4 + 1 - K \), where

\[ c(P, \rho) = 6C(K + 1)! \left( \frac{2\mu}{\rho(\mu - 1)} \right)^{K+1} (P + K)^K. \quad (2.109) \]
Proof. First consider the estimate (2.107). Since for \( \lambda \geq k \) we have \( \sqrt{\lambda^2 - k^2} \leq \lambda \) and \( 0 \leq \lambda - \sqrt{\lambda^2 - k^2} \leq k \leq \lambda + \sqrt{\lambda^2 - k^2} \), we get \( |E_p^+| \leq Ck^{K+1}I_p \), where

\[
I_p = \int_k^\infty \frac{e^{-\sqrt{\lambda^2 - k^2} \rho}}{\sqrt{\lambda^2 - k^2}} \left( \frac{\lambda + \sqrt{\lambda^2 - k^2}}{k} \right)^{M+1} d\lambda
\]  

(2.110)

and \( M = |p| + K \). With the substitution \( v = (\lambda + \sqrt{\lambda^2 - k^2})/k \),

\[
I_p = \int_1^\infty e^{\frac{k\rho}{2}(-v+v^{-1})} v^M dv
\]

\[
\leq \int_1^\infty e^{\frac{k\rho}{2}(-v)} \left( \sum_{j=0}^{M-1} \frac{1}{j!} \left( \frac{k\rho}{2} v^{-1} \right)^j + \frac{1}{M!} \left( \frac{k\rho}{2} v^{-1} \right)^M e^{\frac{k\rho}{2} v^{-1}} \right) v^M dv
\]

\[
\leq \sum_{j=0}^{M-1} \frac{1}{j!} \left( \frac{k\rho}{2} \right)^j \int_1^\infty e^{\frac{k\rho}{2}(-v)} v^{-j} dv + \frac{1}{M!} \left( \frac{k\rho}{2} \right)^M \int_1^\infty e^{\frac{k\rho}{2}(-v+1)} dv
\]

\[
= \sum_{j=0}^{M-1} \frac{(M-j)!}{j!} \left( \frac{k\rho}{2} \right)^{2j-M-1} + \frac{1}{M!} \left( \frac{k\rho}{2} \right)^{M-1}
\]

\[
= M! \left( \frac{k\rho}{2} \right)^{-M-1} \sum_{j=0}^M c_j,
\]

where

\[
c_j = \frac{(M-j)!}{M!j!} \left( \frac{k\rho}{2} \right)^{2j}, \quad j = 0, \ldots, M.
\]  

(2.111)

One can verify that

\[
c_0 = 1, \quad c_1 = \frac{1}{M} \left( \frac{k\rho}{2} \right)^2 \leq \frac{(M-1)}{M}.
\]  

(2.112)

For \( 1 \leq j \leq M-2 \), we have

\[
\frac{c_{j+1}}{c_j} = \frac{(k\rho)^2}{4(j+1)(M-j)} \leq \frac{1}{2}
\]

For \( c_M \) we have

\[
\frac{c_M}{c_{M-1}} = \frac{(k\rho)^2}{4M} \leq 1.
\]
In sum,
\[
\sum_{j=0}^{M} c_j \leq c_0 + 2c_1 \leq 1 + \frac{2(M - 1)}{M} \leq 3,
\] (2.113)
so
\[
|E_p^+| \leq Ck^{K+1}I_p \leq 3Ck^{K+1}M! \left(\frac{k\rho}{2}\right)^{-M-1} \leq 3C (|p| + K)! \left(\frac{2}{\rho}\right)^{K+1} \left(\frac{k\rho}{2}\right)^{-|p|}. \quad (2.114)
\]
To get the expansion (2.96) with \((x, y) = (0, \rho)\) on \([k', \infty)\), by using Lemma 2.11 and (2.114), we have
\[
\int_{k'}^\infty \left| J_p(k\rho') e^{ip\rho} e^{-\sqrt{x^2-k^2} \rho} (-iw(\lambda))^p f(\lambda) \right| d\lambda \leq \frac{1}{|p|!} \left(\frac{k\rho'}{2}\right)^{|p|} \cdot 3C (|p| + K)! \left(\frac{2}{\rho}\right)^{K+1} \left(\frac{k\rho}{2}\right)^{-|p|} \leq 3C (|p| + K)! \left(\frac{2}{\rho}\right)^{K+1} \left(\frac{\rho'}{\rho}\right)^{|p|}.
\]
Now, as in (2.100), by using the Fubini’s theorem, we have the expansion of (2.108),
\[
\int_{k'}^\infty e^{-\sqrt{x^2-k^2} (\rho+y') + i\lambda (-x')} f(\lambda) d\lambda = \int_{k'}^\infty \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\rho} e^{-\sqrt{x^2-k^2} \rho} (-iw(\lambda))^p f(\lambda) d\lambda = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\rho} E_p^+,
\]
with a \(P\)-term truncation error for \(P \geq (k\rho)^2/4 + 1 - K\),
\[
\left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\rho} E_p^+ \right| \leq \sum_{|p| \geq P} 3C (|p| + K)! \left(\frac{2}{\rho}\right)^{K+1} \left(\frac{\rho'}{\rho}\right)^{|p|} \leq c(P, \rho) \left(\frac{\rho'}{\rho}\right)^P.
\]
\[\Box\]

Remark 2.18. The bound of \(|E_p^+|\) in Lemma 2.17 is shown as an analog of the asymptotic behavior of
\[
H_n^{(1)}(x) \sim (n - 1)! (x/2)^n / (i\pi)
\]
for $x > 0$ as $n \to \infty$ [1, (9.3.1)].

2.5.3. Convergence of general integral Bessel-type expansions

In order to obtain the convergence estimate of the integral Bessel-type expansion (2.96) on an infinite interval, we will take two steps. First, the Cagniard–de Hoop transform [16] will be used to convert the general $(x, y)$ case to the $(\rho, 0)$ case as discussed in Lemma 2.17, namely, the complex factor $e^{-\sqrt{\lambda^2 - k^2}y + i\lambda x}$ in (2.96) is converted to $e^{-\sqrt{\lambda^2 - k^2}\rho}$. Second, we deform the new complex contour of integration as a result of the transform to the real axis, see the illustration in Figure 2.4.

2.5.3.1. The Cagniard–de Hoop transform

Given positive real numbers $x$, $y$ and $T$ satisfying $x < Ty$. Let $(\rho, \theta)$ be the polar coordinates of $(x, y)$. Let $\beta = \frac{\pi}{2} - \theta \in (0, \frac{\pi}{2})$, then $y + xi = \rho e^{i\beta}$.

Define an open set

$$\Omega = \{ z \in \mathbb{C} : \Re z > 0, z \notin (0, k]\}.$$  \hspace{1cm} (2.115)

Then the holomorphic Cagniard–de Hoop mapping [16] $\phi : \Omega \to \mathbb{C}$ is given by

$$\phi(z) = z \cos \beta + i\sqrt{z^2 - k^2} \sin \beta.$$ \hspace{1cm} (2.116)

Consider the right branch of the hyperbola

$$\Gamma = \left\{ a + bi : a, b \in \mathbb{R}, \frac{a}{\cos \beta} = \sqrt{\frac{b^2}{\sin^2 \beta} + k^2} \right\}$$ \hspace{1cm} (2.117)

with its vertex $k \cos \beta$ on the real axis, and the upper and lower parts of $\Gamma$ denoted as $\gamma^+$ and $\gamma^-$, respectively, i.e.

$$\Gamma = \gamma^+ \cup \gamma^- \cup \{ k \cos \beta\}.$$ \hspace{1cm} (2.118)
We can easily verify that $\phi((k, +\infty)) = \gamma^+$, and $\phi(\gamma^-) = (k, +\infty)$, namely,

$$
\gamma^+ = \phi((k, +\infty)), \quad \gamma^- = \phi^{-1}((k, +\infty)),
$$

where

$$
\phi^{-1}(w) = w \cos \beta - i\sqrt{w^2 - k^2} \sin \beta.
$$

Usually, $\Gamma$ is known as the Cagniard–de Hoop contour. The two straight lines passing the origin with slopes $\pm \tan \beta$ are the asymptotes of $\gamma^\pm$, respectively.

Define regions to the right of $\Gamma$ in the first and the fourth quadrant, respectively, by

$$
D^\pm = \{ z + t : z \in \gamma^\pm, t \in \mathbb{R}^+ \}.
$$

$D^\pm$ are isomorphic as the following sequence of lemmas show.
Lemma 2.19. Let \( a, b \in \mathbb{R} \) such that \( z = a + bi \in D^- \), then \( \Re \phi(z) > 0, \Im \phi(z) > 0 \).

Proof. Let \( u, v \in \mathbb{R} \) such that \( u + vi = \sqrt{z^2 - k^2} \), then \( uv = ab < 0 \). With the convention of the branch cut (2.10), we have \( u > 0 \), so \( v < 0 \). Recall that \( \beta \in (0, \frac{\pi}{2}) \), we have \( u \sin \beta - b \cos \beta > 0 \) and \( \Re \phi(z) = a \cos \beta - v \sin \beta > 0 \). For \( \Im \phi(z) \), let

\[
Q_1 = (a^2 - b^2 - k^2)^2 + 4a^2b^2, \quad Q_2 = (a^2 - b^2 - k^2) \sin^2 \beta - 2b^2 \cos^2 \beta.
\]

(2.122)

By simple calculation, we have

\[
2u^2 \sin^2 \beta - 2b^2 \cos^2 \beta = \sqrt{Q_1 \sin^2 \beta + Q_2},
\]

so \( \sqrt{Q_1 \sin^2 \beta} = |\sqrt{Q_1 \sin^2 \beta}| > |Q_2| \), which implies

\[
\Im \phi(z) = b \cos \beta + u \sin \beta = \frac{\sqrt{Q_1 \sin^2 \beta + Q_2}}{2(u \sin \beta - b \cos \beta)} > 0.
\]

Lemma 2.20. If \( w \in \gamma^+ \), then \( \phi(z) \neq w \) for any \( z \in D^- \).

Proof. Suppose for contradiction that \( z \in D^- \), \( \phi(z) = w \). Since \( w \in \gamma^+ \), \( \exists x_0 \geq k \) such that \( w = x_0 \cos \beta + i\sqrt{x_0^2 - k^2} \sin \beta \). Therefore, \( x_0 \) and \( z \) are distinct roots of the quadratic equation \( \lambda^2 - 2\lambda w \cos \beta + w^2 = k^2 \sin^2 \beta \) of \( \lambda \). Hence

\[
z = 2w \cos \beta - x_0 = x_0 \cos(2\beta) + i\sqrt{x_0^2 - k^2} \sin(2\beta) \notin D^-
\]

because \( \Im z \geq 0 \), a contradiction.

Lemma 2.21. \( \phi|_{D^-} \) is a bijection to \( D^+ \) with inverse \( \phi^{-1}|_{D^+} \) given by (2.120).

Proof. Define \( \phi' : D^+ \to \mathbb{C} \) by

\[
\phi'(w) = w \cos \beta - i\sqrt{w^2 - k^2} \sin \beta.
\]

(2.123)
It suffices to show $\phi'$ is the inverse of $\phi$ on $D^+$, i.e. $\phi^{-1}|_{D^+} = \phi'$.

First, we will show that $\phi(D^-) \subset D^+$. By Lemma 2.19 and Lemma 2.20, $\phi(D^-)$ is a subset of the first quadrant, and it has no intersection with the hyperbola $\Gamma$. If $w = \phi(z)$ for some $z \in D^-$ and $w \notin D^+$, when we move $z$ horizontally to the left, eventually $z$ touches $\Gamma$ and $\phi(z)$ approaches the positive real axis, so the trajectory of $\phi(z)$, which must be continuous because $\phi$ is holomorphic, crosses $\Gamma$ in the first quadrant, but it contradicts with Lemma 2.20 since the intersection must has its inverse in $D^-$. Similarly (by taking complex conjugates), $\phi'(D^+) \subset D^-$. 

Second, we will show that $\phi$ is bijective on $D^-$ with inverse $\phi'$. Let $a, b \in \mathbb{R}^+$ such that $z = a + bi \in D^-$, then $w = \phi(z) \in D^+$ is one of the roots of the quadratic equation of $\lambda$

$$\lambda^2 - 2\lambda z \cos \beta + z^2 = k^2 \sin^2 \beta. \quad (2.124)$$

Let $u, v \in \mathbb{R}$ such that $\sqrt{z^2 - k^2} = u + vi$, then $u > 0$, the pair of roots are given by

$$\lambda_\pm = (a \cos \beta \mp v \sin \beta) + i(b \cos \beta \pm u \sin \beta). \quad (2.125)$$

By Lemma 2.19, $\Im w = \Im \phi(z) > 0$, so $w = \lambda^+$. Conversely, $z$ is the only root of the quadratic equation $\lambda^2 - 2\lambda w \cos \beta + w^2 = k^2 \sin^2 \beta$ in $D^-$ provided $\phi(z) = w$ by the similar reason, so $\phi$ is injective and $z = \phi'(w)$. Repeat this step for any $w' \in D^+$ and let $z' = \phi'(w')$, we have $\phi$ is surjective and $w' = \phi(\phi'(w'))$. 

2.5.3.2. The general Bessel-type expansion

With the above preparation, we can now prove the expansion (2.96) when $f(\lambda)$ has a polynomial bound in $\Omega$ and $|\lambda|$ is sufficiently large and $\Im \lambda/\Re \lambda$ is bounded. To be specific, we make the following assumptions.
Assumption 2.22. Given $T > 0$, $\epsilon_0 > 0$. Suppose $f(\lambda)$ is a complex function with branch points $\pm k_0, \cdots, \pm k_L$, it is even and is meromorphic in $\mathbb{C}$ excluding the branch cuts of $\sqrt{\lambda^2 - k_\ell^2}$, $0 \leq \ell \leq L$, with poles of order up to one. Assume that

- $f(\lambda)$ has a decomposition
  \[ f(\lambda) = \sum_{r=1}^{n_r} \frac{f_r}{\lambda - \lambda_r} + \tilde{f}(\lambda), \quad \text{and} \quad \tilde{f}(\lambda) = \sum_{c=1}^{n_c} \frac{f_c}{\lambda - \lambda_c} + \bar{\tilde{f}}(\lambda), \]  
  (2.126)

  here $\lambda_r \neq 0$ are all the real poles of $f(\lambda)$ with residue $f_r$, $\lambda_c$ are all the (complex) poles of $f(\lambda)$ in the region

  \[ \Omega_T^+ = \{ a + bi : a > 0, 0 < b < aT \} \]  
  (2.127)

  with residue $f_c$, respectively. Further, we suppose the complex poles in $\Omega_T^+$ have a given bound $\lambda_M$, i.e. $\lambda_M \geq \max_{1 \leq c \leq n_c} |\lambda_c|$.

- $|\tilde{f}(\lambda)| \leq C \left(1 + |\lambda|^K\right)$ for any $\lambda \in \Omega_T^+ \cup \mathbb{R}^+$ satisfying $\max_{0 \leq \ell \leq L} |\lambda - k_\ell| \geq \epsilon_0$, here $C > 0$ and $K \in \mathbb{N} \cup \{0\}$ are given integer constants.

- For $k' = 4k_M + 2\lambda_M + 2\epsilon_0$, $k_M = \max\{k_1, \cdots, k_L\}$, $S = \int_{-k'}^{k'} |\tilde{f}(\lambda)| \, d\lambda < +\infty$.

Lemma 2.23. Let $\mu > 1$, $T > 0$, $\epsilon_0 > 0$ be some given constants, and the function $f(\lambda)$ satisfies Assumption 2.22, and $\tilde{f}(\lambda)$ is so defined with the real poles removed from $f(\lambda)$, $(\rho, \theta)$ and $(\rho', \theta')$ are the polar coordinates of $(x, y)$ and $(x', y')$, respectively, and $\rho > \mu \rho' \geq 0$. Suppose $y > 0$, $y + y' > 0$ and $|x| < Ty$. Then, the integral Bessel-type expansion (2.96) holds on the interval $(k', \infty)$ (by replacing the original $f(\lambda)$) with $\tilde{f}(\lambda)$, with a truncation error estimate for a finite $P$-term truncation

\[ \left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\theta} \int_{k'}^\infty \Psi(\lambda) (-iw(\lambda))^p \tilde{f}(\lambda) \, d\lambda \right| \leq c_+(P, \rho) \left( \frac{\rho'}{\rho} \right)^P, \]  
  (2.128)

for any sufficiently large $P \geq m_+(\rho)$. Here, $m_+(\rho)$ is an (at most) quadratic function of $\rho$, and $c_+(P, \rho)$ is a function having polynomial growth rate in $P$. 

48
Proof. If \( x = 0 \), then \( y = \sqrt{x^2 + y^2} = \rho \), and

\[
|\bar{f}(\lambda)| \leq \sum_{c=1}^{n_c} \frac{|f_c|}{3\lambda_c} + C \left( 1 + |\lambda|^K \right) \leq C_1 \left( 1 + |\lambda|^K \right)
\]

for \( \lambda \in (k', \infty) \), here \( C_1 > 0 \) is a constant only depending on \( f(\lambda) \). By Lemma 2.17, we can choose

\[
m(\rho) = \left( \frac{k\rho}{2} \right)^2 + 1 - K, \quad c_+(P, \rho) = 6C_1(1 + (\frac{2\mu}{\rho(\mu - 1)})^{K+1}(P + K)^K.
\]

If \( x \neq 0 \), without a loss of generality, we assume \( x > 0 \), since the case \( x < 0 \) will follow by taking complex conjugates. Let \( \beta = \frac{\pi}{2} - \theta \), then \( \tan \beta \in (0, T) \). Let \( \kappa \) be the segment from \( \phi(k') \) to \( k' \) (see Figure 2.4), here \( \phi \) is the Cagniard–de Hoop mapping defined in (2.116). One can verify the length of \( \kappa \) is bounded by \( \sqrt{2}k' \), and that \( \lambda_M + k_M \leq |\lambda| \leq \sqrt{2}k' \) and \( |\lambda - k_l| > \epsilon_0 \) for \( \lambda \in \kappa, 0 \leq l \leq L \). Define

\[
E = \int_{\kappa \cup (k', \infty)} \Psi(\lambda) \Psi'(\lambda) \bar{f}(\lambda) d\lambda, \quad G = \int_{\kappa} \Psi(\lambda) \Psi'(\lambda) \bar{f}(\lambda) d\lambda,
\]

(2.129)

we will discuss the expansions for \( E \) and \( G \), separately, then give the integral Bessel-type expansion for \( E - G \). On \( \kappa \) we have the bound of \( \bar{f}(\lambda) \) given by

\[
|\bar{f}(\lambda)| \leq C \left( 1 + |\lambda|^K \right) + \sum_{c=1}^{n_c} \frac{|f_c|}{|\lambda - \lambda_c|} \leq C \left( 1 + (\sqrt{2}k')^K \right) + \sum_{c=1}^{n_c} \frac{|f_c|}{k_M} := C_2. \tag{2.130}
\]

Thus by Lemma 2.15, the Bessel-type expansion for \( G \) is given by

\[
G = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\theta'} G_p, \quad G_p = \int_{\kappa} \Psi(\lambda) (-i w(\lambda))^p \bar{f}(\lambda) d\lambda, \tag{2.131}
\]
with a truncation error

$$\left| \sum_{|p| \geq P} J_p(k\rho')e^{ip\theta'} G_p \right| \leq c_\kappa \left( \frac{\rho'}{\rho} \right)^P \quad \text{for } P \geq m_\kappa(\rho),$$

(2.132)

here $c_\kappa = 2\mu C_2 \cdot \sqrt{2k'/(\mu - 1)}$, $m_\kappa(\rho) = e(\lambda_M + k/2) \rho$. For the contour $\kappa \cup (k', \infty)$, with the substitution $\lambda = \phi(\lambda') = \lambda' \cos \beta + i\sqrt{\lambda^2 - k^2} \sin \beta$ we have

$$\sqrt{\lambda^2 - k^2} = \frac{\lambda \cos \beta - \phi^{-1}(\lambda)}{i \sin \beta} = \frac{\phi(\lambda') \cos \beta - \lambda'}{i \sin \beta} = \sqrt{\lambda'^2 - k^2} \cos \beta + i\lambda' \sin \beta,$$

so

$$\tilde{\Psi}(\lambda) = e^{-\sqrt{\lambda'^2 - k^2} y + i\lambda x} = e^{-(\sqrt{\lambda'^2 - k^2} \cos \beta + i\lambda' \sin \beta)(\rho \cos \beta + i(\lambda' \cos \beta + i\sqrt{\lambda'^2 - k^2} \sin \beta)(\rho \sin \beta)} = e^{-\sqrt{\lambda'^2 - k^2} \rho}.$$

Similarly, $\tilde{\Psi}'(\lambda) = e^{-\sqrt{\lambda'^2 - k'^2} \rho' \sin(\theta' - \beta) - i\lambda' \rho' \cos(\theta' - \beta)}$ and $w(\lambda) = e^{-i\beta} w(\lambda')$. Hence

$$E = \int_{\phi^{-1}(\kappa) \cup \gamma'} e^{-\sqrt{\lambda'^2 - k^2} \rho' \sin(\theta' - \beta) - i\lambda' \rho' \cos(\theta' - \beta)} \tilde{f}(\lambda') d\lambda'$$

(2.133)

where $\gamma' = \phi^{-1}((k', \infty))$ is the lower part of $\gamma$ starting from $\phi^{-1}(k')$ located somewhere on $\gamma^-$, and

$$\tilde{f}(\lambda') = \tilde{f}(\lambda) \frac{d\lambda}{d\lambda'} = \tilde{f}(\phi(\lambda')) \frac{\sqrt{\phi(\lambda')^2 - k^2}}{\sqrt{\lambda'^2 - k^2}}.$$

(2.134)

Since $\phi(\lambda')$ has a polynomial bound, roughly,

$$|\phi(\lambda')| = \left| \lambda' \cos \beta + i\sqrt{\lambda'^2 - k^2} \sin \beta \right| \leq |\lambda'| + \sqrt{|\lambda'|^2 + k^2} \leq 2|\lambda'| + k$$

(2.135)

when $\lambda' \in D^-$ and $|\lambda'|$ is sufficiently large, $\tilde{f}(\lambda')$ also has a polynomial bound of $|\lambda'|$. 50
Next, we proceed to change the contour of the integral $E$ from $\phi^{-1}(\kappa) \cup \gamma'$ back to $(k', \infty)$. Let $\zeta$ be the counterclockwise arc with radius $r$ connecting $\phi^{-1}(\kappa) \cup \gamma'$ and the real axis, parameterized by $\lambda' = re^{i\eta}$, where the range of $\eta$ is a subset of $(-\beta, 0)$. On the arc $\zeta: \lambda' = re^{i\eta}$, as $r \to \infty$, the exponent of the integrand in $E$ satisfies

$$-\sqrt{\lambda'^2 - k^2 (\rho + \rho' \sin(\theta' - \beta))} - i\lambda' \rho' \cos(\theta' - \beta) \sim -\lambda' e^{i\beta} \left( pe^{-i\beta} + i\rho' e^{-i\theta'} \right)$$

$$\sim r \bar{\rho} \exp \left( i \left( \eta + \bar{\theta} + \beta + \frac{\pi}{2} \right) \right)$$

where $(\bar{\rho}, \bar{\theta})$ are the polar coordinates of $(x - x', y + y')$, and the rest of the integrand has a polynomial bound. Since $y + y' > 0$, $\rho > \rho'$, one can verify $\bar{\theta} \in (0, \pi - \beta)$. Then

$$\Re \left\{ r \bar{\rho} \exp \left( i \left( \eta + \bar{\theta} + \beta + \frac{\pi}{2} \right) \right) \right\} \leq r \cdot \max \left\{ -(y + y'), -\rho - \rho' \sin(\theta' - \beta) \right\}$$

for any $\eta \in (-\beta, 0)$, so the integrand on $\zeta$ decays exponentially, and the corresponding integral on $\zeta$ vanishes as $r \to +\infty$. Also notice that there are no poles of $\tilde{f}(\lambda')$ in $D' \subset D^-$, where $D'$ is the region enveloped by $(k', +\infty)$ and $\phi^{-1}(\kappa) \cup \gamma'$, because $\phi$ is a holomorphic function on $D'$ which maps any possible pole in $D'$ to a pole of $f$ in $\phi(D')$, however, for any $\lambda' \in D'$ and any pole $\lambda_c \in \Omega_T^+ \cap \phi(\lambda')$, $|\phi(\lambda')| \geq \lambda_M + k_M > |\lambda_c|$. Hence, by deforming the integration contour in $E$ to the real axis, we have

$$E = E' := \int_{(k', \infty)} e^{-\sqrt{\lambda'^2 - k^2 (\rho + \rho' \sin(\theta' - \beta)) - i\lambda' \rho' \cos(\theta' - \beta)}} \tilde{f}(\lambda') d\lambda'.$$

(2.136)

Now, for $\lambda' \in (k', \infty)$, recall that

$$\tilde{f}(\lambda) = \frac{\sqrt{\lambda^2 - k^2}}{\sqrt{\lambda'^2 - k^2}} \left( \tilde{f}(\phi(\lambda')) + \sum_{c=1}^{n_c} \frac{f_c}{\phi(\lambda') - \lambda_c} \right),$$

for each $\lambda_c$ we have $|\phi(\lambda') - \lambda_c| \geq |\phi(\lambda')| - |\lambda_c| \geq \sqrt{\lambda'^2 - k^2 \sin^2 \beta} - \lambda_M \geq 3k_M$, so using
there exists some constant $C_2 > 0$ such that
\[
|\tilde{f}(\lambda')| \leq \frac{2|\lambda'| + 2k}{3kM} \left( C \left( 1 + (2|\lambda'| + k)^K \right) + \sum_{e=1}^{n_e} \frac{|f_e|}{3kM} \right) \leq C_2 |\lambda'|^{K+1}.
\]
Hence by Lemma 2.17, $E'$ has a series expansion
\[
E' = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip(\theta'-\beta)} \int_{\lambda'}^{\infty} e^{-\sqrt{\lambda'^2-k^2}\rho} (-iw(\lambda'))^p \tilde{f}(\lambda') d\lambda'
\] (2.137)
with a $P$-term truncation error estimate
\[
\left| \sum_{|p| \geq P} J_p(k\rho') e^{ip(\theta'-\beta)} \int_{\lambda'}^{\infty} e^{-\sqrt{\lambda'^2-k^2}\rho} (-iw(\lambda'))^p \tilde{f}(\lambda') d\lambda' \right| \leq c_{E'}(P, \rho) \left( \frac{\rho'}{\rho} \right)^P
\] (2.138)
for $P \geq m_{E'}(\rho) = (k\rho)^2/4 - K$. Here
\[
c_{E'}(P, \rho) = 6C_2(K + 2)! \left( \frac{2\mu}{\rho(\mu - 1)} \right)^{K+2} (P + K + 1)^{K+1}.
\] (2.139)
In the series (2.137), the $p$-th term is
\[
e^{-ip\beta} \int_{(k,\infty)} e^{-\sqrt{\lambda'^2-k^2}\rho} (-iw(\lambda'))^p \tilde{f}(\lambda') d\lambda'
\] (2.140)
\[
e^{-ip\beta} \int_{\phi^{-1}(\kappa)\cup \lambda'} e^{-\sqrt{\lambda'^2-k^2}\rho} (-iw(\lambda'))^p \tilde{f}(\lambda') d\lambda'
\] (2.141)
In the above equation, the first equality is obtained by changing the contour, and on the path $\zeta : \lambda' = re^{i\eta}$ the integrand decays exponentially as $r \to \infty$ as the real part of the exponent
\[
\Re \left( -\sqrt{(re^{i\eta})^2 - k^2}\rho \right) \sim \Re(-re^{i\eta}\rho) \leq -ry,
\]
while the remaining parts have polynomial growth rate. The second equality is by the
substitution from $\lambda'$ to $\lambda$. In total we have proven the series expansion of $E$ given by
$$E = \sum_{p=-\infty}^{\infty} J_p(k\rho')e^{ip\rho'} E_p$$
with a $P$-term truncation error estimate
$$|E - \sum_{|p|<P} J_p(k\rho')e^{ip\rho'} E_p| \leq c_E'(P, \rho) \left( \frac{\rho'}{\rho} \right)^P \text{ for } P \geq m_E'(\rho). \quad (2.142)$$

For each $p$,
$$E_p - G_p = \int_{\kappa \cup (k', \infty)} \Psi(\lambda) (-iw(\lambda))^p \bar{f}(\lambda)d\lambda - \int_{\kappa} \Psi(\lambda) (-iw(\lambda))^p \bar{f}(\lambda)d\lambda$$
$$= \int_{k'}^{\infty} \Psi(\lambda) (-iw(\lambda))^p \bar{f}(\lambda)d\lambda, \quad (2.143)$$

which is the desired expansion function in the Bessel-type expansion (2.96).

Finally, by combining the results (2.132) and (2.142), $\forall P \geq \max\{m_E(\rho), m_\kappa(\rho)\}$,

$$\left| \int_{k'}^{\infty} \Psi(\lambda)\Psi'(\lambda) \bar{f}(\lambda)d\lambda - \sum_{|p|<P} J_p(k\rho')e^{ip\rho'} \int_{k'}^{\infty} \Psi(\lambda) (-iw(\lambda))^p \bar{f}(\lambda)d\lambda \right|$$
$$\leq \left| E - \sum_{|p|<P} J_p(k\rho')e^{ip\rho'} E_p \right| + \left| G - \sum_{|p|<P} J_p(k\rho')e^{ip\rho'} G_p \right| \leq (c_E'(P, \rho) + c_\kappa) \left( \frac{\rho'}{\rho} \right)^P$$

which suggests $c_+(P, \rho) = c_E'(P, \rho) + c_\kappa$ and $m_+(\rho) = \max\{m_E(\rho), m_\kappa(\rho)\}$.

**Theorem 2.24 (the Bessel-type expansion).** Suppose conditions of Lemma 2.23 are satisfied. Further suppose $0 < \rho_m < \rho_M$ are given such that $\rho \in [\rho_m, \rho_M]$. Then, the integral Bessel-type expansion (2.96) holds with a truncation error estimate
$$\left| \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2-k^2(y+y')+i\lambda(x-x')}} f(\lambda)d\lambda - \sum_{|p|<P} J_p(k\rho')e^{ip\rho'} F_p \right| \leq c(P) \left( \frac{\rho'}{\rho} \right)^P \quad (2.144)$$

for some function $c(\cdot)$ with polynomial growth rate when $P$ is sufficiently large, i.e., $P \geq m(\rho_M)$, $m(\rho_M)$ is an at most quadratic function.
Proof. Consider the decomposition of the integral

\[ I = \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2} (y + y') + i\lambda(x - x')} f(\lambda) d\lambda \]

\[ = \sum_{r=1}^{n_r} \tau_r i\pi \Psi(\lambda_r) \Psi'(\lambda_r) f_r + \left( \int_{-\infty}^{-k'} + \int_{-k'}^{k'} + \int_{k'}^{\infty} \right) \Psi(\lambda) \Psi'(\lambda) \bar{f}(\lambda) d\lambda \]  

(2.145)

where each \( \tau_r = \pm 1 \) are determined by the well-posed physical problem, see (2.25). Each term \( I_j \) of the decomposition with index \( j \) has the corresponding Bessel-type expansion, \( j = 0, 1, \cdots, n_r, +, - \). Namely, for each \( I_r \), by Lemma 2.14, by choosing \( c_r = 2\pi \mu |f_r|/(\mu - 1) \) and \( m_r(\rho) = e(|\lambda_r| + k/2)\rho \), the pointwise Bessel-type expansion (2.95) holds

\[ I_r = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\theta'} I_{r,p}, \quad I_{r,p} = \tau_r i\pi \Psi(\lambda_r) (-iw(\lambda_r))^p \]

with the truncation error for a \( P \)-term truncation

\[ \left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\theta'} I_{r,p} \right| \leq c_r \left( \frac{\rho'}{\rho} \right)^P \text{ for } P \geq m_r(\rho). \]  

(2.146)

For \( I_0 \), by Lemma 2.16, by choosing \( c_0 = 2\pi \mu S/(\mu - 1) \) and \( m_0(\rho) = ek'\rho \), the integral Bessel-type expansion (2.96) holds

\[ I_0 = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\theta'} I_{0,p}, \quad I_{0,p} = \int_{-k'}^{k'} \Psi(\lambda) (-iw(\lambda))^p \bar{f}(\lambda) d\lambda \]

with the truncation error for a \( P \)-term truncation

\[ \left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\theta'} I_{0,p} \right| \leq c_0 \left( \frac{\rho'}{\rho} \right)^P \text{ for } P \geq m_0(\rho). \]  

(2.147)

For \( I_+ \) and \( I_- \), by choosing the \( c_+(P, \rho) \) and \( m_+(\rho) \) provided by Lemma 2.23, and \( c_-(P, \rho) = \frac{54}{\rho} \).
c_+(P, \rho) \text{ and } m_-(\rho) = m_+(\rho) \text{ due to the symmetry, the integral Bessel-type expansion (2.96) holds as } I_\pm = \sum_{p=-\infty}^{\infty} J_p(k\rho') e^{ip\rho'} I_{\pm,p}, \text{ where }

I_{+p} = \int_{k'}^{\infty} \Psi(\lambda) (-i w(\lambda))^p \bar{f}(\lambda) d\lambda, \quad I_{-p} = \int_{-\infty}^{-k'} \Psi(\lambda) (-i w(\lambda))^p \bar{f}(\lambda) d\lambda

with the truncation error for a \( P \)-term truncation

\[
\left| \sum_{|p| \geq P} J_p(k\rho') e^{ip\rho'} I_{\pm,p} \right| \leq c_\pm(P, \rho) \left( \frac{\rho'}{\rho} \right)^P \text{ for } P \geq m_\pm(\rho). \tag{2.148}
\]

For each \( p \), the expansion functions add up to \( F_p \) because

\[
F_p = \sum_{r=1}^{n_r} \tau_r i \pi \Psi(\lambda_r) (-i w(\lambda_r))^p f_r + \left( \int_{-\infty}^{-k'} + \int_{-k'}^{k'} + \int_{k'}^{\infty} \right) \Psi(\lambda) (-i w(\lambda))^p \bar{f}(\lambda) d\lambda
\]

\[
= \sum_{r=1}^{n_r} I_{r,p} + I_{-p} + I_{0,p} + I_{+p}.
\]

Hence by adding the series expansions up, for any \( P \geq m(\rho) := \max_j m_j(\rho) \),

\[
\left| I - \sum_{|p| < P} J_p(k\rho') e^{ip\rho'} F_p \right| \leq c(P, \rho) \left( \frac{\rho'}{\rho} \right)^P,
\]

where \( c(P, \rho) := \sum_{r=1}^{n_r} c_r + c_0 + c_+(P, \rho) + c_-(P, \rho) \). Since the only dependence of \( c(P, \rho) \) on \( \rho \) appears in the terms \( c_\pm(P, \rho) \) which reach their upper bounds at \( \rho = \rho_m \), and each \( m_j(\cdot) \) is an increasing function, we conclude that by choosing \( c(P) := c(P, \rho_m) \), the truncation error estimate (2.144) holds for any \( P \geq m(\rho_M) \).

\[\square\]

2.5.4. Proof of Theorem 2.12

Here, only the proof of the ME (2.81) will be given since the others can be similarly treated.
Let \( \tilde{x} = x - x_c, \tilde{y} = \tau^*(y - d^*_\ell) + \tau^*(y_c - d^*_\ell), \tilde{x}' = x' - x_c, \tilde{y}' = \tau^*(y' - y_c), \) and

\[
f(\lambda) = e^{(\sqrt{\lambda^2 - k^2_\ell} - \sqrt{\lambda^2 - k^2_\ell^*})\tau^*(y-d^*_\ell)} \sigma^{**}_{ell}(\lambda)
\] (2.149)

so that the integral (2.21) can be written as

\[
u^{**}_{ell}(r; r') = \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2_\ell^*}(y-d^*_\ell)} - \sqrt{\lambda^2 - k^2_\ell^*} \tau^*(y'-d^*_\ell^*) + i\lambda(x-x') \sigma^{**}_{ell}(\lambda) d\lambda
\]

\[= \int_{-\infty}^{\infty} e^{-\sqrt{\lambda^2 - k^2_\ell^*}(\tilde{y}+\tilde{y}') + i\lambda(\tilde{x}-\tilde{x}')} f(\lambda) d\lambda.
\]

With the assumption that the sources, the targets and the centers are bounded in a given box, and that \( |y_c-d^*_\ell| \) has a nonzero lower bound, there exists fixed \( T > 0 \) such that \( |\tilde{x}| < T\tilde{y} \). By Theorem 2.9, \( \sigma^{**}_{ell}(\lambda) \) has a polynomial bound in the region \( \Omega_T = \{ a + bi : a > 0, -aT < b < aT \} \) when \( \Re \lambda \) is sufficiently large, and has a finite number of poles in \( \Omega_T \), which easily imply the same for \( f(\lambda) \). With the decomposition (2.126), when neighborhoods of each branch point \( k_\ell \) with a sufficiently small radius \( \epsilon_0 > 0 \) are excluded from \( \Omega_T \), \( \tilde{f}(\lambda) \) is finite and hence has polynomial bound. Replacing \( x, y, x', y', k \) in Theorem 2.24 by \( \tilde{x}, \tilde{y}, \tilde{x}', \tilde{y}', k_\ell \) finishes the proof of (2.81).

For the LE (2.82), similarly, choose \( \tilde{x} = x^l_c - x', \tilde{y} = \tau^*(y^l_c - d^*_\ell) + \tau^*(y' - d^*_\ell), \tilde{x}' = x^l_c - x, \)
\( \tilde{y}' = \tau^*(y - y^l_c), \) \( k = k_\ell \) and \( f(\lambda) = e^{(\sqrt{\lambda^2 - k^2_\ell} - \sqrt{\lambda^2 - k^2_\ell^*})\tau^*(y'-d^*_\ell^*)} \sigma^{**}_{ell}(\lambda).
\]

For the M2L (2.83), for each LE coefficient \( L^*_{m}(r^l_c, r) \), choose \( \tilde{x} = x^l_c - x_c, \tilde{y} = \tau^*(y_c - d^*_\ell) + \tau^*(y_c - d^*_\ell) \), \( \tilde{x}' = x^l_c - x, \tilde{y}' = \tau^*(y' - y_c), \) \( k = k_\ell \) and

\[f(\lambda) = e^{(\sqrt{\lambda^2 - k^2_\ell} - \sqrt{\lambda^2 - k^2_\ell^*})\tau^*(y'-d^*_\ell^*)} \sigma^{**}_{ell}(\lambda) (i\omega_\ell(\lambda)^{-1})^m.\]

For the L2L (2.84), for each LE coefficient \( L^*_{m}(r_c, r) \), choose \( \tilde{x} = x^l_c - x', \tilde{y} = \tau^*(y_c - d^*_\ell) + \tau^*(y' - d^*_\ell) \), \( \tilde{x}' = x^l_c - x_c, \tilde{y}' = \tau^*(y^l_c - y^l_c), \) \( k = k_\ell \) and

\[f(\lambda) = e^{(\sqrt{\lambda^2 - k^2_\ell} - \sqrt{\lambda^2 - k^2_\ell^*})\tau^*(y'-d^*_\ell^*)} \sigma^{**}_{ell}(\lambda) (i\omega_\ell(\lambda)^{-1})^m.\]
Remark 2.25. (Dependence of convergence estimate on the number of layer interfaces $L$) As pointed out in Theorem 2.9, each reflection/transmission coefficient $\sigma_{i\ell}^{*}(\lambda)$ is asymptotically sublinear as $\lambda \to \infty$ regardless of $L$ (the number of the interfaces). When applying Theorem 2.24 and Theorem 2.23, the bounds of $\sigma_{i\ell}^{*}(\lambda)$ are assumed of the same polynomial order as $\lambda \to \infty$. Therefore, as $L$ increases, the required terms for truncation, namely, the $m(\rho_M)$ in the proof of Theorem 2.24 has linear dependence on the distribution of the poles as shown by Lemma 2.14, while the leading term $(k\rho_M/2)^2$ remains unchanged.

2.6. Conclusion

Far-field expansions of ME, LE as well as M2L and L2L translation operators are derived for the LMGF of the 2-D Helmholtz equation, and the exponential convergence rates are proven. The analysis shows that the convergence of ME and LE for the reaction field components depends on the distance between the target and the equivalent polarization source. This fact shows how the ME and LE for the layered media can be used in the traditional FMM framework, and such an approach has been implemented for the succeeding works in Chapters 3–4, see also [76, 78, 79].
Chapter 3

FMM for the Helmholtz equation and the linearized Poisson–Boltzmann equation in 3-D layered media

The content in this chapter has been published in the following journal papers in collaboration with Bo Wang and Wei Cai:


The authors thank Min Hyung Cho and Johannes Tausch for helpful discussions.

3.1. Introduction

In the free space, the FMM for Helmholtz equation solving wave interaction problems has been proposed in [14], with a few remarkable succeeding work such as in [13] for high frequency wave propagation, where the pairwise particle interactions are expanded using Graf’s addition theorems with spherical Hankel functions and spherical Bessel functions. As a corollary, by using the modified spherical Hankel functions and the modified spherical Bessel functions, the Green’s function in the free space, also referred as the Yukawa potential or screened Coulomb potential (cf. [45, 50, 51, 8]), has been developed in [36, 43].

For problems in layered media, a number of numerical methods using the LMGF of the Helmholtz equation for wave interaction has been mentioned in Section 1.2. It is worth reiterating the inadequacies of these approaches compared to the method to be elaborated
in this section, for the purpose of developing a hierarchical numerical method: the insufficient use of the layer structure and the LMGF, and of the far-field relationship in general layered media.

In this chapter, we study the LMGF of the Helmholtz equation and the linearized Poisson–Boltzmann (PB) equation and develop the FMM algorithms for these differential equations with numerical validation. As guided by the numerical analysis work of the FMM for 2-D Helmholtz equation in Chapter 2, we will begin with general theories of the LMGF in Section 3.2, then propose the far-field expansions for the LMGF in Section 3.3, and the FMM framework in Section 3.4. The implementation details with a few optimization are then discussed in Section 3.5. Numerical tests in Section 3.6 verifies the efficiency and the accuracy of the developed FMM in layered media.

3.2. The Green’s function in 3-D layered media

We consider the differential equations presented in the united form for both the Helmholtz equation and the linearized PB equation

$$\nabla^2 u(r; r') + k^2 u(r; r') = -\delta(r - r'),$$

(3.1)

where the partial derivatives are taken on the target coordinates $r = (x, y, z)$, $\delta$ is the Dirac delta function, and $k$ is a piecewisely constant number, with the interface conditions

$$[au] = 0, \quad [b \frac{\partial u}{\partial n}] = 0, \quad z = d_0, \cdots, d_{L-1},$$

(3.2)

where $a, b > 0$ are piecewisely constant parameters for each layer. Certain additional conditions are applied to the LMGF:

1. For the Helmholtz equation, $k \in \mathbb{R}^+$ refers to the wave number in each layer, and the Green’s function is restricted by the upward/downward outgoing radiation conditions [11] of field propagation. An example of the interface conditions for acoustic wave...
propagation is given by $a = 1, b = \rho^{-1}$, where $\rho$ is the density of the medium, and $u(r; r')$ refers to the pressure. Another example arising from the Maxwell’s equations to be discussed in Chapter 5 adopts $a = 1, b = \mu^{-1}$ and $a = 1, b = \varepsilon^{-1}$, where $\mu$ and $\varepsilon$ are the piecewisely constant relative permittivity and the relative permeability of the layered media, respectively.

2. For the linearized PB equation, $k = i\lambda$, where $\lambda \in \mathbb{R}^+$ refers to the inverse Debye–Hückle screen length measuring the ionic length in solvents. The Green’s function is restricted by a decaying condition

$$u(r; r') \to 0 \quad \text{as} \quad r = |r| \to \infty.$$ \hspace{1cm} (3.3)

In continuum electrostatics, $u(r; r')$ refers to the electrostatic potential, and the interface conditions are given by $a = 1, b = \varepsilon$, where $\varepsilon$ is the dielectric constant.

**Remark 3.1.** Other works related with electrostatics often have the LMGF defined differing by a factor from the one in (3.1), such as in [51, 69], where the dielectric constant $\varepsilon_{\ell'}$ is divided by the Dirac delta function in the right-hand side. They can be equivalently treated both theoretically and practically since $\varepsilon_{\ell'}$ is piecewisely constant.

3.2.1. Reaction field decomposition of the LMGF in the frequency domain

With either of the Sommerfeld radiation condition [71] or (3.3), the free-space Green’s function satisfying the differential equation (3.1) is given by

$$G^f(r; r') = \frac{e^{ik_{\ell'}|r-r'|}}{4\pi|r-r'|}. \hspace{1cm} (3.4)$$

Like in the Definition 2.1 of the reaction field for the 2-D case, we define the reaction field in 3-D as follows.
Definition 3.2 (Reaction field). The reaction field of the LMGF is defined by

\[ u^r(r; r') = u(r; r') - \delta_{\ell,\ell'} G^f(r; r'), \tag{3.5} \]

where \( G^f(r; r') \) was defined in (3.4).

To derive the reaction field decomposition like in the 2-D case (2.20), we take the 2-D Fourier transform from \((x - x', y - y')\) to \((k_x, k_y)\)

\[ f(r) = \frac{1}{4\pi^2} \iint_{\mathbb{R}^2} e^{ik_x(x-x') + ik_y(y-y')} \hat{f}(k_x, k_y, z) dk_x dk_y. \tag{3.6} \]

For the rest of this chapter (and this thesis), we use \((k_\rho, \alpha)\) as the polar coordinate pair of \((k_x, k_y)\), i.e.,

\[ k_x = k_\rho \cos \alpha, \quad k_y = k_\rho \sin \alpha. \tag{3.7} \]

In the frequency domain of the Fourier transform, the reaction field satisfies the ordinary differential equation

\[ \frac{\partial^2 \hat{u}^r(k_x, k_y, z; r')}{\partial z^2} + (k^2 - k_x^2 - k_y^2) \hat{u}^r(k_x, k_y, z; r') = 0, \tag{3.8} \]

while the interface conditions are shown as

\[ \begin{bmatrix} a(\hat{u}^r + \hat{G}^f) \end{bmatrix} = 0, \quad \begin{bmatrix} b \left( \frac{\partial \hat{u}^r}{\partial n} + \frac{\partial \hat{G}^f}{\partial n} \right) \end{bmatrix} = 0, \quad z = d_0, \cdots, d_{L-1}. \tag{3.9} \]

The Fourier transform of \( \hat{G}^f(k_x, k_y, z; r') \) is given by

\[ \hat{G}^f(k_x, k_y, z; r') = -\frac{1}{2} \frac{e^{ik_\rho |z-z'|}}{ik_\rho z}, \tag{3.10} \]
where for $\ell = 0, \cdots, L$, $k_{\ell z}$ is defined by

$$k_{\ell z} = \sqrt{k^2_{\ell} - k^2_p}. \quad (3.11)$$

The branch cut of the square root is taken to ensure a nonnegative imaginary part, i.e. for any $z = re^{i\theta}$ with $z \geq 0, \theta \in [0, 2\pi)$,

$$\sqrt{z} = \sqrt{r} e^{i\theta/2}. \quad (3.12)$$

In particular, for the linearized PB equation, $k_{\ell z} = i\sqrt{\lambda^2 + k^2}$.

Most properties discussed in Section 2.2 for the LMGF of the 2-D Helmholtz equation also hold for the 3-D reaction field $\hat{u}$. Here we state the corresponding result for the reaction field decomposition, without reiterating the derivation and the proof.

**Definition 3.3 (Reaction field decomposition).** The reaction field decomposition of $\hat{u}$ in the frequency domain is defined by

$$u^r(r; r') = \frac{1}{4\pi^2} \int \int e^{ik_x(x-x') + ik_y(y-y')} \sum_{*,* \in \{\uparrow, \downarrow\}} \sigma_{\ell \ell'}^{*\ast}(k_x, k_y; x', y') e^{\tau^* i k_{\ell z}(z - d^*_{\ell})} e^{\tau^{*\ast} i k_{\ell z'}(z' - d^*_{\ell'})} dk_x dk_y \quad (3.13)$$

where

$$\tau^\uparrow = +1, \quad \tau^\downarrow = -1, \quad (3.14)$$

$d^*_{\ell}$ is the same as in (2.16),

$$d^\uparrow_{\ell} = d_{\ell}, \quad 0 \leq \ell \leq L - 1, \quad d^\downarrow_{L} = -\infty, \quad (3.15)$$

$$d^\uparrow_{\ell} = d_{\ell-1}, \quad 1 \leq \ell \leq L, \quad d^\downarrow_{0} = +\infty. \quad (3.16)$$


Each reaction field component

\[ u_{\ell \ell}^*(r; r') = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell \ell}^*(r, r'; k, k') \sigma_{\ell \ell}^*(k_\rho; x', y') dk_x dk_y, \]  

(3.17)

where \( \sigma_{\ell \ell}^*(k_\rho; x', y') \) are the reflection/transmission coefficients of the layered media in the frequency domain, and

\[ \mathcal{E}_{\ell \ell}^*(r, r'; k, k') = e^{ik_\rho(x-x') + ik_\rho(y-y') + \tau^* k_\rho z (z-d_\rho) + \tau k_\rho z (z-d_\rho')}. \]  

(3.18)

The reflection/transmission coefficients \( \sigma_{\ell \ell}^*(k_\rho; x', y') \) in fact are independent from \( x' \) and \( y' \), and are rotationally symmetric on the \((k_x, k_y)\)-plane, which are direct corollaries of the recursive formulas in the coming section, where the linear systems whose unknowns consist of \( \sigma_{\ell \ell}^*(k_\rho; x', y') \) do not depend on \( x' \) and \( y' \), and are rotationally invariant on the \((k_x, k_y)\)-plane. Thus, we can simply denote

\[ \sigma_{\ell \ell}^*(k_\rho; x', y') = \sigma(k_\rho) \]  

(3.19)

in the above reaction field decomposition.

Again,

\[ \sigma_{00}^* = \sigma_{L0}^* = \sigma_{L0}^* = \sigma_{L0}^* = 0, \quad u_{00}^* = u_{L0}^* = u_{L0}^* = u_{L0}^* = 0, \quad *, * \in \{\uparrow, \downarrow\}. \]  

(3.20)

Poles of the reflection/transmission coefficients are understood in a similar way as in Remark 2.3.

3.2.2. Recursive formula for the reflection/transmission coefficients

Here we introduce a recursive solver for calculating the reflection/transmission coefficients. In the derivation, we fix the target \( r' \) together with the target layer index \( \ell' \). Define
\[ \vartheta = -\frac{1}{2 i k_{PZ}} \]  

(3.21)

\[ e_\ell = e^{i k_{PZ}(d_{\ell-1} - d_{\ell})}, \quad 1 \leq \ell \leq L - 1, \quad e_0 = e_L = 1, \]  

(3.22)

\[ \gamma^\pm_\ell = \frac{a_\ell}{a_{\ell-1}} \pm \frac{b_\ell k_{PZ}}{b_{\ell-1} k_{PZ-1}}, \quad 1 \leq \ell \leq L, \]  

(3.23)

\[ \sigma^*_\ell = \sigma_{\ell\ell'}^* e^{i k_{PZ}(z' - d_{\ell'})} + \sigma_{\ell\ell'}^* e^{i k_{PZ}(d_{\ell'-1} - z')}, \quad 0 \leq \ell \leq L, \]  

(3.24)

where \( d_{-1} = +\infty \) and \( d_L = -\infty \). The interface equations (3.9) can be expanded like in (2.26), but in 3 different cases:

1. If \( \ell \neq \ell' \) and \( \ell \neq \ell' - 1 \), at \( z = d_{\ell} \),

\[ a_{\ell-1} \left( \sigma^\dagger_{\ell-1} + e_{\ell-1} \sigma^\dagger_{\ell-1} \right) = a_\ell \left( e_\ell \sigma^\dagger_\ell + \sigma^\dagger_\ell \right), \]

\[ b_{\ell-1} k_{PZ-1} \left( \sigma^\dagger_{\ell-1} - e_{\ell-1} \sigma^\dagger_{\ell-1} \right) = b_\ell k_{PZ} \left( e_\ell \sigma^\dagger_\ell - \sigma^\dagger_\ell \right); \]  

(3.25)

2. If \( \ell = \ell' + 1 \), at \( z = d_{\ell} \),

\[ a_{\ell-1} \left( \sigma^\dagger_{\ell-1} + e_{\ell-1} \sigma^\dagger_{\ell-1} + \vartheta e^{i k_{PZ}(z' - d_{\ell'})} \right) = a_\ell \left( e_\ell \sigma^\dagger_\ell + \sigma^\dagger_\ell \right), \]

\[ b_{\ell-1} k_{PZ-1} \left( \sigma^\dagger_{\ell-1} - e_{\ell-1} \sigma^\dagger_{\ell-1} - \vartheta e^{i k_{PZ}(z' - d_{\ell'})} \right) = b_\ell k_{PZ} \left( e_\ell \sigma^\dagger_\ell - \sigma^\dagger_\ell \right); \]  

(3.26)

3. If \( \ell = \ell' \), at \( z = d_{\ell} \),

\[ a_{\ell-1} \left( \sigma^\dagger_{\ell-1} + e_{\ell-1} \sigma^\dagger_{\ell-1} \right) = a_\ell \left( e_\ell \sigma^\dagger_\ell + \sigma^\dagger_\ell + \vartheta e^{i k_{PZ}(d_{\ell'-1} - z')} \right), \]

\[ b_{\ell-1} k_{PZ-1} \left( \sigma^\dagger_{\ell-1} - e_{\ell-1} \sigma^\dagger_{\ell-1} \right) = b_\ell k_{PZ} \left( e_\ell \sigma^\dagger_\ell - \sigma^\dagger_\ell + \vartheta e^{i k_{PZ}(d_{\ell'-1} - z')} \right). \]  

(3.27)
Define matrices
\[
S^{-1} = S^{(L+1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad S^{(\ell)} = \frac{1}{2e_\ell} \begin{bmatrix} e_\ell & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (a_\ell)^{-1} & (b_\ell k_{\ell z})^{-1} \\ (a_\ell)^{-1} & -(b_\ell k_{\ell z})^{-1} \end{bmatrix}, \quad 0 \leq \ell \leq L, \quad (3.28)
\]
and the transition matrices
\[
T_{\ell-1,\ell} = (2e_{\ell-1})^{-1} T_{\ell-1,\ell}, \quad 1 \leq \ell \leq L, \quad (3.29)
\]
where
\[
\tilde{T}_{\ell-1,\ell} = \begin{bmatrix} e_{\ell-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma^{+}_{\ell} & \gamma^{-}_{\ell} \\ \gamma^{-}_{\ell} & \gamma^{+}_{\ell} \end{bmatrix} \begin{bmatrix} e_{\ell} & 0 \\ 0 & 1 \end{bmatrix}.
\quad (3.30)
\]

The interface equations (3.25)–(3.27) can be solved in a unified recursive form
\[
\begin{bmatrix} \sigma_{\ell-1}^{\uparrow} \\ \sigma_{\ell-1}^{\downarrow} \end{bmatrix} = T_{\ell-1,\ell} \begin{bmatrix} \sigma_{\ell}^{\uparrow} \\ \sigma_{\ell}^{\downarrow} \end{bmatrix} + \delta_{\ell, \ell' + 1} S^{(\ell')} \begin{bmatrix} -a_{\ell'} \\ b_{\ell'} k_{\ell'z} \end{bmatrix} \varphi e^{i k_{\ell'z}(z'-d_{\ell'})} + \delta_{\ell, \ell} S^{(\ell'-1)} \begin{bmatrix} a_{\ell'} \\ b_{\ell'} k_{\ell'z} \end{bmatrix} \varphi e^{i k_{\ell'z}(d_{\ell'-1}-z')}
\quad (3.31)
\]
for $\ell = 1, \ldots, L$. Let
\[
A^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^{(\ell)} = \tilde{T}_{0,1} \cdots \tilde{T}_{\ell-1,\ell}, \quad 1 \leq \ell \leq L, \quad (3.32)
\]
\[
C^{(\ell)} = \prod_{j=0}^{\ell-1} \frac{1}{2e_j}, \quad 0 \leq \ell \leq L. \quad (3.33)
\]
By concatenating the recursion formula (3.31) for $\ell = 1, \cdots, t \leq L$, we have

\[
\begin{bmatrix}
\sigma_0^t \\
\sigma_0^\ell \\
\sigma^t \\
\sigma^\ell \\
\end{bmatrix} = C^{(t)} \Delta^{(t)} \begin{bmatrix}
\sigma^t \\
\sigma^\ell \\
\end{bmatrix} + 1_{\{t < \ell\}} C^{(\ell)} \Delta^{(\ell)} S^{(\ell)} \begin{bmatrix}
-a^\ell \\
b^\ell k_{\ell z} \\
\end{bmatrix} \varphi e^{ik_{\ell z}(z'-d_\ell)}
\]

\[\text{ (3.34)}\]

\[+ 1_{\{0 < \ell \leq t\}} C^{(\ell-1)} \Delta^{(\ell-1)} S^{(\ell-1)} \begin{bmatrix}
a^\ell \\
b^\ell k_{\ell z} \\
\end{bmatrix} \varphi e^{ik_{\ell z}(d_{\ell-1}-z')}.
\]

Note the indicators $1_{\{t < \ell\}}$ and $1_{\{0 < \ell \leq t\}}$ suggest when the following term should be counted in the sum.

One of the major numerical challenges for solving the linear system is the fact that $|e_\ell| \ll 1$ as $k_\rho \to \infty$ in the future numerical integration. Since $\sigma_0^t = \sigma_0^\ell = 0$, we can solve $\sigma^t_0$ using the second row of (3.34) with $t = L$. To recursively calculate $\sigma^\ell_0$ for $\ell = L-1, \cdots, 0$, we should carefully choose the formula used between the recursion (3.31) and its accumulated form (3.34), so that $e_\ell$ terms are kept in the numerator of any fraction. The detailed formulas are listed in Algorithm 3.1. The linear systems of $\sigma^\ell_0$ is guaranteed to be solvable for $k_\rho \in (\max_{0 \leq \ell \leq L} k_\ell, +\infty)$, which is a corollary of Theorem 2.5 for the 3-D case. For other values of $k_\rho$ the solvability can be guaranteed via analytic extension, except the (isolated) poles.

3.3. Far-field expansions of the reaction field

In this section, we derive the generalized Funk–Hecke formula [55] for dealing with the plane waves (and the equivalent exponential forms for the linearized PB equation) in the frequency domain, which are the mathematical foundation of the far-field expansions for LMGF in 3-D. Derivation of the formulas consists of two steps, the first is a generalized plane wave expansion, the second is a generalized Legendre addition theorem. Formulas of the far-field expansions including ME, LE and M2L will be presented in the end.
Algorithm 3.1: Recursively solving $\sigma_{\ell\ell'}$ using (3.31) and (3.34).

\[
\begin{align*}
\sigma_{\ell\ell'}^{\uparrow\uparrow} &= \sigma_{\ell\ell'}^{\downarrow\downarrow} = 0; \\
& \text{for } \ell = L \to 1 \text{ do} \\
\sigma_{\ell\ell'}^{\downarrow\uparrow} &= -\frac{A_{21}^{(\ell)}}{A_{22}^{(\ell)}} \sigma_{\ell\ell'}^{\uparrow\uparrow} - 1_{\{\ell' = \ell \}} \frac{\vartheta}{C_{0}^{(\ell)}} \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] A_{0}^{(\ell)} S_{\ell}^{(\ell')} \left[ \begin{array}{c} -a_{\ell'} \\ b_{\ell'} \end{array} \right]; \\
\sigma_{\ell\ell'}^{\uparrow\downarrow} &= -\frac{A_{21}^{(\ell)}}{A_{22}^{(\ell)}} \sigma_{\ell\ell'}^{\uparrow\uparrow} - 1_{\{0 < \ell' \leq \ell \}} \frac{\vartheta}{C_{0}^{(\ell)}} \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] A_{1}^{(\ell-1)} \left( 2e_{\ell' - 1} S_{\ell' - 1} \right) \left[ \begin{array}{c} a_{\ell'} \\ b_{\ell'} \end{array} \right]; \\
\sigma_{\ell-1,\ell'}^{\uparrow\uparrow} &= T_{11}^{\ell-1,\ell} \sigma_{\ell\ell'}^{\uparrow\uparrow} + T_{12}^{\ell-1,\ell} \sigma_{\ell\ell'}^{\downarrow\uparrow} + \delta_{\ell,\ell'+1} \left( -S_{11}^{(\ell')} a_{\ell'} + S_{12}^{(\ell')} b_{\ell'} k_{\ell' \ell} \right); \\
\sigma_{\ell-1,\ell'}^{\downarrow\downarrow} &= T_{11}^{\ell-1,\ell} \sigma_{\ell\ell'}^{\downarrow\downarrow} + T_{12}^{\ell-1,\ell} \sigma_{\ell\ell'}^{\downarrow\uparrow} + \delta_{\ell,\ell'} \left( S_{11}^{(\ell-1)} a_{\ell'} + S_{12}^{(\ell-1)} b_{\ell'} k_{\ell' \ell} \right); \\
\end{align*}
\]

end

$\sigma_{0\ell'}^{\uparrow\uparrow} = \sigma_{0\ell'}^{\downarrow\uparrow} = 0$;

We begin with a brief review of the ME for the free-space Green’s function, which is also known as the $h$–expansion. Given a source center $r_{c}$, let $(r'_{c}, \theta'_{c}, \phi'_{c})$ and $(r_{c}, \theta_{c}, \phi_{c})$ be the spherical coordinates of $r' - r_{c}$ and $r - r_{c}$, respectively. The $h$–expansion for the spherical Hankel function of the first kind is given by

\[
h_{0}^{(1)}(k|\vec{r} - \vec{r}'|) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} M_{nm} h_{n}^{(1)}(kr_{c}) Y_{n}^{m}(\theta_{c}, \phi_{c}),
\]

where $Y_{n}^{m}$ are spherical harmonic functions defined later in (3.44),

\[
M_{nm} = 4\pi j_{n}(kr'_{c}) Y_{n}^{m}(\theta'_{c}, \phi'_{c})
\]

are the ME coefficients, and the ME basis functions are formed by spherical Hankel functions with increasing orders. Note that $G^{i}(\vec{r}, \vec{r}') = \frac{i k}{4\pi} h_{0}^{(1)}(k|\vec{r} - \vec{r}'|)$ differs only by a constant factor. The $h$–expansion separates the summed information of the source and the target from a function, hence is also known as an addition theorem [55].
By using the modified spherical Bessel functions and the modified spherical Hankel functions, one can derive the ME that is mathematically equivalent to the $h$–expansion for the free-space Green’s function of the linearized PB equation [36, 43].

In the following discussion, the far-field expansions of the LMGF will be proposed with a similar structure that can indeed be extended to the free-space case.

3.3.1. The generalized Funk–Hecke formula

For any $k, r \in \mathbb{R}^3$, recall the plane wave expansion (PWE) using the spherical Bessel functions of the first kind $j_n$ and the Legendre polynomials $P_n$ [81],

$$e^{ik \cdot r} = \sum_{n=0}^{\infty} (2n + 1)i^n j_n(kr)P_n(\hat{k} \cdot \hat{r}),$$  \hspace{1cm} (3.41)

where $k = |k|$, $r = |r|$, and $\hat{k}$, $\hat{r}$ are unit vectors in $\mathbb{R}^3$ that have the same direction as $k$ and $r$, respectively. In the above series expansion, each Legendre polynomial can be further expanded into a finite sum by the well-known Legendre addition theorem

$$P_n(\hat{k} \cdot \hat{r}) = \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} \frac{1}{n-m} \frac{Y^m_n(\theta_k, \phi_k)}{Y^m_n(\theta, \phi)} Y^m_n(\theta, \phi),$$  \hspace{1cm} (3.42)

where $(r, \theta, \phi)$ form the spherical coordinate triple of $r$ such that

$$\hat{r} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta),$$  \hspace{1cm} (3.43)

$(k, \theta_k, \phi_k)$ form the spherical coordinate triple of $k$, $Y^m_n$ are the spherical harmonics defined as

$$Y^m_n(\theta, \phi) = (-1)^m \sqrt{\frac{2n + 1 (n - m)!}{4\pi (n + m)!}} P^m_n(\cos \theta) e^{im\phi} = \hat{P}^m_n(\cos \theta) e^{im\phi},$$  \hspace{1cm} (3.44)

where each $P^m_n$ is an associated Legendre functions, and $\hat{P}^m_n$ is its normalization. The series ex-
pansions (3.41) and (3.42) together completes a separation of the information of \( k \) and the information of \( r \) from the exponential left-hand side of (3.41), resulting in one double-sum series expansion, known as the Funk–Hecke formula [55, 40] due to the orthogonality of the spherical harmonics and the spherical Bessel functions.

Let \( k_\ell = (k_x, k_y, k_\ell z) \). The following lemmas provide various generalizations of the PWE (3.41).

**Lemma 3.4.** If \( k_\ell \in \mathbb{R}^+ \), then

\[
e^{i k_\ell \cdot r} = \sum_{n=0}^{\infty} (2n + 1)^n j_n(k_\ell r) P_n(\hat{k}_\ell \cdot \hat{r}),
\]

(3.45)

where \( \hat{k}_\ell = k_\ell^{-1} k_\ell \).

**Proof.** We will prove

\[
e^{i(k_\ell \cdot r)z} = \sum_{n=0}^{\infty} (2n + 1)^n j_n(k_\ell r) P_n(z), \quad \forall z \in \mathbb{C}
\]

(3.46)

and apply \( z = \hat{k}_\ell \cdot \hat{r} \).

From the PWE (3.41) we know the equality holds for any real \( z \in [-1, 1] \). We will prove the other cases by analytical extension. First, notice that the left-hand side \( e^{i(k_\ell \cdot r)z} \) is an entire function of \( z \). Then, we give an upper bound for each term in the expansion of the right-hand side. By [1, (9.1.62)],

\[
|j_n(k_\ell r)| \leq \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(n + \frac{3}{2})} \left(\frac{k_\ell r}{2}\right)^n = \frac{2^n}{(2n + 1)!!} \left(\frac{k_\ell r}{2}\right)^n.
\]

(3.47)

By the properties of the Legendre polynomials that \( P_n \) is a \( n \)-th polynomial with leading term coefficient \( \frac{(2n-1)!!}{n!} \) and with \( n \) distinct roots \( z_1, \cdots, z_n \) on the real interval \([-1, 1]\), we have

\[
|P_n(z)| = \frac{(2n-1)!!}{n!} \prod_{j=1}^{n} \left| z - z_j \right| \leq \frac{(2n-1)!!}{n!}(|z| + 1)^n, \quad \forall z \in \mathbb{C}.
\]

(3.48)
In total, we have
\[
\sum_{n=0}^{\infty} |(2n+1)i^n j_n(k_{\ell}r) P_n(z)| \leq \sum_{n=0}^{\infty} \frac{(2n+1) \cdot 2^n (2n-1)!!}{(2n+1)!! n!} \left( \frac{k_{\ell}r}{2} |z| + 1 \right)^n = e^{k_{\ell}r(|z|+1)} < \infty.
\] (3.49)

It’s easy to show that the right-hand side of (3.46) converges uniformly in any compact subset of \( \mathbb{C} \), hence converges to an entire function of \( z \) on \( \mathbb{C} \). By the analytic extension theory, the proof is finished.

\[\square\]

**Lemma 3.5 (Generalized PWE).** For any \( k_{\ell} \in \mathbb{C}, k_{\ell} \neq 0 \),
\[
e^{ik_{\ell}\cdot r} = \sum_{n=0}^{\infty} (2n+1)i^n j_n(k_{\ell}r) P_n(\hat{k}_{\ell} \cdot \hat{r}),
\] (3.50)

where \( \hat{k}_{\ell} = k_{\ell}^{-1} k_{\ell} \).

**Proof.** We will show the identity
\[
e^{i(\kappa r)z} = \sum_{n=0}^{\infty} (2n+1)i^n j_n(\kappa r) P_n(z), \quad \forall \kappa \in \mathbb{C},
\] (3.51)

where \( z = \hat{k}_{\ell} \cdot \hat{r} \), and apply \( \kappa = k_{\ell} \) to finish the proof. This time, we consider the analytic extension from (3.46), where the case \( \kappa \in \mathbb{R}^+ \) has been covered. Similar to the proof of Lemma 3.4, first notice that the left-hand side \( e^{i(\kappa r)z} \) is an entire function of \( \kappa \). Then, by [1, (9.1.62)], the estimate
\[
|j_n(\kappa r)| \leq \frac{2^n}{(2n+1)!!} \left( \frac{|\kappa r|}{2} \right)^n e^{\frac{3|\kappa r|}{2}}
\] (3.52)

together with the estimates on \( P_n(z) \) in the proof of Lemma 3.4 leads to the upper bound estimate
\[
\sum_{n=0}^{\infty} |(2n+1)i^n j_n(\kappa r) P_n(z)| \leq e^{r(|z|+1) + r|3\kappa|} < +\infty.
\] (3.53)

It’s easy to show that the right-hand side of (3.51) converges uniformly in any compact subset of \( \mathbb{C} \), hence converges to an entire function of \( \kappa \) on \( \mathbb{C} \). By the analytic extension theory, the proof is finished.

\[\square\]
theory, the proof is finished.

Before the discussion on a generalized Legendre addition theorem, we take a loot at the following useful lemma.

**Lemma 3.6.** Let \( f(x), g(x) \) be polynomials satisfying

\[
f(x) + \sqrt{1 - x^2} g(x) = 0, \quad \forall x \in [-1, 1].
\]  

(3.54)

Then, \( f \equiv g \equiv 0 \).

**Proof.** If \( f \equiv 0 \) or \( g \equiv 0 \), the proof is trivial. Suppose for contradiction that there exist a nontrivial polynomial \( f \) and a nontrivial polynomial \( g \) such that (3.54) holds. Without loss of generality, suppose nontrivial polynomials \( f \) and \( g \) minimize the nonnegative integer value

\[
\text{deg}(f) + \text{deg}(g)
\]  

(3.55)

subject to (3.54). Since \( f(-1) = f(1) = 0 \) and \( f \not\equiv 0 \), there exists a polynomial \( f_1(x) \not\equiv 0 \) such that

\[
f(x) = (1 - x^2)f_1(x),
\]  

(3.56)

which immediately follows that

\[
g(x) + \sqrt{1 - x^2} f_1(x) = 0, \quad \forall x \in (-1, 1).
\]  

(3.57)

By taking the limits as \( x \to \pm 1 \), the above equality also holds for \( x = \pm 1 \). Therefore, \( g \) and \( f_1 \) are another pair of nontrivial polynomials subject to (3.54), which break the minimization assumption, because

\[
0 \leq \text{deg}(g) + \text{deg}(f_1) = \text{deg}(g) + (\text{deg}(f) - 2) < \text{deg}(f) + \text{deg}(g).
\]

\[\square\]
The generalized Legendre addition theorem is stated as follows.

**Lemma 3.7 (Generalized Legendre addition theorem).** For any $k_\ell \in \mathbb{C}$, $k_\ell \neq 0$,

$$P_n(\hat{k}_\ell \cdot \hat{r}) = \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y_n^m(\theta, \phi) \hat{P}_n^m \left( \frac{k_{\ell z}}{k_\ell} \right) e^{i m\alpha},$$

(3.58)

where $\hat{k}_\ell = k^{-1}_\ell k_\ell$, $\alpha$ is the polar angle of $(k_x, k_y)$ as stated in (3.7).

**Proof.** Here only the first equality is proven. The second equality can be dealt with in a similar way. Define

$$D(w) = P_n \left( \sqrt{1 - w^2} \sin \theta \cos (\alpha - \phi) + w \cos \phi \right) - \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y_n^m(\theta, \phi) \hat{P}_n^m (w) e^{i m\alpha}. \quad (3.59)$$

For any $w \in [0, 1]$, by the Legendre addition theorem (3.42) with the real unit vectors

$$\hat{k} = (\sqrt{1 - w^2} \cos \alpha, \sqrt{1 - w^2} \sin \alpha, w),$$

$$\hat{r} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta), \quad (3.60)$$

we have $D(w) = 0$.

Since $P_n$ is a polynomial, each normalized associated Legendre function $\hat{P}_n^m (w)$ is either a polynomial of $w$ or such a polynomial multiplied by $\sqrt{1 - w^2}$, there exist polynomials $D_0(w)$ and $D_1(w)$ such that

$$D(w) = D_0(w) + \sqrt{1 - w^2} D_1(w). \quad (3.61)$$

By Lemma 3.6, $D_0(w) \equiv D_1(w) \equiv 0$.

By letting $w = k_{\ell z}/k_\ell$ in (3.61), the proof is finished. $\Box$

With Lemma 3.5 and Lemma 3.7, we arrive at the generalized Funk–Hecke formula.
Theorem 3.8 (Generalized Funk–Hecke formula). Let \( r \in \mathbb{R}^3 \) have spherical coordinates \((r, \theta, \phi)\). Let \( k_z \) be any nonzero complex number. Let \( k = (k_x, k_y, k_z) \). Then, we have the following series expansions

\[
e^{ik_zr} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j_n(k_z r) Y_n^m(\theta, \phi) \cdot i^n \hat{\mathcal{P}}_n \left( \frac{k_z}{k} \right) e^{i\alpha}, \quad (3.62)
\]

\[
e^{-ik_zr} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j_n(k_z r) Y_n^m(\theta, \phi) \cdot i^n \hat{\mathcal{P}}_n \left( \frac{k_z}{k} \right) e^{-i\alpha}. \quad (3.63)
\]

3.3.2. The far-field expansions and translations

The far-field expansions for each reaction field component is then proposed using (3.62) and (3.63) to separate (3.18).

To derive the ME, we split the difference \( r - r' = (r - r_c) + (r_c - r') \) for each one source \( r' \) in the exponential term of the integrand in (3.17), namely,

\[
E^{*\ast}_{\ell\ell'}(r, r', k) = E^{*\ast}_{\ell\ell'}(r, r_c, k) \cdot \exp \left( ik_x(x_c - x') + ik_y(y_c - y') + ik_z\tau^*(z' - z_c) \right), \quad (3.64)
\]

where the source center \( r_c = (x_c, y_c, z_c) \) should be on the same side of the interface \( z = d_z^c \), i.e. \( z_c - d_z^c \) and \( z' - d_z' \) should have the same sign. Let \((r'_c, \theta'_c, \phi'_c)\) be the spherical coordinates of \( r' - r_c \). Then, the triple \((x_c - x', y_c - y', \tau^*(z' - z_c)) \in \mathbb{R}^3 \) has the corresponding spherical coordinates \((r'_c, \frac{\pi}{2} + \theta' - \frac{\pi}{2}, \phi'_c + \pi)\). By using (3.62),

\[
\exp \left( ik_x(x_c - x') + ik_y(y_c - y') + ik_z\tau^*(z' - z_c) \right) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j_n(k_z r'_c) Y_n^m \left( \frac{\pi}{2} + \tau^* \left( \frac{\theta'_c - \frac{\pi}{2}}{2} \right), \phi'_c + \pi \right) i^n \hat{\mathcal{P}}_n \left( \frac{k_z}{k'} \right) e^{i\alpha} \quad (3.65)
\]

\[
= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j_n(k_z r'_c) Y_n^m(\theta'_c, \phi'_c) i^n(-1)^m(\tau^*)^{m+n} \hat{\mathcal{P}}_n \left( \frac{k_z}{k'} \right) e^{i\alpha}.
\]

Let

\[
M_{nm}(r', r_c) = 4\pi j_n(k_z r'_c) Y_n^m(\theta'_c, \phi'_c), \quad (3.66)
\]

73
the ME of the reaction field component \( u_{\ell'\ell}^{**}(\mathbf{r}; \mathbf{r}') \) is given by

\[
\begin{align*}
    u_{\ell'\ell}^{**}(\mathbf{r}; \mathbf{r}') &= \frac{1}{4\pi^2} \int \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell}^{**}(\mathbf{r}, \mathbf{r}', k_\rho) \sigma_{\ell'\ell}^{**}(k_\rho) dk_x dk_y \\
    &= \frac{1}{4\pi^2} \int \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell}^{**}(\mathbf{r}, \mathbf{r}_c, k_\rho) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} M_{nm}(\mathbf{r}', \mathbf{r}_c) i^n (-1)^m (\tau^*)^m \hat{F}_n \left( \frac{k_{\ell'z}}{k_{\ell'}} \right) e^{im\alpha} \sigma_{\ell'\ell}^{**}(k_\rho) dk_x dk_y \\
    &\approx \sum_{n=0}^{P} \sum_{m=-n}^{n} M_{nm}(\mathbf{r}', \mathbf{r}_c) F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c),
\end{align*}
\]

(3.67)

where

\[
F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c) = \frac{1}{4\pi^2} \int \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell}^{**}(\mathbf{r}, \mathbf{r}_c, k_\rho) i^n (-1)^m (\tau^*)^m \hat{F}_n \left( \frac{k_{\ell'z}}{k_{\ell'}} \right) e^{im\alpha} \sigma_{\ell'\ell}^{**}(k_\rho) dk_x dk_y
\]  

(3.68)

are the ME basis function, and \( M_{nm}(\mathbf{r}', \mathbf{r}_c) \) serve as the ME coefficients. When there are multiple sources \( \mathbf{r}'_j \) with strength \( q_j \) provided, \( j = 1, \cdots, M \), the ME coefficients are the weighted sums

\[
\sum_{j=1}^{M} q_j M_{nm}(\mathbf{r}'_j, \mathbf{r}_c), \quad 0 \leq n \leq P, \quad -n \leq m \leq n.
\]

(3.69)

The ME coefficients for the LMGF are identical to the ones for the free-space Green’s functions (3.40) with wave number \( k_{\ell'} \), see also [76, 79]. In fact, if we treat the free-space problem as a two-layer problem with the same physical parameters, the ME basis function \( F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c) \) defined above will serve as an integral representation of the corresponding free-space ME basis function.

To derive the LE, we split the difference \( \mathbf{r} - \mathbf{r}' = \mathbf{r}_c - \mathbf{r}' + \mathbf{r} - \mathbf{r}_c^l \) for each one target \( \mathbf{r} \), namely,

\[
\mathcal{E}_{\ell'\ell}^{**}(\mathbf{r}, \mathbf{r}', k_\rho) = \mathcal{E}_{\ell'\ell}^{**}(\mathbf{r}_c, \mathbf{r}', k_\rho) \cdot \exp \left( i k_{\ell z} (x - x_{\ell c}) + i k_{\ell y} (y - y_{\ell c}) + i k_{\ell z} \tau^* (z - z_{\ell c}) \right),
\]

(3.70)
where the target center \( r^t_c = (x^t_c, y^t_c, z^t_c) \) should be on the same side of the interface \( z = d^t \), i.e. \( z - d^t \) and \( z^t_c - d^t \) should have the same sign. Let \((r^t, \theta^t, \phi^t)\) be the spherical coordinates of \( r - r^t_c \). Then, the triple \((x - x^t_c, y - y^t_c, \tau^*(z - z^t_c))\) has the spherical coordinates \((r^t, \frac{\pi}{2} + \tau^*(\theta^t - \frac{\pi}{2}), \phi^t)\). By using (3.63),

\[
\exp \left( ik_x(x - x^t_c) + ik_y(y - y^t_c) + ik_z \tau^*(z - z^t_c) \right) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j_n(k \ell r^t) Y^m_n \left( \frac{\pi}{2} + \tau^*(\theta^t - \frac{\pi}{2}), \phi^t \right) \cdot i^n \hat{P}_n \left( k \ell z \right) e^{-i\alpha} \tag{3.71}
\]

Let

\[
K_{nm}(r, r^t_c) = j_n(k \ell r^t) Y^m_n(\theta^t, \phi^t), \tag{3.72}
\]

the LE of the reaction field component \( u^{**}_{\ell\ell'}(r; r') \) is given by

\[
u^{**}_{\ell\ell'}(r; r')
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}^{**}_{\ell\ell'}(r, r'; k_{\rho}) \sigma^{**}_{\ell\ell'}(k_{\rho}) dk_{\rho} dk_y
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}^{**}_{\ell\ell'}(r^t_c, r', k_{\rho}) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} K_{nm}(r, r^t_c) \cdot 4\pi i^n(\tau^*)^{n+m} \hat{P}_n \left( k \ell z \right) e^{-i\alpha} \sigma^{**}_{\ell\ell'}(k_{\rho}) dk_{\rho} dk_y
\approx \sum_{n=0}^{P} \sum_{m=-n}^{n} K_{nm}(r, r^t_c) L^{**}_{nm}(r', r^t_c), \tag{3.73}
\]

where

\[
L^{**}_{nm}(r', r^t_c) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}^{**}_{\ell\ell'}(r^t_c, r', k_{\rho}) 4\pi i^n(\tau^*)^{n+m} \hat{P}_n \left( k \ell z \right) e^{-i\alpha} \sigma^{**}_{\ell\ell'}(k_{\rho}) dk_{\rho} dk_y \tag{3.74}
\]

are the LE coefficients, and \( K_{nm} \) are the LE basis functions. The LE basis functions for the LMGF are identical to the ones for the free-space Green’s functions with wave number \( k_{\ell} \), see [76, 79].
To derive the M2L translation, we split the exponent

$$E_{\ell'\ell}^{*\star}(r_c, r', k_\rho) = E_{\ell'\ell}^{*\star}(r_c, r_c, k_\rho) \cdot \exp \left( ik_z(x_c - x') + ik_y(y_c - y') + ik_{\ell' z} x^* (z' - z_c) \right) \quad (3.75)$$

in the integral form (3.74) of each LE coefficient $L_{nm}^{*\star}(r', r_c')$ with the same expansion as in (3.65). Similar as the derivation of the ME, we have

$$L_{nm}^{*\star}(r', r_c') \approx \sum_{n'=-m'}^{P} \sum_{m'} A_{nm; n'm'}^{*\star}(r_c, r_c') M_{n'm'}^{*\star}(r', r_c') \quad (3.76)$$

where

$$A_{nm; n'm'}^{*\star}(r_c, r_c') = \frac{1}{\pi} \int_{\mathbb{R}^2} E_{\ell'\ell}^{*\star}(r_c, r_c, k_\rho) i^{n+n'+2m}(\tau^*)^{n+m}(\tau^*)^{n'+m'} \frac{\hat{P}_n}{k_{\ell'}} \frac{\hat{P}_{n'}}{k_{\ell}} e^{i(m'-m)\alpha} \sigma_{\ell'\ell}^{*\star}(k_\rho) dk_x dk_y \quad (3.77)$$

are the M2L translation coefficients.

The M2M is the same as the free-space versions, as the ME coefficients are identical. For the L2L, although the LE coefficients $L_{nm}^{*\star}(r', r_c')$ are distinct from the free-space versions, the LE basis functions remain the same. Since the L2L can be interpreted as the linear transform for a change of the LE basis, the translation is equivalent to the free-space case. Details of the free-space implementation can be found in [36, 43, 14].

### 3.3.2.1. Numerical validation of exponential convergence

Define the polarization distance

$$D_{\ell'\ell}^{*\star}(r, r') = \sqrt{(x - x')^2 + (y - y')^2 + (\tau^* (z - d_{\ell}^*) + \tau^* (z' - d_{\ell'}^*))} \quad (3.78)$$

like in the 2-D case (2.80).
We will numerically verify the exponential convergence of the derived ME (3.67) by calculating the reaction field component $u_{11}^{↓↓}$ in a 3-layer medium, and comparing the convergence with the far-field ratio determined with the polarization distance (3.78).

The medium has 3 layers divided by interfaces $z = d_0 = 0.5$ and $z = d_1 = -0.5$. For testing the Helmholtz case, we set

$$k_1 = 2.0, \quad k_2 = 3.0, \quad k_3 = 4.7,$$

(3.79)

while for the linearized PB case we set

$$\lambda = 2.0, \quad \lambda_2 = 3.0, \quad \lambda_3 = 4.7,$$

(3.80)

while $k_\ell = i\lambda_\ell$. The interface conditions are equipped with parameters

$$a_0 = a_1 = a_2 = 1.0, \quad b_0 = 1.0, \quad b_1 = 3.0, \quad b_2 = 5.0.$$

(3.81)

The ME center is fixed at the origin, i.e., $r_c = (0.0, 0.0, 0.0)$. The following 3 target-source pairs are used in the numerical test:

1. $r = (0.5, 0.3, -0.2), \quad r' = (-0.1, -0.2, 0.4)$;
2. $r = (-0.5, 0.7, -0.2), \quad r' = (-0.1, 0.2, 0.0)$;
3. $r = (0.0, 0.1, -0.8), \quad r' = (0.1, 0.0, 0.0)$.

For each pair, we numerically compute the relative error

$$e_P^{ME} = \left| u_{11}^{↓↓}(r; r') - \sum_{n=0}^{P} \sum_{m=-n}^{n} M_{mm}(r', r_c) F_{nm}^{↓↓}(r, r_c) \right| \left/ \left| u_{11}^{↓↓}(r; r') \right| \right|,$$

(3.82)
For comparison, we define the reference exponential convergence ratio

\[ r_{ME} = \frac{|r' - r_c|}{D_{11}^{++}(r, r_c)}. \]  

(3.83)

Figure 3.1 shows the relative errors \( e_{P}^{ME} \), \( P = 3, 4, \cdots, 12 \) for the Helmholtz equation (left) and the linearized PB equation (right), respectively. The relative errors are compared with the reference exponential convergence rates indicated by the corresponding colored dashed lines with slopes \( \log_{10} r_{ME} \), respectively, which shows that the relative errors decay at the expected exponential rates, determined by the polarization distance (3.78).

Figure 3.1: Relative errors of ME for the reaction field component \( u_{11}^{++} \) from the LMGF of the Helmholtz equation and the linearized PB equation for \( P = 3, 4, \cdots, 12 \) in a 3-layer medium for three cases, compared to the reference exponential convergence rates indicated by the polarization distance.

3.4. The FMM framework

Due to the exponential convergence result verified in the previous section, we can develop a similar framework for the FMM for LMGF like in the 2-D case with the concept of polarization sources like in (2.92).
Definition 3.9 (Polarization source). Let $r_2 = (x_2, y_2, z_2)$ be any source point in layer $\ell'$ (hence satisfying $\tau^*(z_2 - d_\ell^*) > 0$). Let

$$P_{\ell\ell'}^{*\star} : r_2 \mapsto \tilde{r}_2 = (x_2, y_2, d_\ell^* - \tau^*\tau^*(z_2 - d_\ell^*))$$

(3.84)

be a bijective mapping. $P_{\ell\ell'}^{*\star}(r_2)$ is called the polarization source of $r_2$.

Given target particles $r_i$ in layer $\ell$, and source particles $r_j'$ with strength $q_j$ in layer $\ell'$, the FMM for the reaction field

$$u_{\ell\ell'}^{*\star}(r_i) = \sum_j q_j u_{\ell\ell'}^{*\star}(r_i; r_j')$$

(3.85)

can be described in Algorithm 3.2.

Then, by joining the interaction for all the target-source layer pairs $(\ell, \ell')$, together with the free-space part using a conventional approach, the overall algorithm is summarized in Algorithm 3.3.

It is worth mentioning that the FMM box used in Algorithm 3.2 should place its center on the interface between the target layer and the polarization source layer as shown by Figure 3.2, so that children boxes won’t lie across the interface.

Remark 3.10. Algorithm 3.2 and Algorithm 3.3 separate the multi-layer problem into a number of target-source layer pairs. For each reaction field component, the FMM is reduced to a local problem with two layers. This is true even when the targets and the sources come from the same layer.

3.5. Implementation details for efficiency

In this section we discuss a few methods to improve the efficiency as well as the accuracy in the numerical implementation of the FMM for LMGF.
Algorithm 3.2: FMM for reaction field component $u_{\ell\ell'}^\star(r_i)$.

Form the polarized source layer with each source $r'_j$ mapped to $P_{\ell\ell'}^\star(r'_j)$;
Generate an adaptive hierarchical octree with levels $0 \rightarrow H$ with polarization 
sources $P_{\ell\ell'}^\star(r'_j)$ and original targets $r_i$;

// Initialize the output
for source particle $r_i$ do
    $u_{\ell\ell'}^\star(r_i) = 0$;
end

// Sanity check: skip if reaction field is prohibited
if $(\ell, \star)$ or $(\ell', \star) \in \{(0, \downarrow), (L, \uparrow)\}$ then return;

// The upward pass
for $\ell = H \rightarrow 0$ do
    for box $j$ on source tree level $\ell$ do
        if $j$ is a leaf node then
            Form the ME using (3.67);
        else
            Form the ME by merging children’s expansions using the free-space M2M 
            translation, e.g. [76, (13)] or [79, (3.5)];
        end
    end
end

// The downward pass
for $\ell = 1 \rightarrow H$ do
    for box $j$ on target tree level $\ell$ do
        Shift the LE of $j$’s parent to $j$ using the free-space L2L translation, e.g. [76, 
        (13)] or [79, (3.5)];
        Collect the contribution from the interaction list using the M2L (3.76);
    end
end

// LE evaluation
for leaf node $j$ do
    for source particle $r_i$ in node $j$ do
        Evaluate the LE at $r_i$ with (3.73), add add the result to $u_{\ell\ell'}^\star(r_i)$;
    end
end

// Neighboring direct interactions
for source particle $r_i$ do
    for target particle $r'_j$ in a neighboring leaf box of $r_i$ do
        Add the direct interaction $q_j u_{\ell\ell'}^\star(r_i; r'_j)$ to $u_{\ell\ell'}^\star(r_i)$;
    end
end
Figure 3.2: Polarization sources and boxes in the source tree of FMM.

3.5.1. Conversion to single integrals

In the ME, LE and M2L formulas derived above, the reaction field component \(u_{\ell\ell'}^{*}(r; r')\) and other various functions are given in the double integral form inherited from the 2-D Fourier transform (3.6). These double integrals can be reduced to single integrals of \(k_{\rho}\) for efficient computation. Namely, suppose \(m \in \mathbb{Z}\), \((\tilde{x}, \tilde{y}) \in \mathbb{R}^2\) with polar coordinates \((\tilde{\rho}, \tilde{\phi})\).

By using the identity of the Bessel functions of the first kind

\[
\int_{0}^{2\pi} e^{ik_{\rho}\tilde{\rho}\cos(\alpha-\tilde{\phi}) + im(\alpha-\tilde{\phi})} d\alpha = 2\pi i^m J_m(k_{\rho}\tilde{\rho}),
\]

which is a Sommerfeld-type integral, or is interpreted as a Hankel transform of order \(|m|\). For
Algorithm 3.3: FMM in 3-D layered media

// The free-space part
for $\ell = 0 \rightarrow L$ do
    For source particles $r_i$ and target particles in layer $\ell$, compute $u^f(r_i)$ with the classic FMM for the Helmholtz equation [14] or the linearized PB equation [36, 43] in the free space;
end

// The reaction fields
for $\ell = 0 \rightarrow L$ do
    for $\ell' = 0 \rightarrow \ell$ do
        For source particle $r_i$ in layer $\ell$ and target particles in layer $\ell'$, compute $u_{\ell\ell'}^{\uparrow\uparrow}(r_i)$, $u_{\ell\ell'}^{\uparrow\downarrow}(r_i)$, $u_{\ell\ell'}^{\downarrow\uparrow}(r_i)$ and $u_{\ell\ell'}^{\downarrow\downarrow}(r_i)$ with Algorithm 3.2;
    end
end

// Summing up
for $\ell = 0 \rightarrow L$ do
    for source particle $r_i$ in layer $\ell$ do
        $u(r_i) = u^f(r_i) + \sum_{\ell'=0}^{L} \left( u_{\ell\ell'}^{\uparrow\uparrow}(r_i) + u_{\ell\ell'}^{\uparrow\downarrow}(r_i) + u_{\ell\ell'}^{\downarrow\uparrow}(r_i) + u_{\ell\ell'}^{\downarrow\downarrow}(r_i) \right)$;
    end
end

$m < 0$, one can use the formula $J_{-m}(k\rho\tilde{\rho}) = (-1)^m J_m(k\rho\tilde{\rho})$ to avoid repeated computation.

3.5.2. Real arguments for the linearized Poisson–Boltzmann equation

For the linearized PB equation where $k_\ell = i\lambda_\ell$, one may use modified spherical Bessel functions $i_n$ instead of using $j_n$ with a purely imaginary parameter. Namely,

$$j_n(k_\ell r) = i^n i_n(\lambda_\ell r). \tag{3.88}$$

In the evaluation of the associated Legendre functions

$$\hat{P}_n^m \left( \frac{k_{\ell z}}{k_\ell} \right), \tag{3.89}$$
one can write
\[
\frac{k_{\ell z}}{k_\ell} = \frac{i\sqrt{\lambda^2 + k_\rho^2}}{i\lambda_\ell} = \frac{\sqrt{\lambda^2 + k_\rho^2}}{\lambda_\ell}. \tag{3.90}
\]

However, implementations of \( \hat{P}_n^m(x) \) have different branch cuts affecting the values at \( x \in [1, \infty) \), which we must specify. Recall the definition for \( x \in (-1, 1) \),
\[
\hat{P}_n^m(x) = \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} (1-x^2)^\frac{m}{2} \frac{d^{n+m}}{dx^{n+m}} \left((x^2-1)^n\right)}, \tag{3.91}
\]
where the factor \((1-x^2)^\frac{m}{2}\) is multivalued if extended for \( x \in \mathbb{C} \) when \( m \) is an odd number.

In the Helmholtz case where \( k_\ell > 0 \), if \( k_\rho \) is chosen from the real interval \((0, k_\ell)\), we adopt the value
\[
(1-x^2)^\frac{1}{2} \bigg|_{x=k_{\ell z}/k_\ell} = \frac{k_\rho}{k_\ell}. \tag{3.92}
\]
As a natural extension, for the linearized PB case, we specify
\[
(1-x^2)^\frac{1}{2} \bigg|_{x=k_{\ell z}/k_\ell} = \frac{k_\rho}{k_\ell} = -i\frac{k_\rho}{\lambda_\ell}. \tag{3.93}
\]
Note that in a recursive scheme evaluating the associated Legendre functions, the term \((1-x^2)^\frac{1}{2}\bigg|_{x=k_{\ell z}/k_\ell}\) can be calculated only once.

3.5.3. Contraction of associated Legendre functions

The M2L usually takes the most computational cost in the overall algorithm, where we calculate the coefficients \( A^{**}_{nm;n'm'} \) defined in (3.77) with numerical integration methods such as the Gauss–Legendre quadrature. The contraction of associated Legendre functions in the integrand will reduce the number of numerical integration from \( O(P^4) \) to \( O(P^2) \), based on the observation that each function
\[
\hat{P}_n^m \left( \frac{k_{\ell z}}{k_\ell} \right) = \left( \frac{k_{\ell z}}{k_\ell} \right)^{\eta_{nm}} Q_n^m(k_\rho; k_\ell), \tag{3.94}
\]
where $Q^m_n$ is an $(n - \eta_{nm})$-th order polynomial of $k_\rho$ whose coefficients are polynomials of $k_\ell$, and

$$\eta_{nm} = (n + m) \mod 2. \quad (3.95)$$

We can then expand $A_{nm;n'm'}^{**}$ in (3.77) into a linear combination of

$$\int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^*(r^l_c, r_c, k_\rho) \left(\frac{k_{\ell z}}{k_{\ell'}}\right)^{\eta_{nm}} \left(\frac{k_{\ell' z}}{k_{\ell}}\right)^{\eta_{n'm'}} k^j_\rho e^{i(m'm'-m)\alpha} \sigma_{\ell\ell'}^*(k_\rho) dk_z dk_y \quad (3.96)$$

for $0 \leq j \leq n + n' - \eta_{nm} - \eta_{n'm'}$ with constant coefficients. Computation of these constants has $O(P^5)$ complexity for each pair of target-source layers $(\ell, \ell')$, but they can be pre-computed regardless of the $O(N)$ sized number of the particles.

The coefficients are found in the following derivation [76, (82)–(89)]. We first consider the case for $n \geq m \geq 0$ and $n' \geq m' \geq 0$. Let

$$c_{nm} = \sqrt{\frac{2n + 1}{4\pi} \frac{(n - m)!}{(n + m)!}}, \quad a_{nm}^j = \frac{(-1)^{n-j}(2j)!}{2^n j!(n-j)!(2j - n - m)!}. \quad (3.97)$$

Then

$$\frac{c_{00}(x^2 - 1)^n}{2^nn!} = \sum_{k=0}^{n} a_{00}^j x^{2j}, \quad c_{nm} \frac{d^{n+m}}{dx^{n+m}} (x^2 - 1)^n = \sum_{j=\left\lceil \frac{n+m}{2} \right\rceil}^{n} a_{nm}^j x^{2j-n-m}. \quad (3.98)$$

Let

$$r = \left\lceil \frac{n + |m|}{2} \right\rceil, \quad r' = \left\lceil \frac{n' + |m'|}{2} \right\rceil, \quad (3.99)$$

the polynomial $Q^m_n(k_\rho; k_\ell)$ can be written by

$$Q^m_n(k_\rho; k_\ell) = \sum_{s=0}^{n-r} b_{nm}^s \left(\frac{k_\rho}{k_\ell}\right)^{m+2s}, \quad (3.100)$$

84
where each coefficient

\[ b_{nm}^t = \sum_{j=q}^{n} \frac{(-1)^s q_{nm}(j-r)!}{s!(j-r-s)!}, \quad q = \max \left( \left[ \frac{n+m}{2} \right], s+r \right). \tag{3.101} \]

Furthermore,

\[ \hat{P}_n^m \left( \frac{k_{\ell z}}{k_{\ell'}} \right) \hat{P}_{n'}^{m'} \left( \frac{k_{\ell z}}{k_{\ell'}} \right) = \left( \frac{k_{\ell z}}{k_{\ell'}} \right)^{\eta_{nm}} \left( \frac{k_{\ell z}}{k_{\ell'}} \right)^{\eta_{n'm', n-r+n'-r'}} \sum_{s=0}^\infty \frac{\alpha_{nn'm'm'}^s k_{m+m'+2s}}{k_{\ell}^{s-t} k_{\ell'}^{s-t}} \right), \tag{3.102} \]

where

\[ \alpha_{nn'm'm'}^s = \sum_{t=\max(s-n'+r',0)}^{\min(s,n-r)} \frac{b_{nn'}^t b_{n'm'}^{s-t}}{K_{m+2t}^{m'+2(s-t)}}. \tag{3.103} \]

For \(-n \leq m < 0\) or \(-n' \leq m' < 0\), similar formulas can be obtained by using the fact \(\hat{P}_n^m(z) = (-1)^m \hat{P}^{-m}_n(z)\) for \(m < 0\). In general, we have

\[ \hat{P}_n^m \left( \frac{k_{\ell z}}{k_{\ell'}} \right) \hat{P}_{n'}^{m'} \left( \frac{k_{\ell z}}{k_{\ell'}} \right) = \left( \frac{k_{\ell z}}{k_{\ell'}} \right)^{\eta_{nm}} \left( \frac{k_{\ell z}}{k_{\ell'}} \right)^{\eta_{n'm', n-r+n'-r'}} \sum_{s=0}^\infty \frac{\alpha_{nn'm'm'}^s k_{|m|+|m'|+2s}}{k_{\ell}^{s-t} k_{\ell'}^{s-t}} \right), \tag{3.104} \]

and

\[ \alpha_{nn'm'm'}^s = \sum_{t=\max(s-n'+r',0)}^{\min(s,n-r)} \frac{\tau_{nm} \tau_{n'm'} b_{n|m|}^{s-t} b_{n'|m'|}^{s-t}}{K_{m+2t}^{m'+2(s-t)}} \right), \quad \tau_{\nu} = \begin{cases} 1, & \nu \geq 0, \\ (-1)^\nu, & \nu < 0. \end{cases} \tag{3.105} \]

3.5.4. Scaling factors

When the truncation index \(P\) goes large, and when box sizes vary a lot across levels, the special functions could have extreme values, and the evaluation becomes less accurate. Multiplying by scaling factors can ease such difficulties. For example, in the software package DASHMM [25], the scaling factor for each ME coefficient \(M_{nm}(r', r_c)\) for the free-space FMM is \(S^{-n}\), where \(S\) is the length of edge of the box where the sources locate.
3.6. Numerical results

In this section, we present numerical results to demonstrate the performance of the proposed FMM for the Helmholtz equation and the linearized PB equation in 3-D layered media. Data of numerical test results are from published journal papers [76] and [79], respectively, in collaboration with Bo Wang and Wei Cai. Bo Wang has been the primary contributor in the development of the code implementation for 3-D FMM in layered media.

3.6.1. Test environment

The FMM for Helmholtz equation and for linearized PB equation in 3-D layered media are implemented based on an open-source adaptive FMM package DASHMM [25] on a workstation with two Xeon E5-2699 v4 2.2 GHz processors (each having 22 cores and 44 threads) and 500GB RAM using the GCC compiler version 6.3.

We test the algorithms in a medium with three layers, with interfaces placed at $d_0 = 0$, $d_1 = -1.2$. Particles are set to be uniformly distributed in irregular domains which are obtained by shifting the domain determined by $r = 0.5 - a + \frac{a}{8}(35\cos^4\theta - 30\cos^2\theta + 3)$ with $a = 0.1, 0.15, 0.05$ to new centers $(0,0,0.6)$, $(0,0,-0.6)$ and $(0,0,-1.8)$, respectively (see Figure 3.3 for the cross section of the domains). All particles are generated by keeping the uniformly distributed particles in a larger cube within corresponding irregular domains.

In both tests, let $u_\ell(r_i)$ be the field at $r_i$ from layer $\ell$, including both the free-space part and the reaction field part, while in the free-space part the interaction with itself is excluded. Let $\tilde{u}_\ell(r_i)$ be the approximated values of $u_\ell(r_i)$ calculated by the FMM. Define $\ell^2$ and maximum errors as

$$\text{Err}_{\ell}^2 := \sqrt{\frac{\sum_{i=1}^{N_\ell} |u_\ell(r_i) - \tilde{u}_\ell(r_i)|^2}{\sum_{i=1}^{N_\ell} |u_\ell(r_i)|^2}}, \quad \text{Err}_{\ell}^{\text{max}} := \max_{1 \leq i \leq N_\ell} \frac{|u_\ell(r_i) - \tilde{u}_\ell(r_i)|}{|u_\ell(r_i)|},$$

(3.106)

where $N_\ell$ is the number of target particles in layer $\ell$. To test the accuracy of the FMM, we
put $N = 912 + 640 + 1296$ particles in the irregular domains in the layers, see Figure 3.3.

For the computation of reaction field components, the numerical integrals for M2L are pre-computed with the contraction of Legendre functions presented in the previous section.

3.6.2. Numerical results for FMM for Helmholtz equation in 3-D layered media

The wave numbers for each layer are given by $k_0 = 1.2$, $k_1 = 1.5$, $k_2 = 1.8$. Convergence rates against the truncation index $P$ are depicted in Figure 3.4 (a). The CPU time in seconds for the computation of all three free-space components \( \{u_{\ell}^f(r_i)\}_{\ell=0}^2 \), three selected reaction components \( \{u_{00}^{\uparrow\uparrow}, u_{11}^{\uparrow\uparrow}, u_{22}^{\downarrow\downarrow} \} \) and all sixteen reaction components $u_{\ell\ell'}^{*}(r_i)$ with truncation index $P = 4$ are compared in Figure 3.4 (b) for up to 3 millions particles.

CPU time with multiple cores is given in Table 3.1.
3.6.3. Numerical results for FMM for linearized PB equation in 3-D layered media

In the layered medium, the dielectric constants \( \{ \varepsilon_\ell \}_{\ell=0}^2 \) and the inverse Debye–Huckel lengths \( \{ \lambda_\ell \}_{\ell=0}^2 \) are set to be

\[
\varepsilon_0 = 1.0, \quad \varepsilon_1 = 8.6, \quad \varepsilon_2 = 20.5, \quad \lambda_0 = 1.2, \quad \lambda_1 = 0.5, \quad \lambda_2 = 2.1.
\]

Convergence rates against \( p \) are depicted in Fig. 3.5 (a). We test the FMM for up to 3 millions particles, and the CPU time for the computation of all three free space components \( \{ u_\ell^f(\mathbf{r}_i) \}_{\ell=0}^2 \), three selected reaction components \( \{ u_{00}^{\uparrow\uparrow}, u_{11}^{\uparrow\uparrow}, u_{22}^{\downarrow\downarrow} \} \) and all sixteen reaction components \( u_{\ell\ell}^{**}(\mathbf{r}_i) \) with truncation index \( P = 5 \) are compared in Figure 3.5 (b).

CPU time with multiple cores is given in Table 3.2.

3.6.4. Summary

It has been shown by both numerical tests that the FMM for all of the selected reaction field components have \( O(N) \) complexity, while the CPU time for the computation of reaction
<table>
<thead>
<tr>
<th>cores</th>
<th>number of particles</th>
<th>time for all ${u_\ell^J }_{\ell=0}$</th>
<th>time for ${u_{00}^{\uparrow\uparrow}, u_{11}^{\uparrow\uparrow}, u_{22}^{\downarrow\downarrow}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>618256</td>
<td>46.22</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>121.3</td>
<td>13.32</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>285.4</td>
<td>15.15</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>312.2</td>
<td>16.51</td>
</tr>
<tr>
<td>6</td>
<td>618256</td>
<td>8.75</td>
<td>6.48</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>22.64</td>
<td>7.58</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>53.50</td>
<td>7.94</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>58.55</td>
<td>8.56</td>
</tr>
<tr>
<td>36</td>
<td>618256</td>
<td>2.04</td>
<td>6.51</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>5.05</td>
<td>6.83</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>11.62</td>
<td>7.34</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>12.71</td>
<td>7.22</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of CPU time (in seconds) with multiple CPU cores for FMM for Helmholtz equation in 3-D layered media, the truncation index $P = 4$, data from [76].

components has a much smaller linear scaling constant compared to the free-space part. This is due to the fact that when targets and sources come from the same layer, the polarization sources are better separated from the targets in the reaction field, when compared to the free-space computation where short-range interactions must be directly calculated.

Since the reaction field components can be computed independently, it is straightforward to implement a version of code that computes them in parallel. However, as shown by Table 3.1 and Table 3.2, due to the small proportion of CPU time spent on the reaction field components, potential speedup of the parallel computation is mainly decided by the free-space part.

3.7. Conclusion

In this chapter, we have presented the fast multipole method for the efficient calculation of the discretized integral operators for the Helmholtz equation and the linearized Poisson–Boltzmann equation in 3-D layered media. The layered media Green’s functions
Figure 3.5: Performance of FMM for linearized PB equation in 3-D layered media, picture from [79]. The notations Φ in [79] refers to the field $u$ in this thesis.

are decomposed into the free-space part and four types of reaction field components. Using the spectral form of the layered media Green’s functions and the Funk–Hecke identities, we developed multipole expansion of $O(p^2)$ terms for the far field of the reaction field components, which can be associated with polarization sources at specific locations for each type of the reaction field components. Multipole-to-local translation operators are also developed for the reaction fields. As a result, the traditional ME-based FMM can be applied to both the free space part and the reaction field components, once the polarization sources are used together with the original sources. Due to the separation of the polarization and the original source charges by a material interface, the computational cost from the reaction field parts is only a fraction of that of the FMM for the free space part. For a given layered structure, some one-time pre-computations of numerical integrals for the translation operator will be required, otherwise, computing the wave interactions of many sources in layered media basically costs the same as that for the wave interaction in the free space.

For the future work, we will carry out error estimates of the FMM for the Helmholtz equation and the linearized PB equation in 3-D layered media, which require an error analysis for the newly proposed MEs and M2L operators for the reaction components, following our
<table>
<thead>
<tr>
<th>cores</th>
<th>number of particles</th>
<th>(\ell = 0)</th>
<th>(\ell')</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(u^{(f)}_t)</td>
<td>(u^{(f')}_t)</td>
</tr>
<tr>
<td>1</td>
<td>618256</td>
<td>40.36</td>
<td>17.06</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>86.72</td>
<td>62.47</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>269.05</td>
<td>74.93</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>292.42</td>
<td>81.47</td>
</tr>
<tr>
<td>6</td>
<td>618256</td>
<td>7.653</td>
<td>3.613</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>16.29</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>50.72</td>
<td>15.52</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>54.85</td>
<td>17.27</td>
</tr>
<tr>
<td>36</td>
<td>618256</td>
<td>2.042</td>
<td>1.639</td>
</tr>
<tr>
<td></td>
<td>1128556</td>
<td>4.308</td>
<td>4.459</td>
</tr>
<tr>
<td></td>
<td>1862568</td>
<td>14.94</td>
<td>6.104</td>
</tr>
<tr>
<td></td>
<td>2861288</td>
<td>15.21</td>
<td>7.673</td>
</tr>
</tbody>
</table>

Table 3.2: Comparison of CPU time (in seconds) with multiple CPU cores for FMM for linearized PB equation in 3-D layered media, the truncation index \(P = 5\), data from [79].

results in the 2-D case in Chapter 2.
Chapter 4
FMM for Laplace’s equation in 3-D layered media

The content in this chapter has been included in the following published journal paper


and the following submitted paper under the name


in collaboration with Bo Wang and Wei Cai.

4.1. Introduction

The fast multipole method for Laplace’s equation in the homogeneous free space was proposed by L. Greengard and V. Rokhlin in [35]. The algorithm has been used in the development of FastCap (cf. [58]) to accelerate the matrix solver for discretized integral equations, where the free-space Green’s function of the Laplace’s equation is involved. However, when using the free-space Green’s functions to treat the dielectric material interfaces in the IC design, unknowns representing the polarization charges from the dielectric inhomogeneities have to be introduced over the infinite material interfaces, thus creating unnecessary unknowns and contributing to larger linear systems. These extra unknowns over material interfaces can be avoided by using the LMGF of the Laplace’s equation in the formulation of the integral equations. To find fast algorithms to solve the discretized linear system, image
charges are used to approximate the LMGF [22, 2, 3], converting the reaction potential to the free-space Coulomb potential from the charges and their images, thus, the free space FMM can be used [44, 37, 32]. Apparently, such approaches are limited to the ability of finding image charge approximation for the layered media Green’s function. Unfortunately, finding such an image approximation can be challenging, if not impossible, for general layered media.

In this chapter, we develop the FMM for charge interactions in layered media, which can be then used in fast iterative solvers for Laplace’s equations through integral equation methods with a LMGF. We first study the LMGF in Section 4.2, then the far-field expansions are derived for the reaction field components in Section 4.3. A FMM framework similar to the ones from previous chapters is given in Section 4.4. With a few practical techniques discussed in Section 4.5, the numerical tests in Section 4.6 verifies the efficiency and the accuracy of the proposed FMM. Finally, the exponential convergence of the far-field expansion is proven in the ending sections.

4.2. The Green’s function in layered media

The LMGF of the Laplace’s equation can be treated as the limiting case of the linearized Poisson–Boltzmann equation discussed in Chapter 3 as \( k_\ell \to 0 \). Here we repeat the necessary steps to derive its representation in the frequency domain and the reaction field decomposition.

The LMGF satisfies the differential equation

\[
\nabla^2 u (\mathbf{r}; \mathbf{r}') = -\delta (\mathbf{r} - \mathbf{r}'),
\]

(4.1)

where the partial derivatives are taken on the target coordinates \( \mathbf{r} = (x, y, z) \), \( \delta \) is the Dirac delta function, with the interface conditions

\[
[au] = 0, \quad [b \partial u / \partial \mathbf{n}] = 0, \quad z = d_0, \cdots, d_{L-1},
\]

(4.2)
where $a, b > 0$ are piecewisely constant parameters for each layer. Typically, when $u(\mathbf{r}; \mathbf{r}')$ refers to the electrostatic potential, we adopt $a = 1$ and $b = \varepsilon$, where $\varepsilon = \varepsilon_\ell$ is the dielectric constant in layer $\ell$. In addition, the LMGF is restricted by a decaying condition

$$u(\mathbf{r}; \mathbf{r}') \to 0 \quad \text{as} \quad r = |\mathbf{r}| \to \infty. \quad (4.3)$$

*Remark 4.1.* Other works related with the Laplace’s equation often have the LMGF defined differing by a factor from the one in (4.1), such as in [59, 89]. They can be equivalently treated both theoretically and practically.

### 4.2.1. Reaction field decomposition of the LMGF in the frequency domain

Like in Definition 2.1 and Definition 3.2, we exclude the free-space Green’s function to form the reaction field for the LMGF of the Laplace’s equation.

**Definition 4.2 (Reaction field).** The reaction field of the LMGF is defined by

$$u^r(\mathbf{r}; \mathbf{r}') = u(\mathbf{r}; \mathbf{r}') - \delta_{\ell,\ell'} G^f(\mathbf{r}; \mathbf{r}'), \quad (4.4)$$

where

$$G^f(\mathbf{r}; \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \quad (4.5)$$

is the free-space Green’s function of the Laplace’s equation.

It is obvious that the reaction field $u^r$ also satisfies the decaying condition (4.3).

To derive the reaction field decomposition like in (2.20) and (3.13), we take the 2-D Fourier transform (3.6). In the frequency domain, the reaction field satisfies the ordinary differential equation

$$\frac{\partial^2 u^r(k_x, k_y, z; \mathbf{r}')}{\partial z^2} = 0, \quad (4.6)$$
while the interface conditions are shown as

\[
\left[ a(\hat{u}^r + \hat{G}^f) \right] = 0, \quad \left[ b \left( \frac{\partial \hat{u}^r}{\partial \mathbf{n}} + \frac{\partial \hat{G}^f}{\partial \mathbf{n}} \right) \right] = 0, \quad z = d_0, \cdots, d_{L-1},
\]  

(4.7)

where

\[
\hat{G}^f(k_x, k_y, z; r') = \frac{e^{-k_\rho|z-z'|}}{2k_\rho}.
\]  

(4.8)

Note that \( \hat{G}^f(k_x, k_y, z; r') \) is singular at \( k_x = k_y = 0 \). However, when switching to the polar coordinates \( dk_x dk_y = k_\rho dk_\rho d\alpha \), the singularity at \( k_\rho = 0 \) will vanish. The same property also applies to the succeeding derivation of the LMGF in the frequency domain.

Most properties discussed in Section 2.2 and Section 3.2 for the LMGF of the Helmholtz equation also hold for the Laplace’s equation. Here we state the corresponding result for the reaction field decomposition.

**Definition 4.3 (Reaction field decomposition).** The reaction field decomposition of \( \hat{u}^r \) in the frequency domain is defined by

\[
u^r(r; r') = \frac{1}{4\pi^2} \int \int_{\mathbb{R}^2} e^{ik_x(x-x') + ik_y(y-y')} \sum_{*,* \in \{\uparrow, \downarrow\}} k_\rho^{-1} \sigma_{\ell \ell'}^{*\,*}(k_x, k_y; x', y') e^{-k_\rho(\tau^\uparrow(z-d^\uparrow_\ell) + \tau^\downarrow(z'-d^\downarrow_\ell'))} dk_x dk_y
\]

\[= \sum_{*,* \in \{\uparrow, \downarrow\}} u^r_{\ell \ell'}(r; r'), \]  

(4.9)

where

\[
\tau^\uparrow = +1, \quad \tau^\downarrow = -1,
\]  

(4.10)

d^*_\ell is the same as in (2.16),

\[
d^1_\ell = d_\ell, \quad 0 \leq \ell \leq L - 1, \quad d^1_L = -\infty,
\]

(4.11)

\[
d^1_\ell = d_{\ell-1}, \quad 1 \leq \ell \leq L, \quad d^1_0 = +\infty.
\]  

(4.12)
Each reaction field component

\[
    u_{\ell'\ell}^*(r; r') = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell}^*(r, r'; k_\rho) k_\rho^{-1} \sigma_{\ell'\ell}^{**}(k_x, k_y; x', y') dk_x dk_y, \tag{4.13}
\]

where \( \sigma_{\ell'\ell}^{**}(k_x, k_y; x', y') \) are the reflection/transmission coefficients of the layered media in the frequency domain, and

\[
    \mathcal{E}_{\ell'\ell}^*(r, r'; k_\rho) = e^{i k_x (x-x') + i k_y (y-y') - k_\rho (z - d_\ell^* + z' - d_{\ell'}^*)}. \tag{4.14}
\]

If compared to the definitions in previous chapters, the reflection/transmission coefficients \( \sigma_{\ell'\ell}^{**}(k_x, k_y; x', y') \) defined here has one less \( k_\rho \) factor, which is for the convenience of developing further theories, and will be compensated when switching to the polar coordinates of \((k_x, k_y)\). These coefficients are in fact independent from \( x' \) and \( y' \), and are rotationally symmetric on the \((k_x, k_y)\)-plane, which are direct corollaries of the recursive formulas in the coming section, where the linear systems whose unknowns consist of \( \sigma_{\ell'\ell}^{**}(k_x, k_y; x', y') \) do not depend on \( x' \) and \( y' \), and are rotationally invariant on the \((k_x, k_y)\)-plane. Thus, we can simply denote

\[
    \sigma_{\ell'\ell}^{**}(k_x, k_y; x', y') = \sigma(k_\rho) \tag{4.15}
\]

in the above reaction field decomposition.

Again, there are prohibited field propagation directions

\[
    \sigma_0^\downarrow = \sigma_0^\uparrow = \sigma_{L0}^\downarrow = \sigma_{L0}^\uparrow = 0, \quad u_{00}^\downarrow = u_{00}^\uparrow = u_{L0}^\downarrow = u_{L0}^\uparrow = 0, \quad *, * \in \{\uparrow, \downarrow\}. \tag{4.16}
\]

4.2.2. Recursive formula for the reflection/transmission coefficients

A recursive algorithm can be developed to compute the reaction field given \( k_\rho \) like in Algorithm 3.1. Following the same derivation, the algorithm is given below.
Define $\vartheta = 1/2$,

$$
e_\ell = e^{-k_\rho(d_{\ell-1} - d_\ell)}, \quad 1 \leq \ell \leq L - 1, \quad e_0 = e_L = 1,
$$

(4.17)

$$
C^{(\ell)} = \prod_{j=0}^{\ell-1} \frac{1}{2e_j}, \quad 0 \leq \ell \leq L.
$$

(4.18)

$$
\gamma_\ell^\pm = \frac{a_\ell}{a_{\ell-1}} \pm \frac{b_\ell}{b_{\ell-1}}, \quad 1 \leq \ell \leq L,
$$

(4.19)

$$
\mathcal{S}^{(-1)} = \mathcal{S}^{(L+1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathcal{S}^{(\ell)} = \frac{1}{2e_\ell} \begin{bmatrix} e_\ell & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (a_\ell)^{-1} & (b_\ell)^{-1} \\ (a_\ell)^{-1} & -(b_\ell)^{-1} \end{bmatrix}, \quad 0 \leq \ell \leq L,
$$

(4.20)

the transition matrices

$$
\mathbb{T}^{\ell-1,\ell} = (2e_{\ell-1})^{-1} \tilde{\mathbb{T}}^{\ell-1,\ell}, \quad 1 \leq \ell \leq L
$$

(4.21)

where

$$
\tilde{\mathbb{T}}^{\ell-1,\ell} = \begin{bmatrix} e_{\ell-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_\ell^+ & \gamma_\ell^- \\ \gamma_\ell^- & \gamma_\ell^+ \end{bmatrix} \begin{bmatrix} e_\ell & 0 \\ 0 & 1 \end{bmatrix},
$$

(4.22)

and

$$
\tilde{A}^{(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \tilde{A}^{(\ell)} = \tilde{\mathbb{T}}^{0,1} \cdots \tilde{\mathbb{T}}^{\ell-1,\ell}, \quad 1 \leq \ell \leq L.
$$

(4.23)

Algorithm 4.1 solves the reflection/transmission coefficients with the above notations.

Unlike the Helmholtz equation, the reflection/transmission coefficient functions $\sigma_{\ell\ell'}^{*}(k_\rho)$ have no poles in the right half complex plane. This can be interpreted as a corollary of
Algorithm 4.1: Recursively solving $\sigma_{L\ell'}^\ast$.  

\[ \sigma_{L\ell'}^\uparrow = \sigma_{L\ell'}^\downarrow = 0; \]

for $\ell = L \to 1$ do

\[ \sigma_{L\ell'}^\uparrow = -\frac{A_{21}(L)}{A_{22}(L)} \sigma_{L\ell'}^\uparrow - 1_{\ell' < \ell} \frac{\partial}{\partial C(\ell')} \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] A(\ell') S(\ell') \left[ \begin{array}{c} -a_{\ell'} \\ b_{\ell'} \end{array} \right]; \quad (4.24) \]

\[ \sigma_{L\ell'}^\downarrow = -\frac{A_{21}(L)}{A_{22}(L)} \sigma_{L\ell'}^\downarrow - 1_{0 < \ell' \leq \ell} \frac{\partial}{\partial C(\ell')} \left[ \begin{array}{c} 0 \\ 1 \end{array} \right] A(\ell'-1) \left( 2e_{\ell'-1} S(\ell'-1) \right) \left[ \begin{array}{c} a_{\ell'} \\ b_{\ell'} \end{array} \right]; \quad (4.25) \]

\[ \sigma_{L-1\ell'}^\uparrow = T_{11}^{\ell-1,\ell} \sigma_{L\ell'}^\uparrow + T_{12}^{\ell-1,\ell} \sigma_{L\ell'}^\downarrow + \delta_{\ell,\ell'+1} \left( -S_{11}(\ell') a_{\ell'} + S_{12}(\ell') b_{\ell'} \right); \quad (4.26) \]

\[ \sigma_{L-1\ell'}^\downarrow = T_{11}^{\ell-1,\ell} \sigma_{L\ell'}^\uparrow + T_{12}^{\ell-1,\ell} \sigma_{L\ell'}^\downarrow + \delta_{\ell,\ell'} \left( S_{11}(\ell') a_{\ell'} + S_{12}(\ell') b_{\ell'} \right); \quad (4.27) \]

end

$\sigma_{0\ell'}^\uparrow = \sigma_{0\ell'}^\downarrow = 0;$

Theorem 2.5 for the 2-D Helmholtz equation.

Theorem 4.4 (Holomorphic in the right half complex plane). For any $k_\rho \in \mathbb{C}$ satisfying $\Re k_\rho \geq 0$, Algorithm 4.1 uniquely solves $\sigma_{L\ell'}^\ast(k_\rho)$ for all reaction field components, hence $\sigma_{L\ell'}^\ast(k_\rho)$ is holomorphic in the right half plane.

For the asymptotic estimates, the following theorem is a corollary of Theorem 2.9 for the 2-D Helmholtz equation.

Theorem 4.5 (Constant bound). There exists $C > 0$ such that for any $k_\rho \in \mathbb{C}$, $\Re k_\rho \geq 0$,

\[ |\sigma_{L\ell'}^\ast(k_\rho)| \leq C. \quad (4.28) \]

Proof. Recall the proof of Theorem 2.5, we have $|e_\ell| \leq 1$, and the fact that

\[ A_{22}^2 - A_{21}^2 \geq \prod_{j=1}^{\ell} (|\gamma_j^+|^2 - |\gamma_j^-|^2) = 4 \ell \frac{a_\ell b_\ell}{a_0 b_0} > 0 \quad (4.29) \]

by induction. On the other hand, upper bounds of entries of $A(\ell)$ can be found from the
definition with a forward recursion. Then, Algorithm 4.1 can be used to give estimates of all the reflection/transmission coefficients. Namely, in (4.24)–(4.27), we have estimates

\[
\left| \frac{C^{(\ell')}}{C^{(\ell)}} \begin{bmatrix} 0 & 1 \end{bmatrix} A^{(\ell')} S^{(\ell')} \begin{bmatrix} -a_{\ell'} \\ b_{\ell'} \end{bmatrix} \right| \leq 2|A_{22}^{(\ell')}| \prod_{j=\ell'+1}^{\ell-1} |e_j|,
\]

\[
\left| \frac{C^{(\ell')}}{C^{(\ell)}} \begin{bmatrix} 0 & 1 \end{bmatrix} A^{(\ell'-1)} \left(2e_{\ell'-1}S^{(\ell'-1)}\right) \begin{bmatrix} a_{\ell'} \\ b_{\ell'} \end{bmatrix} \right| \leq \left(|A_{21}^{(\ell'-1)} \gamma_{\ell'}^+ e_{\ell'-1}| + |A_{22}^{(\ell'-1)} \gamma_{\ell'}^-|\right) \prod_{j=\ell'}^{\ell-1} |e_j|,
\]

\[
|T_{11}^{\ell-1,\ell'}| = \frac{|\gamma_{\ell'}^+| e_{\ell}}{2}, \quad |T_{12}^{\ell-1,\ell'}| = \frac{|\gamma_{\ell'}^-|}{2}, \quad -S_{11}^{(\ell')} a_{\ell'} + S_{12}^{(\ell')} b_{\ell'} = 0, \quad |S_{11}^{(\ell'-1)} a_{\ell'} + S_{12}^{(\ell'-1)} b_{\ell'}| = \frac{|\gamma_{\ell'}^+|}{2}.
\]

These estimates show the existence of a global upper bound by recursion.

4.3. Far-field expansions of the reaction field and the FMM framework

In this section, we derive a series expansion in the frequency domain modified from the generalized Funk–Hecke formula for the 3-D Helmholtz equation and the linearized PB equation, since in Lemma 3.5 and in Theorem 3.8 we excluded the case when the wave number \( k \equiv 0 \). Then, the far-field expansions for ME, LE and M2L will be proposed accordingly.

We begin with a brief review of the ME for the free-space Green’s function. With a source center \( r_c \), the free-space Green’s function has Taylor series

\[
G^f(r; r') = \frac{1}{4\pi |r - r'|} = \frac{1}{4\pi |r - r_c|} \sum_{n=0}^{\infty} P_n(\cos \gamma_c) \left( \frac{|r' - r_c|}{|r - r_c|} \right)^n
\]

provided the far-field condition \(|r' - r_c| < |r - r_c|\) holds, where \( \gamma_c \) is the angle between
vectors \( \mathbf{r} - \mathbf{r}_c \) and \( \mathbf{r}' - \mathbf{r}_c \). The Taylor series (4.30) has exponential convergence

\[
\left| G_f(\mathbf{r}; \mathbf{r}') - \frac{1}{4\pi |\mathbf{r} - \mathbf{r}_c|} \sum_{n=0}^{P} P_n(\cos \gamma_c) \left( \frac{|\mathbf{r}' - \mathbf{r}_c|}{|\mathbf{r} - \mathbf{r}_c|} \right)^n \right| \leq \frac{(4\pi)^{-1}}{|\mathbf{r} - \mathbf{r}_c| - |\mathbf{r}' - \mathbf{r}_c|} \left( \frac{|\mathbf{r}' - \mathbf{r}_c|}{|\mathbf{r} - \mathbf{r}_c|} \right)^{P+1}.
\]

Using the Legendre addition theorem (3.42), \( P_n(\cos \gamma_c) \) is further expanded into a series that separates the angular information from the target and from the source. Specifically, let \((r_c, \theta_c, \phi_c)\) and \((r'_c, \theta'_c, \phi'_c)\) be the spherical coordinate triples of \( \mathbf{r} - \mathbf{r}_c \) and \( \mathbf{r}' - \mathbf{r}_c \), respectively, then

\[
P_n(\cos \gamma_c) = \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y^m_n(\theta'_c, \phi'_c) Y^m_n(\theta_c, \phi_c).
\]

In total, the ME for the free-space Green’s function is given by

\[
G_f(\mathbf{r}; \mathbf{r}') \approx \sum_{n=0}^{P} \sum_{m=-n}^{n} \frac{1}{2n + 1} (r'_c)^n Y^m_n(\theta'_c, \phi'_c) \cdot (r_c)^{-n-1} Y^m_n(\theta_c, \phi_c),
\]

where \((r_c)^{-n-1} Y^m_n(\theta_c, \phi_c)\) serves as the ME basis.

In the following discussion, the far-field expansions of the LMGF will be proposed with a similar structure that can indeed treated as an extension from the free-space case.

4.3.1. The Modified Funk–Hecke formula

Let \( \mathbf{r} = (x, y, z) \in \mathbb{R}^3 \) with spherical coordinates \((r, \theta, \phi)\). Recall Lemma 3.4 where we have the generalized PWE for the Helmholtz case with any wave number \( k > 0 \),

\[
e^{ik \cdot \mathbf{r}} = \sum_{n=0}^{\infty} (2n + 1)i^n j_n(kr) P_n(\hat{k} \cdot \hat{r}),
\]

where \( \hat{k} = k^{-1}(ik_x, ik_y, ik_z) \), \( \hat{r} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \). Consider the limit as \( k \to 0+0 \).

For the asymptotic behavior of the spherical Bessel function \( j_n \),

\[
j_n(kr) \sim k^n r^n (2n + 1)!!.
\]
For the $P_n(\hat{k} \cdot \hat{r})$ part, write $P_n(x) = \frac{(2n-1)!!}{n!} x^n + q_{n-1}(x)$, where $q_{n-1}(x)$ is a polynomial consisting of the non-leading terms of $P_n(x)$ with order up to $n - 1$, we have

$$k^n P_n(\hat{k} \cdot \hat{r}) \sim \frac{(2n-1)!!}{n!} (\hat{k} \cdot \hat{r})^n. \tag{4.36}$$

In total, we get a modified version of Lemma 3.4 as

$$e^{k_\rho k_0 \cdot \hat{r}} = \sum_{n=0}^{\infty} \frac{\nu^n (k_\rho k_0 \cdot \hat{r})^n}{n!}, \tag{4.37}$$

where

$$k_0 = (i \cos \alpha, i \sin \alpha, -1). \tag{4.38}$$

Although (4.37) is merely the Taylor series, it shows a hint to derive the further separation of the angular parts of $k$ and $r$. Recall in Lemma 3.7,

$$P_n(\hat{k} \cdot \hat{r}) = \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y^m_n(\theta, \phi) \widehat{P}_n^m \left( \frac{k_z}{k} \right) e^{i m \alpha} \quad \text{or} \quad \frac{4\pi}{2n + 1} \sum_{m=-n}^{n} Y^m_n(\theta, \phi) \widehat{P}_n^m \left( \frac{k_z}{k} \right) e^{-i m \alpha}. \tag{4.39}$$

Since each associated Legendre function satisfies

$$\widehat{P}_n^m \left( \frac{k_z}{k} \right) = \frac{1}{2^n n!} \sqrt{\frac{2n + 1}{4\pi} \frac{(n - m)!}{(n + m)!}} \left( \frac{k_\rho}{k} \right)^m \left[ \frac{d^{n+m}(x^2 - 1)^n}{dx^{n+m}} \right] \bigg|_{x = k_z}, \tag{4.40}$$

similar as in (4.36), as $k \to 0 + 0$,

$$k^n \widehat{P}_n^m \left( \frac{k_z}{k} \right) \sim \frac{1}{2^n n!} \sqrt{\frac{2n + 1}{4\pi} \frac{(n - m)!}{(n + m)!}} \frac{(2n)!}{(n - m)!} k_\rho \cdot (2n)! (ik_\rho)^{n-m}. \tag{4.41}$$

Substituting into (4.39) and (4.37), the expansion of the $(k_\rho k_0 \cdot \hat{r})^n / n!$ factor in (4.37) takes
the form

\[
\frac{(k_p k_0 \cdot \hat{r})^n}{n!} = \sum_{m=-n}^{n} \frac{1}{2n+1} Y_{n}^{m}(\theta, \phi)c_{nm} i^{2n-m} k_{\rho}^{n} e^{i m \alpha}
\]

\[
= \sum_{m=-n}^{n} \frac{1}{2n+1} Y_{n}^{m}(\theta, \phi)c_{nm} i^{2n-m} k_{\rho}^{n} e^{-i m \alpha},
\]

(4.42)

where

\[
c_{nm} = \sqrt{\frac{4\pi (2n+1)}{(n+m)! (n-m)!}}.
\]

(4.43)

It can be seen as a modified Legendre addition theorem. The modified Funk–Hecke formula is then given by

\[
e^{k_p k_0 \cdot r} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{r^n}{2n+1} Y_{n}^{m}(\theta, \phi)c_{nm} i^{2n-m} k_{\rho}^{n} e^{i m \alpha}
\]

\[
= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{r^n}{2n+1} Y_{n}^{m}(\theta, \phi)c_{nm} i^{2n-m} k_{\rho}^{n} e^{-i m \alpha},
\]

(4.45)

4.3.2. The far-field expansions and translations

With (4.44) and (4.45), the far-field expansions are proposed as follows.

For the ME, we split the exponential term (4.14) as

\[
E_{\ell\ell}^{\star}(r, r', k_{\rho}) = E_{\ell\ell}^{\star}(r, r_c, k_{\rho}) \cdot \exp(i k_x (x_c - x') + i k_y (y_c - y') - k_{\rho}\tau^*(z' - z_c)),
\]

(4.46)

where the source center \( r_c = (x_c, y_c, z_c) \) should be on the same side of the interface \( z = d'_{\ell \ell} \).

Let \((r'_c, \theta'_c, \phi'_c)\) be the spherical coordinates of \( r' - r_c \). Then, the triple \((x_c - x', y_c - y', \tau^*(z' - z_c))\)
\( \mathbf{z}_c \) has the spherical coordinates \((r'_c, \frac{\pi}{2} + \tau^* (\theta'_c - \frac{\pi}{2}), \phi'_c + \pi)\). By using (4.44),

\[
\exp (i k_x (x_c - x') + i k_y (y_c - y') - k_\rho \tau^* (z' - z_c))
\]

\[
= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{(r'_c)^n}{2n+1} Y_n^m \left( \frac{\pi}{2} + \tau^* \left( \theta'_c - \frac{\pi}{2} \right), \phi'_c + \pi \right) c_{nm} i^{2n-m} k_\rho^n e^{i \sigma}
\]

(4.47)

Let

\[
M_{nm}(r'_c, r_c) = \frac{(r'_c)^n}{2n+1} Y_n^m(\theta'_c, \phi'_c)
\]

(4.48)

which takes the same form as the free-space ME coefficient in (4.33) (c.f. [77, 78]), the ME of the reaction field component \( u^{*\star}_{\ell\ell}(r; r') \) is given by

\[
u^{*\star}_{\ell\ell}(r; r') \]

\[
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathbf{E}^{*\star}_{\ell\ell}(r, r', k_\rho) k_\rho^{-1} \sigma^{*\star}_{\ell\ell}(k_\rho) dk_x dk_y
\]

\[
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathbf{E}^{*\star}_{\ell\ell}(r, r_c, k_\rho) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} M_{nm}(r'_c, r_c) (-1)^m (\tau^*)^{m+n} c_{nm} i^{2n-m} k_\rho^{n-1} \sigma^{*\star}_{\ell\ell}(k_\rho) \]

\[
\approx \sum_{n=0}^{P} \sum_{m=-n}^{n} M_{nm}(r'_c, r_c) F^{*\star}_{nm}(r, r_c),
\]

(4.49)

where

\[
F^{*\star}_{nm}(r, r_c) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathbf{E}^{*\star}_{\ell\ell}(r, r_c, k_\rho) (\tau^*)^{m+n} c_{nm} i^{2n+m} k_\rho^{n-1} e^{i \sigma} \]

\[
= \sum_{j=1}^{M} q_j M_{nm}(r'_j, r_c), \quad 0 \leq n \leq P, \quad -n \leq m \leq n.
\]

(4.51)
In fact, if we treat the free-space problem as a two-layer problem with trivial interface conditions, the ME basis function \( F_{nm}(\mathbf{r}, \mathbf{r}_c) \) defined above will serve as an integral representation of the corresponding free-space ME basis function \( (\mathbf{r}_c)^{n-1}Y_n^m(\theta_c, \phi_c) \) in (4.33). This can be understood by using \( M_{nm} \) as a basis in a reversed situation, where we have the uniqueness of \( F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c) \) due to the orthogonality of spherical harmonic functions.

To derive the M2M translation, suppose the source center is shifted from \( \mathbf{r}_c \) to \( \tilde{\mathbf{r}}_c \). Let \((\tilde{r}, \tilde{\theta}, \tilde{\phi})\) be the spherical coordinates of \( \tilde{\mathbf{r}}_c - \mathbf{r}_c \). By using (4.44),

\[
\exp(i k_x (\tilde{x}_c - x_c) + i k_y (\tilde{y}_c - y_c) - k_\rho \tau^*(z_c - \tilde{z}_c))
= \sum_{n' = 0}^{\infty} \sum_{m' = -n'}^{n'} \frac{(\tilde{r})^{n'}}{2n' + 1} Y_{n'}^m(\tilde{\theta}, \tilde{\phi}) (\tau^*)^n m' c_{n'n'm'} k_\rho e^{i m' \alpha} (4.52)
\]

\[
= \sum_{n' = 0}^{\infty} \sum_{m' = -n'}^{n'} \frac{(\tilde{r})^{n'}}{2n' + 1} Y_{n'}^m(\tilde{\theta}, \tilde{\phi}) (-\tau^*)^n m' c_{n'n'm'} k_\rho e^{i m' \alpha},
\]

each ME basis function \( F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c) \), with the assumption that the order of integration and summation over \( n' \) can be exchanged, can be expanded as

\[
F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c)
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell''}(\mathbf{r}, \tilde{\mathbf{r}}_c, k_\rho) \sum_{n' = 0}^{\infty} \sum_{m' = -n'}^{n'} \frac{(\tilde{r})^{n'}}{2n' + 1} Y_{n'}^m(\tilde{\theta}, \tilde{\phi}) (-\tau^*)^n m' c_{n'n'm'} k_\rho e^{i m' \alpha} (4.53)
\]

\[
(\tau^*)^{m+n} c_{nm} 2^{n+1} k_\rho n-1 e^{i m \alpha} \sigma_{\ell'\ell''}(k_\rho) dk_x dk_y
= \sum_{n' = 0}^{\infty} \sum_{m' = -n'}^{n'} T_{nm;n'm'}^{M2M} F_{n+n',m+m'}^{**}(\mathbf{r}, \tilde{\mathbf{r}}_c),
\]

where

\[
T_{nm;n'm'}^{M2M} = \begin{cases} 
\frac{(\tilde{r})^{n'}}{2n' + 1} \frac{Y_{n'}^m(\tilde{\theta}, \tilde{\phi}) (-1)^{n'+m'} c_{nm} c_{n'm'} k_\rho n-1 e^{i m \alpha}}{c_{n+n',m+m'}}, & |m| \leq n, \ |m'| \leq n', \\
0, & \text{else.}
\end{cases} (4.54)
\]
Then, assuming the order of infinite summations can be exchanged,

\[
\begin{align*}
    u_{\ell\ell'}^*(r; r') &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} M_{nm}(r', r_c) F_{nm}^{**}(r, r_c) \\
    &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} M_{nm}(r', r_c) \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} T_{nm,n'-n',m'=-m'}^{M2M} F_{n+n',m+m'}^{**}(r, \tilde{r}_c) \\
    &= \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{M}_{\nu\mu} F_{\nu\mu}^{**}(r, \tilde{r}_c),
\end{align*}
\]

(4.55)

where

\[
\tilde{M}_{\nu\mu} = \sum_{n=0}^{\nu} \sum_{m=-n}^{n} M_{nm}(r', r_c) T_{nm,\nu-n,\mu-m}^{M2M}
\]

(4.56)

are the results of the M2M translation. Following the discussion by the end of the derivation of ME, (4.56) holds for the free-space case as well, which implies

\[
M_{\nu\mu}(r, \tilde{r}_c) = \tilde{M}_{\nu\mu}
\]

(4.57)

due to the orthogonality of the spherical harmonic functions from the free-space ME basis functions. Therefore, the M2M is lossless with a finite-term linear transform. In fact, the M2M derived above is equivalent to the free-space version, see [77, 78, 35].

To derive the LE, we split the exponential term (4.14) as

\[
E_{\ell\ell'}^{**}(r, r', k_\rho) = E_{\ell\ell'}^{**}(r_c, r', k_\rho) \cdot \exp \left( ik_x (x - x_c) + ik_y (y - y_c) - k_\rho \tau^* (z - z_c) \right)
\]

(4.58)

where the target center \( r_c^l = (x_c^l, y_c^l, z_c^l) \) should be on the same side of the interface \( z = d_c^l \). Let \((r^l, \theta^l, \phi^l)\) be the spherical coordinates of \( r - r_c^l \). Then, the triple \((x - x_c^l, y - y_c^l, \tau^* (z - z_c^l))\)
has the spherical coordinates \((r^l, \frac{\pi}{2} + \tau^* (\theta^l - \frac{\pi}{2}), \phi^l)\). By using (4.45),

\[
\exp \left( i k_x (x - x^l_c) + i k_y (y - y^l_c) - k_\rho \tau^* (z - z^l_c) \right)
= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (r^l)^n Y^m_n \left( \frac{\pi}{2} + \tau^* \left( \theta^l - \frac{\pi}{2} \right), \phi^l \right) \frac{c_{nm} k_\rho^{n+1}}{2n+1} \frac{1}{2} \phi_{nm}(\theta^l, \phi^l) e^{-i \frac{m}{2} \pi} e^{-i \alpha}.
\]

(4.59)

Let

\[
K_{nm}(r, r^l_c) = (r^l)^n Y^m_n \left( \theta^l, \phi^l \right),
\]

(4.60)

the LE of the reaction field component \(u_{\ell\ell'}^{**}(r; r')\) is given by

\[
u_{\ell\ell'}^{**}(r; r') = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^{**}(r, r', k_\rho) k_\rho^{-1} \sigma_{\ell\ell'}^{**}(k_\rho) dk_x dk_y
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^{**}(r^l_c, r, k_\rho) \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (r^l)^n Y^m_n \left( \theta^l, \phi^l \right) \frac{c_{nm} k_\rho^{n+1}}{2n+1} \frac{1}{2} \phi_{nm}(\theta^l, \phi^l) e^{-i \frac{m}{2} \pi} e^{-i \alpha} \sigma_{\ell\ell'}^{**}(k_\rho) dk_x dk_y
≈ \sum_{n=0}^{P} \sum_{m=-n}^{n} K_{nm}(r, r^l_c) L_{nm}^{**}(r, r^l_c),
\]

(4.61)

where

\[
L_{nm}^{**}(r^l, r^l_c) = \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^{**}(r^l_c, r^l, k_\rho) \frac{c_{nm} k_\rho^{n+1}}{2n+1} \frac{1}{2} \phi_{nm}(\theta^l, \phi^l) e^{-i \frac{m}{2} \pi} e^{-i \alpha} \sigma_{\ell\ell'}^{**}(k_\rho) dk_x dk_y
\]

(4.62)

are the LE coefficients, and \(K_{nm}\) are the LE basis functions. The LE basis functions for the LMGF are identical to the ones for the free-space Green’s functions, see [77, 78, 35].
To derive the L2L translation, suppose the target center is shifted from \( r_c^l \) to \( \tilde{r}_c^l \). Let \((\vec{r}^l, \vec{\theta}^l, \vec{\phi}^l)\) be the spherical coordinates of \( \tilde{r}_c^l - r_c^l \). Similar as the M2M, by using (4.45),

\[
\exp \left( i k_x (\vec{x}_c^l - x_c^l) + i k_y (\vec{y}_c^l - y_c^l) - k_\rho \tau^* (\vec{z}_c^l - z_c^l) \right)
= \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} (\vec{r}^l)^n m' \left( \vec{\theta}^l, \vec{\phi}^l \right) \frac{c_{n'm'}}{2 n' + 1} (\tau^*)^{n'+m'} 12 n'-m' k_\rho e^{-i m' \alpha}.
\]

(4.63)

With the assumption that the order of integration and summation can be interchanged, each LE coefficient has the expansion

\[
L_{nm}^{**}(r', \tilde{r}_c^l) = \frac{1}{4\pi^2} \int \int_{\mathbb{R}^2} \mathcal{E}_{\ell'\ell}(r_c^l, r', k_\rho) \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} (\vec{r}^l)^n m' \left( \vec{\theta}^l, \vec{\phi}^l \right) \frac{c_{n'm'}}{2 n' + 1} (\tau^*)^{n'+m'} 12 n'-m' k_\rho e^{-i m' \alpha}.
\]

\[
\sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} T_{nm;n'm'}^{L2L} L_{n+n',m+m'}^{**}(r', r_c^l),
\]

(4.64)

where

\[
T_{nm;n'm'}^{L2L} = \begin{cases} (\vec{r}^l)^n m' \left( \vec{\theta}^l, \vec{\phi}^l \right) (2n + 2n' + 1) c_{nm} c_{n'm'} \frac{c_{n+m} c_{n+m'}}{(2n+1)(2n'+1)} & \text{if } |m| \leq n, |m'| \leq n', \\
0, & \text{else}.
\end{cases}
\]

(4.65)

Then, assuming the order of infinite summations can be exchanged,

\[
u_{\ell'\ell}(r; r') = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} K_{nm}(r, \tilde{r}_c^l) L_{nm}^{**}(r', \tilde{r}_c^l)
= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} K_{nm}(r, \tilde{r}_c^l) \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} T_{nm;n'm'}^{L2L} L_{n+n',m+m'}^{**}(r', r_c^l)
= \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \tilde{K}_{\mu\nu} L_{\mu\nu}^{**}(r', r_c^l),
\]

(4.66)

(4.67)

(4.68)
where
\[
\tilde{K}_{\nu\mu} = \sum_{n=0}^{\nu} \sum_{m=-n}^{n} K_{nm}(\mathbf{r}, \mathbf{r}'_c) T_{nm;\nu-n,\mu-m}^{L2L}.
\] (4.69)

Following the discussion from the M2M part, we can see
\[
\tilde{K}_{\nu\mu} = K_{\nu\mu}(\mathbf{r}, \mathbf{r}'_c).
\] (4.70)

Therefore, the L2L is also lossless with a finite-term linear transform. Write the L2L in terms of the LE coefficients, we have
\[
\sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} K_{\nu\mu}(\mathbf{r}, \mathbf{r}'_c)L_{\nu\mu}^{**}(\mathbf{r}', \mathbf{r}'_c)
\]
\[
= \sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} \sum_{n=0}^{n} \sum_{m=-n}^{n} K_{nm}(\mathbf{r}, \mathbf{r}'_c)T_{nm;\nu-n,\mu-m}^{L2L}L_{\nu\mu}^{**}(\mathbf{r}', \mathbf{r}'_c)
\]
\[
= \sum_{n=0}^{P} \sum_{m=-n}^{n} K_{nm}(\mathbf{r}, \mathbf{r}'_c) \sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} T_{nm;\nu-n,\mu-m}^{L2L}L_{\nu\mu}^{**}(\mathbf{r}', \mathbf{r}'_c),
\] (4.71)

i.e.
\[
L_{nm}(\mathbf{r}', \mathbf{r}'_c) = \sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} T_{nm;\nu-n,\mu-m}^{L2L}L_{\nu\mu}^{**}(\mathbf{r}', \mathbf{r}'_c).
\] (4.72)

The basis transform (4.69) is indeed the same as the free-space version [35].

To derive the M2L translation, by applying (4.47) to each LE coefficient, we have
\[
L_{nm}^{**}(\mathbf{r}', \mathbf{r}'_c)
\]
\[
= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^{*\nu}(\mathbf{r}'_c, \mathbf{r}', k_\rho) \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \frac{(r'_c)^{\nu}}{2\nu + 1} Y_{\nu}^{\mu}(\theta'_c, \phi'_c) (-1)^{\mu}(\tau^*)^{\nu+\mu} c_{\nu\mu}1^{2\nu-\mu}k_\rho e^{i\nu}. \]
\[
\frac{c_{nm}}{2n+1} (\tau^*)^{n+m} i^{2n-m} k_\rho^{n-1} e^{-i\mu} \sigma_{\ell\ell'}^{*\nu}(k_\rho) dk_x dk_y
\]
\[
\approx \sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} A_{nm;\nu\mu}^{**}(\mathbf{r}, \mathbf{r}'_c) M_{\nu\mu}(\mathbf{r}', \mathbf{r}'_c),
\] (4.73)
where the M2L translation coefficients are

\[
A_{nm,\nu\mu}^{**}(r_c, r'_c) = \frac{a_{nm,\nu\mu}^{**}}{4\pi^2} \int_{\mathbb{R}^2} \mathcal{E}_{\ell\ell'}^{**}(r_c, r_c, k_r) k_r^{n+\nu-1} e^{i(\mu-m)\alpha} \sigma_{\ell\ell'}^{**}(k_r) dk_x dk_y, \tag{4.74}
\]

\[
a_{nm,\nu\mu}^{**} = \frac{1}{2n+1} (-1)^{\mu-i2n-2\nu-\mu}(\tau^*)^{n+m}(\tau^*)^{\nu+m} c_{nm} c_{\nu\mu}. \tag{4.75}
\]

4.4. The FMM framework

The FMM framework for the Laplace’s equation in 3-D layered media is proposed mostly identical to the previous algorithms for the Helmholtz cases in Section 2.4 and in Section 3.4. We inherit the definition of the polarization sources from Definition 3.9. In fact, the exponential convergence of the far-field expansions will be proven to be regarding the polarization like in Section 2.5, see Section 4.7.

Given target particles \( r_i \) in layer \( \ell \), and source particles \( r'_j \) with strength \( q_j \) in layer \( \ell' \), the FMM for the reaction field

\[
u_{\ell\ell'}^{**}(r_i) = \sum_j q_j u_{\ell\ell'}^{**}(r_i; r'_j) \tag{4.76}
\]

can be described in Algorithm 4.2.

Then, by joining the interaction for all the target-source layer pairs \((\ell, \ell')\), together with the free-space part using a conventional approach, the overall algorithm is summarized in Algorithm 4.3.

Again, the FMM box used in Algorithm 4.2 should place its center on the interface between the target layer and the polarization source layer as shown by Figure 3.2, so that children boxes won’t lie across the interface.
Algorithm 4.2: FMM for reaction field component $u_{\ell'\ell}^{**}(r_i)$.

Form the polarized source layer with each source $r'_j$ mapped to $P_{\ell'\ell}^{**}(r'_j)$; Generate an adaptive hierarchical octree with levels $0 \rightarrow H$ with polarization sources $P_{\ell'\ell}^{**}(r'_j)$ and original targets $r_i$;

// Initialize the output
for source particle $r_i$ do
\hspace{1em} $u_{\ell'\ell}^{**}(r_i) = 0$;
end

// Sanity check: skip if reaction field is prohibited
if $(\ell, \star)$ or $(\ell', \star) \in \{(0, \downarrow), (L, \uparrow)\}$ then return;

// The upward pass
for $\ell = H \rightarrow 0$ do
\hspace{1em} for box $j$ on source tree level $\ell$ do
\hspace{2em} if $j$ is a leaf node then
\hspace{3em} Form the ME using (4.49);
\hspace{2em} else
\hspace{3em} Form the ME by merging children’s expansions using the M2M translation (4.56) or its free-space equivalence (c.f. [35]);
\hspace{2em} end
\hspace{1em} end
\end for

// The downward pass
for $\ell = 1 \rightarrow H$ do
\hspace{1em} for box $j$ on target tree level $\ell$ do
\hspace{2em} Shift the LE of $j$’s parent to $j$ using L2L translation (4.72) or its free-space equivalence (c.f. [35]);
\hspace{2em} Collect the contribution from the interaction list using the M2L (4.73);
\hspace{1em} end
\end for

// LE evaluation
for leaf node $j$ do
\hspace{1em} for source particle $r_i$ in node $j$ do
\hspace{2em} Evaluate the LE at $r_i$ with (4.61), add add the result to $u_{\ell'\ell}^{**}(r_i)$;
\hspace{1em} end
end

// Neighboring direct interactions
for source particle $r_i$ do
\hspace{1em} for target particle $r_j'$ in a neighboring leaf box of $r_i$ do
\hspace{2em} Add the direct interaction $q_{ji}u_{\ell'\ell}^{**}(r_i; r'_j)$ to $u_{\ell'\ell}^{**}(r_i)$;
\hspace{1em} end
end
end
Algorithm 4.3: FMM in 3-D layered media

// The free-space part
for $\ell = 0 \rightarrow L$ do
  For source particles $\mathbf{r}_i$ and target particles in layer $\ell$, compute $u^f(\mathbf{r}_i)$ with the classic FMM [35] in the free space;
end

// The reaction fields
for $\ell = 0 \rightarrow L$ do
  for $\ell' = 0 \rightarrow L$ do
    For source particle $\mathbf{r}_i$ in layer $\ell$ and target particles in layer $\ell'$, compute $u_{\ell\ell'}^{\uparrow\uparrow}(\mathbf{r}_i), u_{\ell\ell'}^{\uparrow\downarrow}(\mathbf{r}_i), u_{\ell\ell'}^{\downarrow\uparrow}(\mathbf{r}_i)$ and $u_{\ell\ell'}^{\downarrow\downarrow}(\mathbf{r}_i)$ with Algorithm 4.2;
  end
end

// Summing up
for $\ell = 0 \rightarrow L$ do
  for source particle $\mathbf{r}_i$ in layer $\ell$ do
    $u(\mathbf{r}_i) = u^f(\mathbf{r}_i) + \sum_{\ell'=0}^{L} \left( u_{\ell\ell'}^{\uparrow\uparrow}(\mathbf{r}_i) + u_{\ell\ell'}^{\uparrow\downarrow}(\mathbf{r}_i) + u_{\ell\ell'}^{\downarrow\uparrow}(\mathbf{r}_i) + u_{\ell\ell'}^{\downarrow\downarrow}(\mathbf{r}_i) \right)$;
  end
end

4.5. Implementation details for efficiency

In this section we present a few methods to improve the efficiency as well as the accuracy in the numerical implementation of the FMM. The conversion from double integral to single integral has been mentioned in Section 3.5.1.

4.5.1. Recursion formula for numerical integration

For the M2L translation, integrals in the form

$$ S^n I_{nm}^{**} = \int_0^\infty J_m(k \rho \rho) \frac{S^n k^n e^{-k z \sigma_{\ell\ell'}^{**}(k \rho)}}{\sqrt{(n + m)! (n - m)!}} dk \rho, \quad 0 \leq m \leq n \leq 2N \quad (4.77) $$

are required to be numerically computed, where $\rho$ and $z$ refer to the horizontal distance and the polarized vertical distance between box centers, respectively, and the scaling factor $S$ is the size of the box where the computation occurs. If $\rho = 0$, then for any $m \geq 1$, $S^n I_{nm}^{**} = 0$,
and the overall computational cost will be $O(N)$. Otherwise, using the recursion formula for Bessel functions

$$J_{m+1}(k_\rho \rho) = \frac{2m}{k_\rho \rho} J_m (k_\rho \rho) - J_{m-1} (k_\rho \rho),$$

(4.78)

we have the forward recurrence formula

$$S^n I_{n,m+1}^{**} = \frac{2m}{\sqrt{(n + m)(n + m + 1)}} \rho \frac{S}{S^{n-1} I_{n-1,m}^{**}} - \frac{\sqrt{(n - m)(n - m + 1)}}{\sqrt{(n + m)(n + m + 1)}} S^n I_{n,m-1}^{**},$$

(4.79)

provided $m \geq 1$ and $n \geq m + 1$. In the FMM where the M2Ls shift between box centers on the same level, when the boxes do not coincide when projected on the $(x, y)$–plane, i.e. when $\rho \neq 0$, the box centers have a horizontal distance of at least $S$, so $S \leq \rho$. Hence,

$$\leq \frac{2m}{\sqrt{(n + m)(n + m + 1)}} \rho S + \frac{\sqrt{(n - m)(n - m + 1)}}{\sqrt{(n + m)(n + m + 1)}} \frac{\sqrt{(n - m)(n - m + 1)}}{\sqrt{(n + m)(n + m + 1)}} < 1,$$

(4.80)

which implies the forward recursion (4.79) is stable. In both cases, the number of numerical integration (4.77) is reduced to $O(N)$.

4.5.2. Changing contour of integration in the complex plane

The integrand of (4.77) may be unfriendly for numerical integration. For instance, when $n > 0$, the factor $k_\rho^n e^{-k_\rho z}$ has zero limits as $k_\rho \to 0$ and as $k_\rho \to \infty$, but its maximum, reached at $k_\rho = n/z$, is $(n/(ez))^n$, which can be very large when $z$ is small, i.e. when targets and sources are close to interfaces. Meanwhile, the Bessel function $J_m (k_\rho \rho)$ introduces persistent oscillation. In certain circumstances, changing the contour of integration to the imaginary axis can be beneficial, knowing the fact that $k_\rho \sigma(k_\rho)$ is also bounded on the imaginary axis (see Theorem 4.5). The contour change is taken by splitting $J_m (k_\rho \rho)$ into the average of a Hankel function pair, then switching to the positive and the negative imaginary axes,
respectively, i.e.,

\[
S_{\ell\ell'}^{n} I_{nm}^{*}(\rho, z) = \frac{1}{2} \int_{0}^{\infty} H_{m}^{(1)}(k_{\rho}\rho) \frac{(Sk_{\rho})^{n} e^{-k_{\rho}z}}{\sqrt{(n + m)!(n - m)!}} \sigma_{\ell\ell'}^{*}(k_{\rho}) dk_{\rho}
+ \frac{(-1)^{m+1}}{2} \int_{-\infty}^{0} H_{m}^{(1)}(k_{\rho}\rho) \frac{(-Sk_{\rho})^{n} e^{k_{\rho}z}}{\sqrt{(n + m)!(n - m)!}} \sigma_{\ell\ell'}^{*}(-k_{\rho}) dk_{\rho}
\]

\[
= \frac{i}{2} \int_{0}^{\infty} H_{m}^{(1)}(i\eta\rho) \frac{(i\eta S)^{n} e^{-i\eta z}}{\sqrt{(n + m)!(n - m)!}} \sigma_{\ell\ell'}^{*}(i\eta) d\eta
- \frac{i}{2} \int_{0}^{\infty} H_{m}^{(1)}(i\eta\rho) \frac{(-1)^{m+1}(-i\eta S)^{n} e^{i\eta z}}{\sqrt{(n + m)!(n - m)!}} \sigma_{\ell\ell'}^{*}(-i\eta) d\eta.
\]

(4.81)

A practical criteria for applying the contour change can be set as \(\rho > z\).

4.6. Numerical results

In this section, we present numerical results to demonstrate the performance of the proposed FMM for the Laplace’s equation in 3-D layered media. Data of numerical test results are from published journal papers [77], in collaboration with Bo Wang and Wei Cai. Bo Wang has been the primary contributor in the development of the code implementation for 3-D FMM in layered media.

The FMM for the Laplace’s equation in 3-D layered media is implemented based on an open-source adaptive FMM package DASHMM [25] on a workstation with two Xeon E5-2699 v4 2.2 GHz processors (each having 22 cores and 44 threads) and 500GB RAM using the GCC compiler version 6.3.

We first test the algorithm in a medium with three layers, with interfaces placed at \(d_0 = 0, d_1 = -1.2\). Charges are set to be uniformly distributed in irregular domains which are obtained by shifting the domain determined by \(r = 0.599 - a + \frac{z}{8}(35 \cos^4 \theta - 30 \cos^2 \theta + 3)\) with \(a = 0.1, 0.15, 0.05\) to new centers \((0, 0, 0.6), (0, 0, -0.6)\) and \((0, 0, -1.8)\), respectively (see Figure 4.1 for the cross section of the domains). All particles are generated by keeping the uniformly distributed particles in a larger cube within corresponding irregular domains. We can see the minimum distance between charges and the interfaces is 0.001 in all three
layers, i.e., the numerical examples have charges very close to the interfaces. The material parameters for the layers are set to be $\varepsilon_0 = 21.2$, $\varepsilon_1 = 47.5$, $\varepsilon_2 = 62.8$. Let $u_\ell(r_i)$ be the field at $r_i$ from layer $\ell$, including both the free-space part and the reaction field part, while in the free-space part the interaction with itself is excluded. Let $\tilde{u}_\ell(r_i)$ be the approximated values of $u_\ell(r_i)$ calculated by the FMM. Define $\ell^2$ and maximum errors as

$$
\text{Err}_\ell^2 := \sqrt{\frac{\sum_{i=1}^{N_\ell} |u_\ell(r_i) - \tilde{u}_\ell(r_i)|^2}{\sum_{i=1}^{N_\ell} |u_\ell(r_i)|^2}}, \quad \text{Err}_\ell^\text{max} := \max_{1 \leq i \leq N_\ell} \frac{|u_\ell(r_i) - \tilde{u}_\ell(r_i)|}{|u_\ell(r_i)|},
$$

where $N_\ell$ is the number of target particles in layer $\ell$. To test the accuracy of the FMM, we put $N = 1168 + 856 + 1504$ particles in the irregular domains in the layers, see Figure 4.1.

Convergence rates against the truncation index $P$ are depicted in Figure 4.2 (a). Clearly, the proposed FMM has a spectral convergence with respect to the truncation index $P$. The CPU time for the computation of all three free space components $\{u_\ell^f(r_i)\}_{\ell=0}^2$ and sixteen reaction components $u_\ell^*(r_i)$ with fixed truncation index $P = 5$ are compared in Figure 4.2 (b) for up to 12 millions charges. It shows that all of them have an $O(N)$ complexity while the CPU time for the computation of reaction components is much shorter than that for free space components due to the fact that most of the equivalent polarization sources are...
Figure 4.2: Performance of FMM for Laplace’s equation in a 3-layer medium, picture from [78]. The notations \( \Phi \) in [78] refers to the field \( u \) in this thesis.

well-separated from the targets, just like the phenomenon we have observed in the Helmholtz and the linearized PB cases in Section 3.6. The CPU times with multiple cores are given in Table 4.1 and they show that the speedup of the parallel computing for reaction components is lower than that for the free space components, as the pre-computation for the numerical integrals \( \{ S_n f_{n\pi}^*(\rho, z) \}_{n=0}^{2p} \) defined in (4.77) have not been implemented in parallel. Here, we only used parallel implementation within the FMM for each reaction field component, as the computation of them is independent from each other.

Figure 4.3: Performance of FMM for Laplace’s equation in a multi-layer medium, picture from [78].
Next, to test a more practical problem, we consider a typical layout of a multi-layered solar cell (cf. [67]). The main materials are Gallium Arsennide (GaAs), Indium Arsenide (InAs) and Silicon (Si). Their relative dielectric constants are 12.9, 15.15 and 2.4, respectively. We will test different cases with up to 32 layers and 32 million charges. In each case, the width of the layers is fixed to be 1.2 and three materials GaAs, InAs and Si are randomly selected in the layered structure. Moreover, the source charges are assumed to be uniformly distributed in cubic boxes of size 1 in each layer. For the test of accuracy, we consider a medium with 16 layers and put 1000 randomly selected source charges in each box. Therefore, the total number of charges is 16000. The errors of the potentials in the 3rd, 7th and 15th layers are depicted in Figure 4.3 (a). We can see that highly accurate results can be obtained with $P$ less than 20. To show the dependence of the CPU time on the number of layers $L$, we test examples of different $L$ and put 1 million randomly selected source charges in each layer. 40 CPU cores are used and the CPU time is depicted in Figure 4.3 (b). The results show that the FMM can handle large number of sources in many layers efficiently. The number of

<table>
<thead>
<tr>
<th>cores</th>
<th>number of charges</th>
<th>time for all ${u^I_{\ell}}_{\ell=0}^2$</th>
<th>time for all ${u^*_{\ell\ell}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4164016</td>
<td>426.66</td>
<td>67.76</td>
</tr>
<tr>
<td></td>
<td>6229016</td>
<td>436.30</td>
<td>57.75</td>
</tr>
<tr>
<td></td>
<td>8883960</td>
<td>455.32</td>
<td>75.97</td>
</tr>
<tr>
<td></td>
<td>12202880</td>
<td>527.56</td>
<td>132.06</td>
</tr>
<tr>
<td>6</td>
<td>4164016</td>
<td>80.78</td>
<td>21.86</td>
</tr>
<tr>
<td></td>
<td>6229016</td>
<td>82.96</td>
<td>20.76</td>
</tr>
<tr>
<td></td>
<td>8883960</td>
<td>86.54</td>
<td>25.74</td>
</tr>
<tr>
<td></td>
<td>12202880</td>
<td>100.08</td>
<td>42.94</td>
</tr>
<tr>
<td>36</td>
<td>4164016</td>
<td>24.97</td>
<td>20.58</td>
</tr>
<tr>
<td></td>
<td>6229016</td>
<td>25.71</td>
<td>18.81</td>
</tr>
<tr>
<td></td>
<td>8883960</td>
<td>26.02</td>
<td>23.45</td>
</tr>
<tr>
<td></td>
<td>12202880</td>
<td>28.51</td>
<td>36.72</td>
</tr>
</tbody>
</table>

Table 4.1: Comparison of CPU time (in seconds) with multiple CPU cores for FMM for Laplace’s equation in 3-D layered media, the truncation index $P = 5$, data from [78].
free space and reaction field components are $L + 1$ and $4L^2$, respectively, when a stratified medium with $L + 1$ layers is considered. Therefore, the CPU time for free space and reaction field components is expected to be $O(L)$ and $O(L^2)$, respectively, as verified in Figure 4.3 (b).

4.7. Proof of exponential convergence of far-field expansions

We study the exponential convergence of the far-field expansions proposed in Section 4.3. Depending on the number of target or source centers involved, these expansions are divided into two types, which are called the single expansion and the double expansion in this section.

**Definition 4.6 (Single expansion).** Let $\sigma(k_\rho)$ be a holomorphic function in the right half plane with upper bound $|\sigma(k_\rho)| \leq 1$. Let $r = (x, y, z), r_1 = (x_1, y_1, z_1)$ be real vectors satisfying

$$|r| > |r_1|, \quad z > 0, \quad z + z_1 > 0. \quad (4.83)$$

Then, we have the single expansion for the integral

$$I_S = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} e^{k_\rho k_0 \cdot (r + r_1)} k_\rho^{-1} \sigma(k_\rho) dk_x dk_y \quad (4.84)$$

$$= \sum_{n=0}^{\infty} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} e^{k_\rho k_0 \cdot r} \frac{(k_\rho k_0 \cdot r_1)^n}{n!} k_\rho^{-1} \sigma(k_\rho) dk_x dk_y. \quad (4.85)$$

**Definition 4.7 (Double expansion).** Let $\sigma(k_\rho)$ be a holomorphic function in the right half plane with upper bound $|\sigma(k_\rho)| \leq 1$. Let $r = (x, y, z), r_1 = (x_1, y_1, z_1), r_2 = (x_2, y_2, z_2)$ be real vectors satisfying

$$|r| > |r_1| + |r_2|, \quad z > 0, \quad z + z_1 + z_2 > 0. \quad (4.86)$$
Then, we have the double expansion for the integral

\[ I_D = \int_{\mathbb{R}^2} e^{k_\rho k_0 \cdot (r_r + r_1 + r_2)} k_\rho^{-1} \sigma(k_\rho) dk_x dk_y \]  

\[ = \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} \int_{\mathbb{R}^2} e^{k_\rho k_0 \cdot r} \frac{(k_\rho k_0 \cdot r_1)^n}{n!} \frac{(k_\rho k_0 \cdot r_2)^\nu}{\nu!} k_\rho^{-1} \sigma(k_\rho) dk_x dk_y. \]  

(4.87)  

(4.88)

The single expansion is corresponding to (4.47) for the ME and (4.59) for the LE, where the further expansion separating the angular terms with index \(-n \leq m \leq n\) is lossless. The double expansion is corresponding to the M2M, L2L and M2L, where two particle centers are involved.

The proof of exponential convergence of these expansions consists of forward and backward contour shifts where the Fubini’s theorem can be applied under the Cagniard–de Hoop transform. Then we will formally apply them to the far-field expansions.

4.7.1. The Cagniard–de Hoop transform

Let \((\rho, \phi)\) be the polar coordinates of \((x, y)\). For \(j = 1, 2\), let \((\rho_j, \phi_j)\) be the polar coordinates of \((x_j, y_j)\). On the \((k_x, k_y)\)-plane, let \(\xi, \eta \in \mathbb{R}\) such that

\[ \xi + i\eta = k_\rho e^{i(\alpha - \phi)} = k_\rho \cos(\alpha - \phi) + ik_\rho \sin(\alpha - \phi), \]  

or equivalently,

\[ k_x = \xi \cos \phi - \eta \sin \phi, \]  

\[ k_y = \xi \sin \phi + \eta \cos \phi. \]  

(4.89)  

(4.90)

With this substitution,

\[ k_\rho k_0 \cdot r = i\xi - \zeta, \]  

\[ k_\rho k_0 \cdot r_j = i\xi_j \cos(\phi - \phi_j) - \eta \rho_j \sin(\phi - \phi_j) - \zeta z_j, \]  

(4.91)
where
\[ \zeta = \sqrt{\xi^2 + \eta^2} = k_\rho. \] (4.92)

To convert the exponent \( i\xi \rho - \zeta z \) to a negative real number, we first switch the integrals on the \((k, k_y)\)-plane to the \((\eta, \xi)\)-plane, then introduce the Cagniard–de Hoop transform for each \( \eta > 0 \), while the case with \( \eta < 0 \) will be similar and the detailed derivation will be omitted.

**Lemma 4.8 (Cagniard–de Hoop transform).** Given \( \eta > 0 \). Given \( r = (x, y, z), r' = (x', y', z') \in \mathbb{R}^3 \) with spherical coordinates \((r, \theta, \phi)\) and \((r', \theta', \phi')\), respectively, satisfying
\[ r > r', \quad z > 0, \quad z + z' > 0. \] (4.93)

Let \( \rho = \sqrt{x^2 + y^2}, \rho' = \sqrt{x'^2 + y'^2} \). Denote by \( \Omega^+ \subset \mathbb{C} \) the complex domain between the real axis and the contour \( \Gamma \) defined by the parametric \( \xi_\pm(t) \) in (4.96). Let \( f(\xi) \) be a holomorphic function in \( \Omega^+ \) with a polynomial bound \( |f(\xi)| \leq C(1 + |\xi|)^m \) for some integer \( m \) and some constant \( C > 0 \), and let
\[ g(\xi) = f(\xi)e^{k_0r'}, = f(\xi)e^{i\xi \rho \cos(\phi - \phi') - \eta \rho \sin(\phi - \phi') - \sqrt{\eta^2 + \xi^2}z'}. \] (4.94)

Then, there holds
\[ \int_{-\infty}^{\infty} g(\xi)e^{i\xi \rho - \sqrt{\eta^2 + \xi^2}z}d\xi = \int_{1}^{\infty} \left( g(\xi_+(t))\zeta_+(t) - g(\xi_-(t))\zeta_-(t) \right) \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} dt, \] (4.95)

where \( \xi_\pm(t), \zeta_\pm(t) \) are defined by the Cagniard–de Hoop transform
\[ \xi_\pm(t) = \frac{\eta}{r}(ip t \pm z \sqrt{t^2 - 1}), \quad \zeta_\pm(t) = \sqrt{\xi_\pm(t)^2 + \eta^2} = \frac{\eta}{r}(zt \pm i p \sqrt{t^2 - 1}). \] (4.96)

**Proof.** When \( \rho = 0 \) the result is a trivial substitution of variables. Now suppose \( \rho > 0 \). For \( \xi \in \mathbb{C} \), define the branch cut of square roots on this complex plane as \((-\infty, 0]\). Then,
Figure 4.4: The Cagniard–de Hoop transform for $\xi$ from the real axis to $\Gamma_+ \cup \Gamma_-$. 

$\zeta = \sqrt{\xi^2 + \eta^2}$ has its branch cut given by the red lines \{i$x$ : $x \in \mathbb{R}, |x| \geq \rho \eta / r$\} lying on the imaginary axis in Figure 4.4. Define a hyperbolic integral path

$$\Gamma = \Gamma_+ \cup \Gamma_-, \quad \Gamma_\pm = \{\xi_\pm(t) : t \geq 1\}.$$ (4.97)

We consider the contour transform for $\xi \in (-\infty, +\infty)$ to $\Gamma$. For any $R > 0$, let $O^+_R$ and $O^-_R$ be the parts of the circle \{\xi : |\xi| = R\} that are bounded by the real axis and $\Gamma_\pm$, respectively (see Figure 4.4). Denote by $\xi(t^\pm_R) = Re^{i\theta^\pm_R}$ the intersections between $O^\pm_R$ and $\Gamma^\pm$, respectively. Then $0 < \theta^+_R < \frac{\pi}{2} < \theta^-_R < \pi$. Since $\Gamma$ has a pair of asymptotes passing through the origin with slopes $\pm \rho / z$, we have

$$\lim_{R \to \infty} \theta^+_R = \tan^{-1} \frac{\rho}{z}, \quad \lim_{R \to \infty} \theta^-_R = \pi - \tan^{-1} \frac{\rho}{z}.$$ (4.98)
We will show that
\[
\lim_{R \to \infty} \int_{O_R^+} g(\xi)e^{i\xi\rho - \sqrt{\eta^2 + \xi^2} z} d\xi = 0. \tag{4.99}
\]
Specifically, since the integrand \( g(\xi)e^{i\xi\rho - \sqrt{\eta^2 + \xi^2} z} = f(\xi)\exp(E(\xi)) \), where \(|f(\xi)|\) has a polynomial bound of \( R \), and
\[
E(\xi) = (i\xi\rho - \sqrt{\eta^2 + \xi^2} z) + (i\xi\rho' \cos(\phi - \phi') - i\eta\rho' \sin(\phi - \phi') - \sqrt{\eta^2 + \xi^2} z'), \tag{4.100}
\]
it suffices to give an estimate on \( \Re E(\xi) \). For any \( \theta_R \in [0, \theta_R^+] \), for \( \xi = R e^{i\theta_R} \), as \( R \to \infty \), we have \( \sqrt{\eta^2 + \xi^2} \sim \xi \), so
\[
\Re E(\xi) \sim -R ((\rho + \rho' \cos(\phi - \phi')) \sin \theta_R + (z + z') \cos \theta_R). \tag{4.101}
\]
If \( \rho \geq \rho' \), then in (4.101),
\[
(\rho + \rho' \cos(\phi - \phi')) \sin \theta_R + (z + z') \cos \theta_R \geq (z + z') \cos \theta_R^+ \sim (z + z') \frac{z}{r}. \tag{4.102}
\]
Otherwise,
\[
(\rho + \rho' \cos(\phi - \phi')) \sin \theta_R + (z + z') \cos \theta_R \\
\geq (\rho - \rho') \sin \theta_R^+ + (z + z') \cos \theta_R^+ \\
\sim r - \left( \frac{\rho - \rho'}{r} \frac{z}{r'} \right) \\
\geq r - r'. \tag{4.103}
\]
Therefore, there exists \( R_0 > 0 \) such that when \( R > R_0 \),
\[
\Re E(\xi) \leq -\frac{R}{2} \min \left\{ \frac{(z + z') z}{r}, r - r' \right\}, \tag{4.104}
\]
121
so the limit (4.99) holds. A similar result can be established for the integral on $O_R$.

As the branch cut guarantees that $g(\xi)e^{i\xi(\sqrt{\eta^2 + \xi^2} - \sqrt{\eta^2})}$ is holomorphic in the domain $\Omega^+$ for any fixed $\eta > 0$, by Cauchy’s integral theorem, (4.95) follows from the facts

$$\int_{-\infty}^{\infty} g(\xi)e^{i\xi(\sqrt{\eta^2 + \xi^2} - \sqrt{\eta^2})} d\xi = \int_{\Gamma} g(\xi)e^{i\xi(\sqrt{\eta^2 + \xi^2} - \sqrt{\eta^2})} d\xi$$  \hspace{1cm} (4.105)

and

$$\frac{d\xi_+(t)}{dt} = \frac{\zeta_+(t)}{\sqrt{t^2 - 1}}, \quad \frac{d\xi_-(t)}{dt} = -\frac{\zeta_-(t)}{\sqrt{t^2 - 1}}.$$  \hspace{1cm} (4.106)

4.7.2. Exponential convergence of single expansions and double expansions

Let $\epsilon$ be any positive real number. Since Lemma 4.8 requires $\eta > 0$, we will estimate the expansions with an interval $\eta \in (-\epsilon, \epsilon)$ excluded, then later take the limit back.

When applying the Cagniard–de Hoop transform, corresponding to the sides of the hyperbola contour $\Gamma_{\pm}$ adopted, we accordingly define

$$s_{j}^\pm(\eta,t) = k_0 \cdot \rho_j \cdot \mathbf{x}_j = i\xi_\pm(t)\rho_j \cos(\phi - \phi_j) - i\eta \rho_j \sin(\phi - \phi_j) - \zeta_\pm(t)z_j, \quad j = 1, 2. \hspace{1cm} (4.107)$$

**Lemma 4.9.** In the domain $(\eta, t) \in (\epsilon, \infty) \times [1, \infty)$,

$$|s_{j}^\pm(\eta, t)| \leq \eta \rho_j t.$$  \hspace{1cm} (4.108)

**Proof.** Obviously

$$|\zeta_\pm(t)| = \frac{\eta}{r} \sqrt{\eta^2 - r^2}.$$  \hspace{1cm} (4.109)

Let

$$w = \frac{\xi_\pm(t)\cos(\phi - \phi_j) - \eta \sin(\phi - \phi_j)}{\zeta_\pm(t)} = \frac{1}{2} \frac{\xi_\pm(t) + i\eta}{\zeta_\pm(t)} e^{i(\phi_j - \phi_j)} + \frac{1}{2} \frac{\xi_\pm(t) - i\eta}{\zeta_\pm(t)} e^{-i(\phi_j - \phi_j)}.$$  \hspace{1cm} (4.110)
Simple calculation reveals

$$\left| \frac{\xi_\pm(t) + i\eta}{\zeta_\pm(t)} \right| = \sqrt{\frac{rt + \rho}{rt - \rho}}, \quad \left| \frac{\xi_\pm(t) - i\eta}{\zeta_\pm(t)} \right| = \sqrt{\frac{rt - \rho}{rt + \rho}}. \quad (4.111)$$

Thus, there exists $\gamma_\pm \in [0, 2\pi)$ for this reciprocal pair

$$\frac{\xi_\pm(t) + i\eta}{\zeta_\pm(t)} = \sqrt{\frac{rt + \rho}{rt - \rho}} e^{i\gamma_\pm}, \quad \frac{\xi_\pm(t) - i\eta}{\zeta_\pm(t)} = \sqrt{\frac{rt - \rho}{rt + \rho}} e^{-i\gamma_\pm}. \quad (4.112)$$

Let $\psi_\pm = \gamma_\pm + \phi - \phi_j$, we then have

$$w = \frac{1}{2} \sqrt{\frac{rt + \rho}{rt - \rho}} e^{i\psi_\pm} + \frac{1}{2} \sqrt{\frac{rt - \rho}{rt + \rho}} e^{-i\psi_\pm} = \frac{rt \cos \psi_\pm + i\rho \sin \psi_\pm}{\sqrt{r^2 t^2 - \rho^2}}, \quad (4.113)$$

so

$$|s_j^\pm(\eta, t)| = |\zeta_\pm(t)| |\rho_j w + iz_j|$$

$$= \frac{\eta}{r} \left| \rho_j rt \cos \psi_\pm + i\rho_j \rho \sin \psi_\pm + iz_j \sqrt{r^2 t^2 - \rho^2} \right|$$

$$= \frac{\eta}{r} \sqrt{(\rho_j rt \cos \psi_\pm)^2 + (\rho_j \rho \sin \psi_\pm + z_j \sqrt{r^2 t^2 - \rho^2})^2}$$

$$= \frac{\eta}{r} \sqrt{(\rho_j \rho + z_j \sqrt{r^2 t^2 - \rho^2} \sin \psi_\pm)^2 + r_j^2 (r^2 t^2 - \rho^2) \cos^2 \psi_\pm} \quad (4.114)$$

$$\leq \frac{\eta}{r} \sqrt{(\rho_j^2 + z_j^2)(\rho^2 + (r^2 t^2 - \rho^2) \sin^2 \psi_\pm) + r_j^2 (r^2 t^2 - \rho^2) \cos^2 \psi_\pm}$$

$$= \eta r_j t.$$

**Lemma 4.10 (Single expansion in the $(\eta, t)$ domain).** With the conditions in Definition 4.6, let

$$I_\epsilon = \int_\epsilon^\infty \int_1^\infty \frac{e^{-\eta t}}{\sqrt{t^2 - 1}} \left( e^{s_j^+ (\eta, t)} \sigma(\zeta_+(t)) - e^{s_j^- (\eta, t)} \sigma(\zeta_-(t)) \right) dt d\eta, \quad (4.115)$$

$$I_\epsilon^{(n)} = \int_\epsilon^\infty \int_1^\infty \frac{e^{-\eta t}}{\sqrt{t^2 - 1}} \left( \frac{(s_j^+ (\eta, t))^n}{n!} \sigma(\zeta_+(t)) - \frac{(s_j^- (\eta, t))^n}{n!} \sigma(\zeta_-(t)) \right) dt d\eta, \quad (4.116)$$
then
\[ I_\epsilon = \sum_{n=0}^{\infty} I^{(n)}_\epsilon, \tag{4.117} \]

with \( P + 1 \) term truncation error
\[ \left| I_\epsilon - \sum_{n=0}^{P} I^{(n)}_\epsilon \right| \leq \pi \frac{1}{r - r_1} \left( \frac{r_1}{r} \right)^{P+1}. \tag{4.118} \]

Proof. Using Lemma 4.9, we have estimates of the absolute integrals
\[
\tilde{I}^{(n)}_\epsilon = \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \left( \frac{|s^+(\eta, t)|^n}{n!} |\sigma(\zeta_+(t))| + \frac{|s^-_1(\eta, t)|^n}{n!} |\sigma(\zeta_-(t))| \right) \, dt \, d\eta \\
\leq 2 \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \frac{(r_1 \eta t)^n}{n!} \, dt \, d\eta \\
\leq 2 \int_{1}^{\infty} \frac{1}{\sqrt{t^2 - 1}} \int_{0}^{\infty} e^{-\eta rt} \frac{(r_1 \eta t)^n}{n!} \, d\eta \, dt \\
= 2 \int_{1}^{\infty} \frac{1}{r} \left( \frac{r_1}{r} \right)^n \frac{dt}{t \sqrt{t^2 - 1}} \\
= \pi \frac{1}{r} \left( \frac{r_1}{r} \right)^n. \tag{4.119} \]

so
\[
\sum_{n=0}^{\infty} \tilde{I}^{(n)}_\epsilon \leq \sum_{n=0}^{\infty} \pi \frac{1}{r} \left( \frac{r_1}{r} \right)^n = \pi \frac{1}{r - r_1}. \tag{4.120} \]

By Fubini’s theorem, (4.117) holds. For a truncated sum at \( n \leq P \), the truncation error
\[
\left| \sum_{n=P+1}^{\infty} I^{(n)}_\epsilon \right| \leq \sum_{n=P+1}^{\infty} \tilde{I}^{(n)}_\epsilon \leq \sum_{n=P+1}^{\infty} \pi \frac{1}{r} \left( \frac{r_1}{r} \right)^n = \pi \frac{1}{r - r_1} \left( \frac{r_1}{r} \right)^{P+1}. \tag{4.121} \]
Lemma 4.11 (Double expansion in the \((\eta, t)\) domain). With the conditions in Definition 4.7, let

\[
I_{\epsilon} = \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \left( e^{s_1^+(\eta,t) + s_2^+(\eta,t)} \sigma(\zeta_+(t)) - e^{s_1^-(\eta,t) + s_2^-(\eta,t)} \sigma(\zeta_-(t)) \right) dt d\eta, \tag{4.122}
\]

\[
I_{\epsilon}^{(n\nu)} = \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \left( \frac{(s_1^+)^n (s_2^+)\nu}{n! \nu!} \sigma(\zeta_+(t)) - \frac{(s_1^-)^n (s_2^-)\nu}{n! \nu!} \sigma(\zeta_-(t)) \right) dt d\eta, \tag{4.123}
\]

then

\[
I_{\epsilon} = \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} I_{\epsilon}^{(n\nu)}, \tag{4.124}
\]

with \((P + 1)^2\) term truncation error

\[
\left| I_{\epsilon} - \sum_{n=0}^{P} \sum_{\nu=0}^{P} I_{\epsilon}^{(n\nu)} \right| \leq \pi \frac{1}{r - r_1 - r_2} \left( \frac{r_1}{r - r_2} \right)^{P+1} + \pi \frac{1}{r - r_1 - r_2} \left( \frac{r_2}{r - r_1} \right)^{P+1}. \tag{4.125}
\]

Proof. Using Lemma 4.9, we have the estimates of the absolute integrals

\[
\tilde{I}_{\epsilon}^{(n\nu)} = \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \left( |s_1^+|^n |s_2^+|^\nu \sigma(\zeta_+(t))| + |s_1^-|^n |s_2^-|^\nu \sigma(\zeta_-(t))| \right) dt d\eta
\]

\[
\leq 2 \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta rt}}{\sqrt{t^2 - 1}} \left( \eta r_1 t^n (\eta r_2 t)^\nu \right) \frac{1}{n! \nu!} dtd\eta
\]

\[
\leq 2 \int_{1}^{\infty} \frac{1}{\sqrt{t^2 - 1}} \int_{0}^{\infty} e^{-\eta rt} \left( \eta r_1 t^n \right) \frac{1}{n! \nu!} dt d\eta
\]

\[
= 2 \int_{1}^{\infty} \frac{(n + \nu)!}{n! \nu!} \frac{r_1^n r_2^\nu}{t^{n+\nu+1}} \frac{dt}{t^2 - 1}
\]

\[
= \pi \frac{(n + \nu)!}{n! \nu!} \frac{r_1^n r_2^\nu}{r^{n+\nu+1}}
\]

so

\[
\sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} \tilde{I}_{\epsilon}^{(n\nu)} \leq \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} \pi \frac{(n + \nu)!}{n! \nu!} \frac{r_1^n r_2^\nu}{r^{n+\nu+1}} = \pi \frac{1}{r - r_1 - r_2}. \tag{4.127}
\]
By Fubini’s theorem, (4.124) holds. For a truncated sum at \( n, \nu \leq P \), the truncation error

\[
\left| I_\epsilon - \sum_{n=0}^{P} \sum_{\nu=0}^{P} I^{(n\nu)}_\epsilon \right| \leq \sum_{n=P+1}^{\infty} \sum_{\nu=0}^{P} i^{(n\nu)}_\epsilon + \sum_{\nu=P+1}^{\infty} \sum_{n=0}^{P} i^{(n\nu)}_\epsilon
\]

\[
\leq \sum_{n=P+1}^{\infty} \frac{n!}{r_{n+1}^{n+1}} \sum_{\nu=0}^{P} \frac{(n+\nu)!}{\nu!} \frac{r_{\nu}^{\nu+1}}{r_{n+1}^{n+1}} + \sum_{\nu=P+1}^{\infty} \frac{n!}{r_{n+1}^{n+1}} \sum_{n=0}^{P} \frac{(n+\nu)!}{\nu!} \frac{r_{\nu}^{\nu+1}}{r_{n+1}^{n+1}}
\]

\[
= \sum_{n=P+1}^{\infty} \frac{n!}{r_{n+1}^{n+1}} \frac{r_{n+1}^{n+1}}{(r - r_2)^{n+1}} + \sum_{\nu=P+1}^{\infty} \frac{n!}{r_{n+1}^{n+1}} \frac{r_{\nu}^{\nu+1}}{(r - r_2)^{\nu+1}}
\]

\[
= \pi \frac{1}{r - r_1 - r_2} \left( \frac{r_1}{r - r_2} \right)^{P+1} + \pi \frac{1}{r - r_1 - r_2} \left( \frac{r_2}{r - r_1} \right)^{P+1}.
\]

(4.128)

**Theorem 4.12 (Exponential convergence of single expansion).** The single expansion

in Definition 4.6 converges with \( P + 1 \) term truncation error

\[
\left| I_S - \sum_{n=0}^{P} \int_{\mathbb{R}^2} e^{k_{x}k_{0}(r + r_1)} k_{p}^{-1} \sigma(k_{p}) dk_{x} dk_{y} \right| \leq 2\pi \frac{1}{r - r_1} \left( \frac{r_1}{r} \right)^{P+1},
\]

(4.129)

where \( r = |r|, r_1 = |r_1| \).

**Proof.** Since \( z > 0 \) and \( z + z_1 > 0 \), these integrals involved are absolutely convergent, so we can change the coordinates of integration from \((k_{x}, k_{y})\) to \((\eta, \xi)\). Let \( \epsilon \) be an arbitrary positive real number. From

\[
I_S = \int_{\mathbb{R}^2} e^{k_{x}k_{0}(r + r_1)} k_{p}^{-1} \sigma(k_{p}) dk_{x} dk_{y} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{k_{x}k_{0}(r + r_1)} \zeta^{-1} \sigma(\zeta) d\xi d\eta,
\]

(4.130)

using Lemma 4.8 with \( r' = r_1 \) and Lemma 4.10,

\[
I^+_S(\epsilon) := \int_{\epsilon}^{\infty} \int_{-\infty}^{\infty} e^{k_{x}k_{0}(r + r_1)} \zeta^{-1} \sigma(\zeta) d\xi d\eta
\]

\[
= \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-\eta t}}{\sqrt{\epsilon^2 - 1}} \left( e^{s^{+}_{\eta}(t)} \sigma(\zeta_{+}(t)) - e^{s^{-}_{\eta}(t)} \sigma(\zeta_{-}(t)) \right) dt d\eta
\]

(4.131)
has $P + 1$ term approximation

$$
\left| I^+_S(\epsilon) - \sum_{n=0}^{P} I^+_{S,n}(\epsilon) \right| \leq \pi \frac{1}{r - r_1} \left( \frac{r_1}{r} \right)^{P+1},
$$

(4.132)

where, by using Lemma 4.8 with $r' = (0, 0, 0)$, each

$$
I^+_{S,n}(\epsilon) = \int_{\epsilon}^{\infty} \int_{1}^{\infty} \frac{e^{-nrt}}{\sqrt{t^2 - 1}} \left( \frac{(s^+_1(\eta, t))^n}{n!} \sigma(\zeta_+ (t)) - \frac{(s^-_1(\eta, t))^n}{n!} \sigma(\zeta_- (t)) \right) \, dt \, d\eta
$$

(4.133)

$$
= \int_{\epsilon}^{\infty} \int_{-\infty}^{\infty} e^{k_\mu k_0 \cdot r} \frac{(k_\mu k_0 \cdot r)^n}{n!} \zeta^{-1} \sigma(\zeta) \, d\xi \, d\eta.
$$

By taking the limit in (4.132) as $\epsilon \to 0$, we get

$$
\left| I^+_S(0) - \sum_{n=0}^{P} I^+_{S,n}(0) \right| \leq \pi \frac{1}{r - r_1} \left( \frac{r_1}{r} \right)^{P+1}.
$$

(4.134)

The same theories can be developed for $\eta < 0$, with a substitution using $-\eta$ for $\eta$. The summation of parts for $\eta \geq 0$ and $\eta < 0$ completes the proof.

Theorem 4.13 (Exponential convergence of double expansion). The double expansion in Definition 4.7 converges with $(P + 1)^2$ term truncation error

$$
\left| I_D - \sum_{n=0}^{P} \sum_{\nu=0}^{P} \int_{\mathbb{R}^2} e^{k_\mu k_0 \cdot r} \frac{(k_\mu k_0 \cdot r_1)^n}{n!} \frac{(k_\mu k_0 \cdot r_2)^\nu}{\nu!} k_\rho^{-1} \sigma(k_\rho) \, dk_\lambda \, dk_\gamma \right| 
$$

(4.135)

$$
\leq 2\pi \frac{1}{r - r_1 - r_2} \left( \frac{r_1}{r - r_2} \right)^{P+1} + 2\pi \frac{1}{r - r_1 - r_2} \left( \frac{r_2}{r - r_1} \right)^{P+1},
$$

where $r = |r|$, $r_1 = |r_1|$, $r_2 = |r_2|$.

The proof resembles the previous one with Lemma 4.11 applied.

4.7.3. Exponential convergence of far-field expansions

Now we apply the exponential convergence of single expansions and double expansions to the far-field expansions proposed in Section 4.3.
Suppose $C > 0$ is an upper bound of the reflection/transmission coefficients $\sigma_{\ell\ell}^*(k_\rho)$ for $k_\rho$ in the right half complex plane, as indicated by Theorem 4.5. Define the polarization distance

$$D_{\ell\ell}^*(\mathbf{r}, \mathbf{r}') = \sqrt{(x-x')^2 + (y-y')^2 + (\tau^*(z - d_\ell^*) + \tau^*(z' - d_{\ell'}^*))} \quad (4.136)$$

like in the 2-D case (2.80) and previous 3-D cases (3.78).

For the ME (4.49), we apply Theorem 4.12 with $\mathbf{r} = (x - x_c, y - y_c, \tau^*(z - d_\ell^*) + \tau^*(z_c - d_{\ell'}^*))$, $\mathbf{r}_1 = (x_c - x', y_c - y', \tau^*(z' - z_c))$ to get

$$\left| u_{\ell\ell}^{**}(\mathbf{r}, \mathbf{r}') - \sum_{n=0}^{P} \sum_{m=-n}^{n} M_{nm}(\mathbf{r}', \mathbf{r}_c) F_{nm}^{**}(\mathbf{r}, \mathbf{r}_c) \right| \leq \frac{1}{2\pi} \frac{1}{D_{\ell\ell}^*(\mathbf{r}, \mathbf{r}_c) - |\mathbf{r}' - \mathbf{r}_c|} \left( \frac{|\mathbf{r}' - \mathbf{r}_c|}{D_{\ell\ell}^*(\mathbf{r}, \mathbf{r}_c)} \right)^{P+1},$$

provided the conditions

$$|\mathbf{r}' - \mathbf{r}_c| < D_{\ell\ell}^*(\mathbf{r}, \mathbf{r}_c), \quad \tau^*(z_c - d_{\ell'}) > 0 \quad (4.138)$$

hold. In comparison with the free-space ME (4.31) where we can treat $\sigma_{\ell\ell}^{**} = 1/2$, the truncation error estimate provides the same result.

For the LE (4.61), we apply Theorem 4.12 with $\mathbf{r} = (x - x_c, y - y_c, \tau^*(z - d_\ell^*) + \tau^*(z_c - d_{\ell'}^*))$, $\mathbf{r}_1 = (x_c - x', y_c - y', \tau^*(z' - z_c))$ to get

$$\left| u_{\ell\ell}^{**}(\mathbf{r}; \mathbf{r}') - \sum_{n=0}^{P} \sum_{m=-n}^{n} K_{nm}(\mathbf{r}, \mathbf{r}_c) L_{nm}^{**}(\mathbf{r}', \mathbf{r}_c) \right| \leq \frac{1}{2\pi} \frac{1}{D_{\ell\ell}^*(\mathbf{r}_c, \mathbf{r}') - |\mathbf{r} - \mathbf{r}_c|} \left( \frac{|\mathbf{r} - \mathbf{r}_c|}{D_{\ell\ell}^*(\mathbf{r}_c, \mathbf{r}')} \right)^{P+1},$$

provided the conditions

$$|\mathbf{r} - \mathbf{r}_c| < D_{\ell\ell}^*(\mathbf{r}_c, \mathbf{r}'), \quad \tau^*(z_c^l - d_{\ell'}) > 0 \quad (4.140)$$

hold.
It has been pointed out that the M2M and L2L translations are lossless, hence in the convergence analysis of M2L (4.73), we can treat it as a directly successor of an ME. We refer to Theorem 4.13 with

\[
\begin{align*}
\mathbf{r} &= (x_l^I - x_c^I, y_l^I - y_c^I, \tau^*(z_l^I - d_l^I) + \tau^*(z_c^I - d_c^I)), \\
\mathbf{r}_1 &= (x - x_l^I, y - y_l^I, \tau^*(z - z_l^I)), \\
\mathbf{r}_2 &= (x_c^I - x', y_c^I - y', \tau^*(z' - z_c^I)),
\end{align*}
\]

so that the M2L translation is interpreted as a double truncation for \( n \leq P \) and for \( \nu \leq P \). Indeed, in the proof of Lemma 4.11 for double expansions in the \((\eta, t)\) domain, the single truncation on \( n \) brings the LE, while the double truncation forms the M2L translation coefficients. The truncation error is given by

\[
\left| u_{\ell_1 \ell_2}^*(\mathbf{r}; \mathbf{r}') - \sum_{n=0}^{P} \sum_{m=-n}^{n} K_{nm}(\mathbf{r}, \mathbf{r}_c^I) \left( \sum_{\nu=0}^{P} \sum_{\mu=-\nu}^{\nu} A_{nm;\nu\mu}(\mathbf{r}_c^I, \mathbf{r}_c^I) M_{\nu\mu}(\mathbf{r}', \mathbf{r}_c^I) \right) \right| \\
\leq \frac{1}{2\pi} \left( \frac{|\mathbf{r}' - \mathbf{r}_c^I|}{D_{\ell_1 \ell_2}^*(\mathbf{r}_c^I, \mathbf{r}_c)} - |\mathbf{r} - \mathbf{r}_c^I| - |\mathbf{r}' - \mathbf{r}_c^I| \right)^{P+1}
\]

provided the conditions

\[
|\mathbf{r} - \mathbf{r}_c^I| + |\mathbf{r}' - \mathbf{r}_c^I| < D_{\ell_1 \ell_2}^*(\mathbf{r}_c^I, \mathbf{r}_c), \quad \tau^*(z_c^I - d_c^I) > 0, \quad \tau^*(z_l^I - d_l^I) > 0
\]

hold. The estimate for M2L is slightly better than the free-space version described in [35], albeit in the FMM algorithm with fixed boxes they are equivalently used.

4.8. Conclusion

In this chapter, we have presented a fast multipole method for the efficient calculation of the interactions between charged particles embedded in 3-D layered media. The layered media Green’s function of the Laplace’s equation is decomposed into a free-space part and
four types of reaction field components. The far-field expansions are derived for the reaction field components with proof of exponential convergence, which consist the FMM together with the concept of polarization sources. The numerical tests validate both the efficiency and the accuracy of the proposed FMM. For systems of large number of charges, the computational cost from the reaction components is only a fraction of that of the FMM for the free space components. Therefore, computing the interactions of many sources in layered media basically costs the same as that for the interactions in the free space, and the proposed FMM scales as in terms of the number of charges in a layered medium and in terms of the number of layers.

For the future work, we will consider applications of the FMM in capacitance extraction of interconnects in VLSI.
A matrix basis formulation for the dyadic Green’s functions of the Maxwell’s equations and the elastic wave equation in layered media

The content in this chapter has been submitted for publication in collaboration with Bo Wang and Wei Cai under the title:


5.1. Introduction

Layered media dyadic Green’s functions (LMDG) of the Maxwell’s equations and the elastic wave equation are commonly used in the integral equation methods for studying wave fields in layered media [16, 8, 12]. These Green’s functions are $3 \times 3$ tensors which satisfy the wave equations and their variants with certain physical transmission conditions across interfaces between layers. When different materials are present, e.g., in a solid-fluid setting, an acoustic wave originating from a source in a non-viscous fluid will be transmitted into elastic waves in a neighboring solid layer. In such cases, a Green’s function in terms of the pressure in the fluid region, representing a compression P-wave, will be used, while in the solid layer a dyadic Green’s function for the 3-dimensional displacement vector will be used, involving an additional S-wave.

A naive derivation of the Green’s functions will have 9 unknown entries of the $3 \times 3$ tensor in each layer whereas the transmission conditions will tangle all the entries together. However, some of the entries are in fact linearly dependent or even identical. To simplify the
derivation and reduce computational cost, a number of formulations have been proposed. For the Maxwell’s LMDG, previous work includes the Sommerfeld potential [71], the transverse potential [31, 57] as well as the Michalski–Zheng formulations [56], the $E_z-H_z$ formulation [48, 17], etc. The Sommerfeld potential and the transverse potential formulations reduce the number of unknowns to 5 while the $E_z-H_z$ approach uses merely 2 scalar variables, based on a TE/TM mode decomposition. For the elastic wave equations, the dyadic Green’s function for the half-space problem was discussed in [33].

The purpose of this chapter is to present a theoretically sound general matrix representation of the $3 \times 3$ LMDG of the Maxwell’s equations and of the elastic wave equation using a linear matrix basis, providing an alternative formulation, in comparison with some previously known results [48, 17]. Mathematical theories are developed to justify the representation for both the electromagnetic and elastic waves. It will be shown that there are several remarkable benefits resulting from the matrix basis formulation (MBF). First, the coefficients of the matrix basis are all radial symmetric in the horizontal directions, so that the evaluation of the reflection/transmission coefficients in the layers are simplified. Second, the Maxwell’s Green’s functions can be naturally decomposed into independent TE and TM components within this formulation, leading to the 2-term $E_z-H_z$ result [48, 17]. Meanwhile, the elastic wave Green’s function is decomposed into S-wave components and P-wave components with corresponding matrix rows. Third, the radial symmetry allows us to apply the FMM in 3-D layered media with minimal modification. We also develop a vector basis formulation which is simplified from the matrix version, used for the LMDG of the mixed-phase elastic wave equations where the source originates in fluid medium.

The rest of this chapter is organized as follows. In Section 5.2, we establish the theories of the matrix basis and propose the MBF. In Section 5.3, the details of the Maxwell’s LMDG are derived, including a 5-term matrix-based general formulation and the concise 2-term formulation. In Section 5.4, the electric and magnetic field LMDGs for a 10-layer problem are numerically calculated using the formulation proposed in this chapter. In Section 5.5,
the details of the LMDG of the elastic wave equation with sources from liquid layers and fluid layers are discussed. A sketch of the FMM and the conclusion are given in the ending sections.

5.2. A matrix basis formulation

In this section, we set up the matrix basis used for the LMDG of the Maxwell’s equations and the elastic wave equations, and develop basic theories regarding the basis coefficients.

We consider the interaction between a given source \( r' = (x', y', z') \) in layer \( \ell' \) and a target \( r = (x, y, z) \) in layer \( \ell \). Parameters of the layered structure are denoted following the convention of previous chapters in this thesis.

Let \( \{k_i\}_{i=1}^{I} \) be all the wave numbers in the layers. Note that there are distinct elastic wave numbers for the S-wave and the P-wave in each solid layer.

The 2-D Fourier transform (3.6) is used for the representation of functions in the frequency domain.

5.2.1. Function fields for the LMDGs

We will construct some fields of functions that will be helpful in representing LMDGs. For the sake of convenience, define

\[
R_0 = [0, \infty) \setminus \{k_i : 1 \leq i \leq I\}. \tag{5.1}
\]

Using the field of complex numbers \( \mathbb{C} \), define

\[
\mathbb{F}_0 = \left\{ f_0(k_\rho) = \frac{p_0(s_i, e_{ij}, e_{iz}, e_{iz'})}{q_0(s_i, e_{ij}, e_{iz}, e_{iz'})} : p_0, q_0 \in \mathbb{P}[\mathbb{C}], q_0(s_i, e_{ij}, e_{iz}, e_{iz'}) \neq 0 \right\}, \tag{5.2}
\]

where

\[
s_i = i \sqrt{k_i^2 - k_\rho^2}, \quad e_{ij} = e^{i \sqrt{k_i^2 - k_\rho^2} j}, \quad e_{iz} = e^{i \sqrt{k_i^2 - k_\rho^2} z}, \quad e_{iz'} = e^{i \sqrt{k_i^2 - k_\rho^2} z'} \tag{5.3}
\]
are continuous complex functions of $k_\rho$ defined on $R_0$, in (5.2) indices ranging in $1 \leq i \leq I$, $0 \leq j \leq L - 1$ are considered, and $\mathbb{P}[\mathbb{C}]$ is the collection of polynomials with coefficients in $\mathbb{C}$. Although the natural domain of any function $f_0$ in $\mathbb{F}_0$ may exclude zeros of the denominator in the definition of (5.2), we claim that the domain of $f_0$ can always be extended to $R_0$ while ranging in the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$.

Consider an analytic extension of the denominator $q_0(s_i, e_{ij}, e_{iz}, e_{iz'})$ as a function of $k_\rho$ with a series of branch cuts (see Figure 5.1)

$$B = \bigcup_{i=1}^{I} \{\pm (k_i - \ell i) : \ell \in [0, \infty)\} \quad (5.4)$$

which leaves the remaining part of the complex plane $\mathbb{C} \setminus B$ connected. By the identity theorem of the complex analysis, we conclude that the denominator is either identical to 0 (which is excluded from the definition), or having isolated zeros in $\mathbb{C} \setminus B$ (hence also isolated in $R_0$). The same properties are also applied to the numerator $p_0(s_i, e_{ij}, e_{iz}, e_{iz'})$. Therefore, we can extend the domain of $f_0$ to $R_0$ by taking the limit from the neighborhoods of each zero.

When treating any element of $\mathbb{F}_0$ as a mapping from $R_0$ to $\overline{\mathbb{C}}$, $\mathbb{F}_0$ forms a field. The proof is trivial.

**Proposition 5.1 (Field structure of $\mathbb{F}_0$).** $\mathbb{F}_0$ is a field of functions mapping $R_0$ to $\overline{\mathbb{C}}$, with function addition and multiplication, where the constant functions 0 and 1 are the
additive identity and the multiplicative identity, respectively.

It is worth mentioning that $k_\rho^2 \in \mathbb{F}_0$ because

$$k_\rho^2 = k_i^2 + s_i^2. \quad (5.5)$$

Also, we can treat elements of $\mathbb{F}_0$ as functions of $k_x$ and $k_y$, with the relationship $k_\rho = \sqrt{k_x^2 + k_y^2}$.

Next, define

$$\mathbb{F} = \left\{ f(ik_x, ik_y) = \frac{p(ik_x, ik_y)}{q(ik_x, ik_y)} : p, q \in \mathbb{P}[\mathbb{F}_0], q(ik_x, ik_y) \not\equiv 0 \right\}, \quad (5.6)$$

where $\mathbb{P}[\mathbb{F}_0]$ is the collection of polynomials with coefficients in $\mathbb{F}_0$. It is clear that $\mathbb{F}$ is a field extension of $\mathbb{F}_0$ with elements $ik_x$ and $ik_y$. We will deduce similar results of $\mathbb{F}$, but with the polar coordinates. Given any $f \in \mathbb{F}$, let $D_f \subset \mathbb{C} \setminus B$ be the set of any $k_\rho$ that makes $\infty$ exist in coefficients of polynomials $p$ and $q$. $D_f$ is known as a discrete set according to the previous discussion on $\mathbb{F}_0$. While also excluding the branch cuts $B$ defined in (5.4) from the complex plane, we have an analytic extension of

$$q(ik_x, ik_y) = q(ik_\rho \cos \alpha, ik_\rho \sin \alpha) := \tilde{q}(k_\rho, \alpha) \quad (5.7)$$

for $(k_\rho, \alpha)$ in the open connected domain $\mathbb{C} \setminus (B \cup D_f) \times \mathbb{C}$. A simple generalization of the identity theorem is sufficient to tell that $\tilde{q}(k_\rho, \alpha)$ is either identical to 0 or having isolated zeros in this domain. Therefore, similarly, any function $f \in \mathbb{F}$ can be extended to a mapping from $(k_\rho, \alpha) \in R_0 \times [0, 2\pi)$ to $\overline{\mathbb{C}}$, and $\mathbb{F}$ is a field in this sense.

Remark 5.2. The coordinates $z$ and $z'$ are in fact redundant in the definition of $\mathbb{F}_0$ and $\mathbb{F}$ for the matrix basis theory. They are included only for the convenience of statements in later sections.
5.2.2. The matrix basis

One of our expectations for the matrix basis is to represent the tensor Green’s functions with all coefficients belonging to \( F_0 \), i.e., information of the polar angle \( \alpha \) would only appear in the matrix basis. For this purpose, based on an observation of derived formulas for the Maxwell’s Green’s functions in a 3-layer problem [19], we propose a matrix basis \( J_1, \cdots, J_9 \) in the frequency domain as follows.

**Proposition 5.3 (The matrix basis).** These matrices form a basis of \( F^{3 \times 3} \):

\[
\begin{align*}
J_1 &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, &
J_2 &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, &
J_3 &= \begin{bmatrix} 0 & 0 & i k_x \\ 0 & 0 & i k_y \\ 0 & 0 & 0 \end{bmatrix}, \\
J_4 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ i k_x & i k_y & 0 \end{bmatrix}, &
J_5 &= \begin{bmatrix} -k_x^2 & -k_x k_y & 0 \\ -k_x & -k_y & 0 \\ 0 & 0 & 0 \end{bmatrix}, &
J_6 &= \begin{bmatrix} 0 & 0 & 0 \\ -i k_y & i k_x & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
J_7 &= \begin{bmatrix} 0 & 0 & i k_y \\ 0 & 0 & -i k_x \\ 0 & 0 & 0 \end{bmatrix}, &
J_8 &= \begin{bmatrix} k_x k_y & k_y^2 & 0 \\ -k_x k_y & k_y & 0 \\ -k_x^2 & -k_x k_y & 0 \end{bmatrix}, &
J_9 &= \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\end{align*}
\]

The proof is trivial.

The above matrix basis will be divided into two groups according to the product properties between them. For any subfield \( \mathbb{K} \subset \mathbb{F} \), we define the following vector spaces with

136
coefficients in \( K \)

\[
\mathcal{R}(K) = \text{span}_K(J_1, \ldots, J_5),
\]

\[
\mathcal{I}(K) = \text{span}_K(J_6, \ldots, J_9),
\]

\[
\mathcal{M}(K) = \text{span}_K(J_1, \ldots, J_5, J_6, \ldots, J_9).
\] (5.9)

**Proposition 5.4.** Let \( K \) be any subfield of \( F \) containing \( k^2 _\rho \). Then, we have

- \( \mathcal{M}(K) = \mathcal{R}(K) \oplus \mathcal{I}(K) \) is the direct sum.
- \( \mathcal{R}(K), \mathcal{M}(K) \) are rings with matrix addition and matrix multiplication.

**Proof.** The direct sum is obvious. For the ring property, notice the identity matrix \( I = J_1 + J_2 \in \mathcal{R}(K) \subset \mathcal{M}(K) \), and the product table of the matrices \( J_1, \ldots, J_9 \)

\[
\begin{bmatrix}
J_1^T & \cdots & J_9^T
\end{bmatrix}^T \cdot \begin{bmatrix}
J_1 & \cdots & J_9
\end{bmatrix} =
\begin{bmatrix}
J_1 & 0 & J_3 & 0 & J_5 & 0 & J_7 & J_8 & J_9
0 & J_2 & 0 & J_4 & 0 & J_6 & 0 & 0 & 0
0 & J_3 & 0 & J_5 & 0 & J_8 - k^2 _\rho J_9 & 0 & 0 & 0
J_4 & 0 & -k^2 _\rho J_2 & 0 & -k^2 _\rho J_4 & 0 & 0 & 0 & J_6
J_5 & 0 & -k^2 _\rho J_3 & 0 & -k^2 _\rho J_5 & 0 & 0 & 0 & J_8 - k^2 _\rho J_9
J_6 & 0 & 0 & 0 & 0 & J_9 & k^2 _\rho J_2 & -k^2 _\rho J_4 & -J_4
0 & J_7 & 0 & -J_8 & 0 & k^2 _\rho J_1 + J_5 & 0 & 0 & 0
J_8 & 0 & k^2 _\rho J_7 & 0 & -k^2 _\rho J_8 & 0 & 0 & 0 & -k^2 _\rho J_1 - J_5
J_9 & 0 & J_7 & 0 & -J_8 & 0 & -J_3 & J_5 & -J_1
\end{bmatrix}
\] (5.10)

which ensures the matrix multiplication is closed in both \( \mathcal{M}(K) \) and \( \mathcal{R}(K) \). \( \square \)
Proposition 5.5 (The product rules). Let $K$ be any subfield of $F$ containing $k^2$.

- If $A \in \mathcal{R}(K)$, $B \in \mathcal{R}(K)$, then $A \cdot B \in \mathcal{R}(K)$.
- If $A \in \mathcal{R}(K)$, $B \in \mathcal{I}(K)$, then $A \cdot B \in \mathcal{I}(K)$.
- If $A \in \mathcal{I}(K)$, $B \in \mathcal{R}(K)$, then $A \cdot B \in \mathcal{I}(K)$.
- If $A \in \mathcal{I}(K)$, $B \in \mathcal{I}(K)$, then $A \cdot B \in \mathcal{R}(K)$.

The behavior in the product rules resembles the products between the real numbers and the pure imaginary numbers, which explains the adopted letters $\mathcal{R}$ and $\mathcal{I}$.

Definition 5.6 (An $\mathcal{R}^0$-matrix basis formulation). Define the linear space

$$\mathcal{R}^0 = \text{span}_{F_0} \{J_1, \cdots, J_5\} = \left\{ \sum_{j=1}^{5} a_j J_j : a_j \in F_0 \right\}. \quad (5.11)$$

The linear expansion of functions in $\mathcal{R}^0$ with matrix basis $J_1, \cdots, J_5$ will be defined as the $\mathcal{R}^0$-matrix basis formulation.

In the coming sections, we will show that this MBF can be used to efficiently express the Green’s functions for the Maxwell’s equations in layered media. Since the tensors Green’s functions are coupled between the layers, we introduce the framework and theories of the block matrices.

For any subfield $K \subset F$ and any $p, q \in \mathbb{N}$, define the linear spaces of block matrices

$$\mathcal{M}_{p \times q}(K) = \left\{ \sum_{j=1}^{9} K_j \otimes J_j : K_j \in K^{p \times q}, 1 \leq j \leq 9 \right\},$$

$$\mathcal{R}_{p \times q}(K) = \left\{ \sum_{j=1}^{5} K_j \otimes J_j : K_j \in K^{p \times q}, 1 \leq j \leq 5 \right\},$$

$$\mathcal{I}_{p \times q}(K) = \left\{ \sum_{j=6}^{9} K_j \otimes J_j : K_j \in K^{p \times q}, 6 \leq j \leq 9 \right\}. \quad (5.12)$$
where $\otimes$ is the Kronecker product. Any $\sum_{j=1}^{9} K_j \otimes J_j \in \mathcal{M}_{p\times q}(\mathbb{K})$ is a $3p \times 3q$ matrix consists of $3 \times 3$ blocks in $\mathcal{M}(\mathbb{K})$. By applying the direct sum decomposition from Proposition 5.4 to each $3 \times 3$ block, we get the direct sum decomposition of the block matrices

$$\mathcal{M}_{p\times q}(\mathbb{K}) = \mathcal{R}_{p\times q}(\mathbb{K}) \oplus \mathcal{I}_{p\times q}(\mathbb{K}).$$

(5.13)

Moreover, the product rules for block matrices are easily generalized as follows.

**Proposition 5.7 (The product rules for block matrices).** Let $p, q, r \in \mathbb{N}$. Let $\mathbb{K}$ be any subfield of $\mathbb{F}$ containing $k^2$.

- If $\bar{A} \in \mathcal{R}_{p\times r}(\mathbb{K})$, $\bar{B} \in \mathcal{R}_{r\times q}(\mathbb{K})$, then $\bar{A} \cdot \bar{B} \in \mathcal{R}_{p\times q}(\mathbb{K})$.
- If $\bar{A} \in \mathcal{R}_{p\times r}(\mathbb{K})$, $\bar{B} \in \mathcal{I}_{r\times q}(\mathbb{K})$, then $\bar{A} \cdot \bar{B} \in \mathcal{I}_{p\times q}(\mathbb{K})$.
- If $\bar{A} \in \mathcal{I}_{p\times r}(\mathbb{K})$, $\bar{B} \in \mathcal{R}_{r\times q}(\mathbb{K})$, then $\bar{A} \cdot \bar{B} \in \mathcal{I}_{p\times q}(\mathbb{K})$.
- If $\bar{A} \in \mathcal{I}_{p\times r}(\mathbb{K})$, $\bar{B} \in \mathcal{I}_{r\times q}(\mathbb{K})$, then $\bar{A} \cdot \bar{B} \in \mathcal{R}_{p\times q}(\mathbb{K})$.

**Proof.** We will only take the second proposition as an example. Suppose

$$\bar{A} = \sum_{j=1}^{5} A_j \otimes J_j, \quad \bar{B} = \sum_{l=6}^{9} B_l \otimes J_l, \quad A_j \in \mathbb{K}^{p \times r}, \quad B_l \in \mathbb{K}^{r \times q},$$

(5.14)

The matrix product is given by

$$\bar{A} \cdot \bar{B} = \sum_{j=1}^{5} \sum_{l=6}^{9} (A_j B_l) \otimes (J_j J_l),$$

(5.15)

where each $A_j B_l \in \mathbb{K}^{p \times q}$ and each $J_j J_l \in \mathcal{I}(\mathbb{K})$ by the product table (5.10). Therefore, $\bar{A} \cdot \bar{B}$ is a $3p \times 3q$ block matrix with every block in $\mathcal{I}(\mathbb{K})$. □

**Proposition 5.8.** For any $p, q \in \mathbb{N}$, $\mathcal{M}_{p\times q}(\mathbb{F}) = \mathbb{F}^{3p \times 3q}$.

The proof is trivial with the fact that $J_1, \ldots, J_9$ form a basis of $\mathbb{F}^{3 \times 3}$. 139
5.2.3. The solution filtering theorem

Just like the fact that any solvable real linear system with a complex solution must also have a real one, the following theorem claims that when the coefficient matrices have the $\mathfrak{R}_0$-matrix basis representation for each of their $3 \times 3$ blocks, there exists a solution with an $\mathfrak{R}_0$-matrix basis representation as well, when solvable.

We begin with a well-known lemma for linear systems of fields.

**Lemma 5.9 (Solution filtering of fields).** Suppose $p, q, r \in \mathbb{N}$, $K$ is a field, $A \in K^{p \times r}$, $B \in K^{p \times q}$ are coefficients of a solvable linear system, i.e. $A \cdot X = B$ for some $r \times q$ matrix $X$. Then, there exists a “filtered” solution $X_0 \in K^{r \times q}$ such that $A \cdot X_0 = B$.

**Proof.** Let $a \leq \min(p, r)$ be the rank of $A$. By applying elementary row and column operations we can transform $A$ in its reduced row echelon form, then into a 0-1 matrix with $I_a$ on the top left corner while all other entries are zero. Hence there exist full-rank matrices $S \in K^{p \times p}$ and $T \in K^{r \times r}$ such that

$$A = S \cdot \begin{bmatrix} I_a & 0_{a \times (r-a)} \\ 0_{(p-a) \times a} & 0_{(p-a) \times (r-a)} \end{bmatrix} \cdot T. \tag{5.16}$$

It immediately follows that $S^{-1}B \in K^{p \times q}$ is a matrix whose bottom $(p-a)$ rows are all zero. Next, we define

$$X_0 = T^{-1} \cdot \begin{bmatrix} I_a & 0_{a \times (p-a)} \end{bmatrix} \cdot (S^{-1}B) \in K^{r \times q}, \tag{5.17}$$

then we have

$$AX_0 = S \begin{bmatrix} I_a & 0_{a \times (p-a)} \end{bmatrix} \cdot (S^{-1}B) \tag{5.18}$$
and
\[ S^{-1}AX_0 = \begin{bmatrix} I_a & 0_{a \times (p-a)} \\ 0_{(p-a) \times q} \end{bmatrix} \cdot (S^{-1}B). \] (5.19)

We can see that \( S^{-1}AX_0 \) has its top \( a \) rows identical to those of \( S^{-1}B \), and its bottom \((p-a)\) rows are zero entries (namely, exactly as \( S^{-1}B \) as noted above). Therefore, \( S^{-1}AX_0 \) and \( S^{-1}B \) are identical, so \( AX_0 = B \).

The main result of this section is an improved version which preserves the matrix basis structure.

**Theorem 5.10 (Solution filtering).** Suppose \( p, q, r \in \mathbb{N}, \bar{A} \in \mathcal{R}_{p \times r}(\mathbb{F}_0) \) and \( \bar{B} \in \mathcal{R}_{p \times q}(\mathbb{F}_0) \) are coefficients of a solvable linear system, i.e. \( \bar{A} \cdot \bar{X} = \bar{B} \) for some \( 3r \times 3q \) matrix \( \bar{X} \). Then, there exists a “filtered” version of block matrix solution \( \bar{X}_0 \in \mathcal{R}_{r \times q}(\mathbb{F}_0) \), i.e., each \( 3 \times 3 \) block of \( \bar{X}_0 \) has an \( \mathcal{R}^0 \)-matrix basis representation, so that \( \bar{A} \cdot \bar{X}_0 = \bar{B} \).

**Proof.** By Proposition 5.8, \( \bar{A} \in \mathbb{F}^{3p \times 3r} \) and \( \bar{B} \in \mathbb{F}^{3p \times 3q} \). By Lemma 5.9, we can suppose \( \bar{X} \in \mathbb{F}^{3r \times 3q} = \mathcal{M}_{r \times q}(\mathbb{F}) \) without loss of generality.

To find the filtered solution in \( \mathcal{R}_{r \times q}(\mathbb{F}_0) \), we begin with an intermediate one in \( \mathcal{R}_{r \times q}(\mathbb{F}) \). Write the direct sum decomposition \( \bar{X} = \bar{X}_1 \oplus \bar{X}_2 \), where \( \bar{X}_1 \in \mathcal{R}_{r \times q}(\mathbb{F}) \) and \( \bar{X}_2 \in \mathcal{I}_{r \times q}(\mathbb{F}) \). By Lemma 5.7 we immediately get \( \bar{A} \cdot \bar{X}_1 \in \mathcal{R}_{p \times q}(\mathbb{F}) \) and \( \bar{A} \cdot \bar{X}_2 \in \mathcal{I}_{p \times q}(\mathbb{F}) \), so

\[ \bar{A} \cdot \bar{X}_1 + \bar{A} \cdot \bar{X}_2 \] (5.20)

is the direct sum decomposition of \( \bar{B} \in \mathcal{R}_{p \times q}(\mathbb{F}_0) \subset \mathcal{M}_{p \times q}(\mathbb{F}) \). Therefore \( \bar{A} \cdot \bar{X}_1 = \bar{B} \).

Then, let

\[ \bar{A} = \sum_{j=1}^{5} A_j \otimes J_j, \quad \bar{X}_1 = \sum_{j=1}^{5} X^1_j \otimes J_j, \quad \bar{B} = \sum_{j=1}^{5} B_j \otimes J_j, \] (5.21)

where each \( A_j \in \mathbb{F}_0^{p \times r} \), \( X^1_j \in \mathbb{F}^{r \times q} \) and \( B_j \in \mathbb{F}_0^{p \times q} \). When treating \( X^1_j \) as the solution to the
linear equation $\bar{A} \cdot \bar{X}_1 = \bar{B}$, the equation is equivalent to

$$\sum_{u=1}^{5} \sum_{v=1}^{5} (A_u X_v^1) \otimes (J_u J_v) = \sum_{j=1}^{5} B_j \otimes J_j,$$

(5.22)

where, using the product table (5.10), among the 25 products only 13 of them will be nonzero to give

$$\sum_{u=1}^{5} \sum_{v=1}^{5} (A_u X_v^1) \otimes (J_u J_v) = (A_1 X_1^1) \otimes J_1 + (A_2 X_2^1 - k_p^2 A_4 X_3^1) \otimes J_2$$

$$+ (A_3 X_2^1 + (A_1 - k_p^2 A_5) X_3^1) \otimes J_3 + (A_4 X_1^1 + A_2 X_4^1 - k_p^2 A_4 X_5^1) \otimes J_4$$

$$+ (A_5 X_1^1 + A_3 X_4^1 + (A_1 - k_p^2 A_5) X_5^1) \otimes J_5.$$

(5.23)

By comparing the coefficients of $J_1, \ldots, J_5$ on each $3 \times 3$ block, the above equation (5.22) is in fact equivalent to the linear system

$$\bar{A} \bar{X}^1 = \bar{B},$$

(5.24)

where a stacked form of $\bar{X}^1$ given below is used

$$\bar{A} = \begin{bmatrix}
A_1 & 0 & 0 & 0 & 0 \\
0 & A_2 & -k_p^2 A_4 & 0 & 0 \\
0 & A_3 & A_1 - k_p^2 A_5 & 0 & 0 \\
A_4 & 0 & 0 & A_2 & -k_p^2 A_4 \\
A_5 & 0 & 0 & A_3 & A_1 - k_p^2 A_5
\end{bmatrix}, \quad \bar{X}^1 = \begin{bmatrix}
X_1 \\
\vdots \\
X_5
\end{bmatrix}, \quad \bar{B} = \begin{bmatrix}
B_1 \\
\vdots \\
B_5
\end{bmatrix}.$$  

(5.25)

Since $\bar{A} \in \mathbb{F}_0^{5p \times 5r}$ and $\bar{B} \in \mathbb{F}_0^{5p \times q}$, by Lemma 5.9, there exists $\bar{X}^0 \in \mathbb{F}^{5r \times q}$ such that $\bar{A} \bar{X}^0 = \bar{B}$. 

142
By writing \( \tilde{X}^0 \) in the stacked form

\[
\tilde{X}^0 = \begin{bmatrix}
X_0^0 \\
\vdots \\
X_5^0
\end{bmatrix}
\]

(5.26)

where each \( X_j^0 \in \mathbb{F}_0^{r \times q} \), we can see that the matrix

\[
\bar{X}_0 = \sum_{j=1}^{5} X_j^0 \otimes J_j \in \mathfrak{M}_{r \times q}(\mathbb{F}_0)
\]

(5.27)

is as desired.

The coming discussion in later sections on the LMDG of the Maxwell’s equations and the elastic wave equation are generally presented in the following route: the required mathematical conditions of the problem, including linear equations derived from the interface conditions and the radiation conditions, are re-formatted using the matrix basis \( J_1, \ldots, J_9 \). The solution filtering theorem then works on the linear system of the tensors in the layers, so that filtered solutions are proven to exist using the matrix basis formulation of \( \mathfrak{M}_0 \). Finally, the formulation helps simplify the required mathematical conditions, by specifying only corresponding basis coefficients.

5.3. Application to the Maxwell’s equations in layered media

In this section we first give a brief introduction about the dyadic Green’s functions of the time-harmonic Maxwell’s equations in the free space, then we will discuss the LMDG and its simplification using the \( \mathfrak{M}_0 \)-matrix basis formulation (5.11).
5.3.1. The dyadic Green’s functions of the time harmonic Maxwell’s equations in the free space

The source-free time harmonic Maxwell’s equations in the free space is given by

\[
\nabla \times \vec{E} = -i\omega \vec{B}, \quad \nabla \times \vec{H} = i\omega \vec{D},
\]

\[
\nabla \cdot \vec{D} = 0, \quad \nabla \cdot \vec{B} = 0,
\]

(5.28)

where \( \vec{D}(r), \vec{E}(r) \) are the electric displacement flux and the electric field, \( \vec{B}(r), \vec{H}(r) \) are the magnetic flux density and the magnetic field, \( \omega \) is the angular frequency in time. The system of Maxwell’s equations is closed by the constitutive relations. In the free space, they are

\[
\vec{D} = \varepsilon \vec{E}, \quad \vec{B} = \mu \vec{H},
\]

(5.29)

where \( \varepsilon \) and \( \mu \) are the constant permittivity and permeability in the free space, respectively.

When dealing with these equations, the Lorenz gauge condition is often used, which allows us to use a vector potential \( \vec{A}(r) \) to represent the electric field \( \vec{E} \) and the magnetic field \( \vec{H} \) as

\[
\vec{E} = -i\omega \left( I + \frac{\nabla \nabla}{k^2} \right) \vec{A}, \quad \vec{H} = \frac{1}{\mu} \nabla \times \vec{A},
\]

(5.30)

where \( I \) is the \( 3 \times 3 \) identity matrix, \( k \) is the wave number defined as

\[
k = \sqrt{\omega^2 \varepsilon \mu}.
\]

(5.31)

From the Maxwell’s equations (5.28), the constitutive relations (5.29) and the Lorenz gauge condition, one can show that \( \vec{A} \) satisfies the Helmholtz equation

\[
\nabla^2 \vec{A} + k^2 \vec{A} = \vec{0}.
\]

(5.32)

The choice of the vector potential is not unique. Indeed, for any function \( \phi \in C^2(\mathbb{R}^3) \)
satisfying the Helmholtz equation $\nabla^2 \phi + k^2 \phi = 0$, we can replace $\vec{A}$ by $\vec{A} + \nabla \phi$ in (5.30) to give exactly the same $\vec{E}$ and $\vec{H}$.

The dyadic Green’s functions for the free space Maxwell’s equations are defined using a $3 \times 3$ potential tensor $G^f_A(r; r')$ such that the electric field dyadic Green’s function $G_E(r; r')$ and the magnetic field dyadic Green’s function $G_H(r; r')$ are represented by

$$G_E = -i \omega \left( I + \frac{\nabla \nabla}{k^2} \right) G^f_A, \quad G_H = \frac{1}{\mu} \nabla \times G^f_A.$$  \hfill (5.33)

According to the equation (5.32) for the vector potential $\vec{A}$, the potential tensor $G^f_A$ is defined as the solution of the Helmholtz equation

$$\nabla^2 G^f_A + k^2 G^f_A = \frac{1}{i \omega} \delta(r - r') I,$$  \hfill (5.34)

with the Silver–Müller radiation conditions [8]

$$|\hat{r} \times \nabla \times G_E(r; r') - i k G_E(r; r')| = O(r^{-2}),$$

$$|\hat{r} \times \nabla \times G_H(r; r') - i k G_H(r; r')| = O(r^{-2}),$$  \hfill (5.35)

as $r = |r| \rightarrow \infty$. Here, $\delta(r)$ is the Dirac Delta function and $\hat{r}$ is the unit direction along $r$.

For the same reason mentioned above, $G^f_A$ is not unique. A commonly used solution to (5.34) is given by

$$G^f_A(r; r') = -\frac{1}{i \omega} \frac{e^{ik|r-r'|}}{4\pi|r-r'|} I = -\frac{1}{i \omega} g^f(r; r') I,$$  \hfill (5.36)

where $g^f(r; r')$ is the free-space Green’s function of the 3-D Helmholtz equation (3.4). It is easy to verify that the $G_E$ and $G_H$ resulted from this $G^f_A$ satisfy the Silver–Müller radiation conditions (5.35).
5.3.2. The dyadic Green’s functions of the time harmonic Maxwell’s equations in layered media

Now suppose the space is horizontally stratified as \( L + 1 \) layers, where we adopt the common notations for the layer structure throughout this thesis. Suppose each layer is homogeneous with constant permittivity \( \varepsilon_\ell \) and constant permeability \( \mu_\ell, \ell = 0, \cdots, L \), respectively. In each layer, the wave number is defined by

\[
k_\ell = \sqrt{\omega^2 \varepsilon_\ell \mu_\ell}.
\]  

(5.37)

5.3.2.1. The equations in the spatial domain

The governing equations in the interior of each layer has the same form as in (5.28)–(5.29), while the following transmission conditions must be satisfied [8] between adjacent layers

\[
[n \times \vec{E}] = 0, \quad [n \cdot \vec{D}] = 0, \quad [n \times \vec{H}] = 0, \quad [n \cdot \vec{B}] = 0.
\]  

(5.38)

\([\cdot]\) is used to represent the jump of the value at the interface, i.e. across the interface \( z = d \),

\[
[f] = \lim_{z \to d^+} f - \lim_{z \to d^-} f.
\]  

(5.39)

The dyadic Green’s functions are again given by the tensor potential \( \mathbf{G}_A \) as

\[
\mathbf{G}_E = -i\omega \left( \mathbf{I} + \frac{\nabla \nabla}{k^2} \right) \mathbf{G}_A, \quad \mathbf{G}_H = \frac{1}{\mu} \nabla \times \mathbf{G}_A.
\]  

(5.40)

The tensor potential \( \mathbf{G}_A \) satisfies the Helmholtz equation

\[
\nabla^2 \mathbf{G}_A + k^2 \mathbf{G}_A = \frac{1}{i\omega} \delta(\mathbf{r} - \mathbf{r}') \mathbf{I},
\]  

(5.41)
while the interface conditions are

\[
\begin{align*}
[n \times G_E] &= 0, \quad [\varepsilon n \cdot G_E] = \vec{0}^T, \quad [n \times G_H] = 0, \quad [\mu n \cdot G_H] = \vec{0}^T.
\end{align*}
\] (5.42)

In horizontally layered media, \( n = e_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \). In addition, the Green’s functions \( G_E \) and \( G_H \) must satisfy upward/downward outgoing radiation conditions [11].

5.3.2.2. The equations in the frequency domain

We will begin with the separation of the \( z \) variable from the tensor potential \( \hat{G}_A \) in the frequency domain of the 2-D Fourier transform, which will lead to a reaction field decomposition. The Fourier transform is taken from \( (x - x', y - y') \) to \( (k_x, k_y) \), as was introduced in (3.6). Notations of the polar coordinate pair \( (k_\rho, \alpha) \) are retained.

In the frequency domain, the gradient operator \( \nabla \) becomes a vector with partial derivative only on the \( z \) coordinate

\[
\hat{\nabla} = \begin{bmatrix} ik_x \\ ik_y \\ \partial_z \end{bmatrix},
\] (5.43)

if followed by a function of \( z \). Operators \( \hat{\nabla} \hat{\nabla} \), \( \hat{\nabla}^2 \) now become \( \hat{\nabla} \hat{\nabla}^T \) and \( \hat{\nabla}^T \hat{\nabla} \), respectively.

Recalling that the right-hand side of the Helmholtz equation (5.41) is nontrivial if and only if \( r' \) is in the same layer as \( r \), i.e. \( \ell = \ell' \), we define the reaction field like in the 2-D and 3-D scalar problems (2.3) and (3.5)

\[
\hat{G}^r_A(k_x, k_y, z; z') = \hat{G}_A(k_x, k_y, z; z') - \delta_{\ell,\ell'} \hat{G}^f_A(k_x, k_y, z; z'),
\] (5.44)

where \( \hat{G}^f_A \) is the Fourier transform of the \( G^f_A \) given in (5.36), and \( \hat{G}^r_A \) is the reaction field,
which satisfies a *homogeneous* Helmholtz equation

\[ \nabla^2 \hat{G}^r_A + k^2 \hat{G}^r_A = 0, \quad \text{i.e.} \quad \partial_{zz} \hat{G}^r_A + (k^2 - k^2_{\rho}) \hat{G}^r_A = 0. \]  

(5.45)

We use the notation

\[ k_z = k_{\ell z} \]

from (3.11) when there’s no need to distinguish the target layer. The general solutions to (5.45), when treated as an ordinary differential equation of \( z \), is given by

\[ \hat{G}^r_A = e^{ik_z z} \hat{G}^r_{A,\ell}(k_x, k_y; z') + e^{-ik_z z} \hat{G}^r_{A,\ell}(k_x, k_y; z'), \]  

(5.46)

where \( \hat{G}^r_{A,\ell} \) and \( \hat{G}^r_{A,\ell} \) are piecewise constants with respect to \( z \), namely,

\[ \hat{G}^r_{A,\ell} = \hat{G}^r_{A,\ell}, \quad \hat{G}^r_{A,\ell} = \hat{G}^r_{A,\ell} \]  

(5.47)

when in layer \( \ell \).

We can also write \( \hat{G}^f_A \) in a similar form

\[ \hat{G}^f_A = \frac{-1}{2\omega k_{\ell z}} e^{ik_{\ell z}z'-z'} I + \frac{-1}{2\omega k_{\ell z}} e^{ik_{\ell z}(z-z')} I + \frac{-1}{2\omega k_{\ell z}} e^{ik_{\ell z}(z'-z)} I \]  

(5.48)

when \( z' \neq z \).

Hence we may alternatively use the expression for the Green’s function for the vector potential

\[ \hat{G}_A = e^{ik_z z} \hat{G}^r_A + e^{-ik_z z} \hat{G}^r_A \]  

(5.49)
where

\[
\hat{G}_A^\uparrow = \delta_{\ell,\ell'} 1_{\{z > z'\}} \frac{-e^{-ik_{\ell'} z'}}{2\omega k_{\ell'} z} I + \hat{G}_A^{\uparrow}, \quad \hat{G}_A^\downarrow = \delta_{\ell,\ell'} 1_{\{z < z'\}} \frac{-e^{ik_{\ell'} z'}}{2\omega k_{\ell'} z} I + \hat{G}_A^{\downarrow},
\]

(5.50)

assuming \(z \neq d_i, 0 \leq i \leq L - 1\) and \(z \neq z'\).

For now, even without using the local coordinates, we call the separation of variable \(z\) in (5.49) the reaction field decomposition of \(\hat{G}_A\). \(e^{\tau ik_z z} \hat{G}_A^\uparrow\) are called the propagation components of \(\hat{G}_A\). \(e^{\tau ik_z z} \hat{G}_A^\downarrow\) are called the reaction components of \(\hat{G}_A\). Similarly for \(\hat{G}_E\) and \(\hat{G}_H\) we will define corresponding terms later after (5.54).

Remark 5.11. In our previous work on the Helmholtz equation in layered media [76, 87, 75], the \(z'\) variable was also separated out, so that each reaction component was further decomposed according to the “propagating direction” of \(z'\).

Next, we will re-format the mathematical conditions of the tensor potential \(\hat{G}_A\), including the interface conditions and the radiation conditions, using the matrix basis \(J_1, \ldots, J_9\).

The LMDGs \(\hat{G}_E\) and \(\hat{G}_H\) in the frequency domain can be calculated via the Fourier transform of (5.40), i.e.,

\[
\hat{G}_E = -i\omega \left( I + \frac{\hat{\nabla} \hat{\nabla}}{k^2} \right) \hat{G}_A, \quad \hat{G}_H = \frac{1}{\mu} \hat{\nabla} \times \hat{G}_A.
\]

(5.51)

Moreover, the general expression (5.49) of \(\hat{G}_A\) implies that

\[
\hat{\nabla} \hat{G}_A = \hat{\nabla}^+ (e^{ik_z z} \hat{G}_A^\uparrow) + \hat{\nabla}^- (e^{-ik_z z} \hat{G}_A^\downarrow),

\hat{\nabla} \times \hat{G}_A = \hat{\nabla}^+ \times (e^{ik_z z} \hat{G}_A^\uparrow) + \hat{\nabla}^- \times (e^{-ik_z z} \hat{G}_A^\downarrow),
\]

(5.52)

where the operators \(\hat{\nabla}^\pm\) are defined as

\[
\hat{\nabla}^\pm := \begin{bmatrix} ik_x & ik_y & \pm ik_z \end{bmatrix}^T.
\]

(5.53)
Thus, the Green’s functions $\hat{G}_E$ and $\hat{G}_H$ can be represented as the linear combinations of reaction components of $\hat{G}_A$, with every coefficient matrices in $\mathbb{9}^0$:

$$\hat{G}_E = -i \omega \left( \begin{bmatrix} 1 & \frac{k^2}{k^2} J_5 + \frac{ik_z}{k^2} J_3 \end{bmatrix} e^{ik_z} \hat{G}_A^\uparrow - i \omega \left( \begin{bmatrix} 1 & \frac{k^2}{k^2} J_5 - \frac{ik_z}{k^2} J_3 \end{bmatrix} e^{-ik_z} \hat{G}_A^\downarrow \right) \right)$$

$$= -i \omega \left( \begin{bmatrix} 1 & \frac{k^2}{k^2} J_2 + \frac{1}{k^2} J_5 + \frac{ik_z}{k^2} J_3 + \frac{ik_z}{k^2} J_4 \end{bmatrix} e^{ik_z} \hat{G}_A^\uparrow \right)$$

$$- i \omega \left( \begin{bmatrix} 1 & \frac{k^2}{k^2} J_2 + \frac{1}{k^2} J_5 - \frac{ik_z}{k^2} J_3 - \frac{ik_z}{k^2} J_4 \end{bmatrix} e^{-ik_z} \hat{G}_A^\downarrow \right)$$

$$\hat{G}_H = \frac{1}{\mu} \hat{\nabla}^+ \times e^{ik_z} \hat{G}_A^\uparrow + \frac{1}{\mu} \hat{\nabla}^- \times e^{-ik_z} \hat{G}_A^\downarrow$$

$$= \frac{1}{\mu} \left( J_6 + J_7 - ik_z J_9 \right) e^{ik_z} \hat{G}_A^\uparrow + \frac{1}{\mu} \left( J_6 + J_7 + ik_z J_9 \right) e^{-ik_z} \hat{G}_A^\downarrow.$$ 

The $n \cdot \nabla$ and $n \times$ operators in interface conditions (5.42), given $n = e_3$, are converted to their equivalent matrix forms in the frequency domain, respectively, as

$$[J_1 \hat{G}_E] = 0, \quad [\varepsilon J_2 \hat{G}_E] = 0, \quad [J_9 \hat{G}_H] = 0, \quad [\mu J_7 \hat{G}_H] = 0. \quad (5.56)$$

Here we only check the first equations of (5.42) and (5.56) as an example. Let $v = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$ be any column of $\hat{G}_E$. The first equation in (5.42) is equivalent to the continuity equations of $n \times v = e_3 \times v = \begin{bmatrix} -v_2 & v_1 & 0 \end{bmatrix}^T$, while in the first equation of (5.56),

$$J_1 v = \begin{bmatrix} v_1 & v_2 & 0 \end{bmatrix}^T. \quad \text{Both of them are equivalent to the continuity equations of } v_1 \text{ and } v_2.$$ 

Specifically, in the brackets of (5.56),

$$J_1 \hat{G}_E = -i \omega \left( J_1 + \frac{1}{k^2} J_5 + \frac{ik_z}{k^2} J_3 \right) e^{ik_z} \hat{G}_A^\uparrow - i \omega \left( J_1 + \frac{1}{k^2} J_5 - \frac{ik_z}{k^2} J_3 \right) e^{-ik_z} \hat{G}_A^\downarrow,$$

$$\varepsilon J_2 \hat{G}_E = -i \omega \varepsilon \left( \frac{k^2}{k^2} J_2 + \frac{ik_z}{k^2} J_4 \right) e^{ik_z} \hat{G}_A^\uparrow - i \omega \varepsilon \left( \frac{k^2}{k^2} J_2 - \frac{ik_z}{k^2} J_4 \right) e^{-ik_z} \hat{G}_A^\downarrow,$$

$$J_9 \hat{G}_H = - \frac{1}{\mu} \left( J_3 - ik_z J_1 \right) e^{ik_z} \hat{G}_A^\uparrow - \frac{1}{\mu} \left( J_3 + ik_z J_1 \right) e^{-ik_z} \hat{G}_A^\downarrow,$$

$$\mu J_7 \hat{G}_H = \left( \frac{k^2}{k^2} J_1 + J_5 \right) e^{ik_z} \hat{G}_A^\uparrow + \left( \frac{k^2}{k^2} J_1 + J_5 \right) e^{-ik_z} \hat{G}_A^\downarrow.$$
The particular choice of \( J_9 \) and \( J_7 \) in (5.56) allows us to just use \( J_1, \ldots, J_5 \) in the above expressions. When imposed on any interface \( z = d_\ell \), each of the equations in (5.56) is a linear equation of \( \hat{G}^*_A,\ell \) and \( \hat{G}^*_A,\ell+1 \) with coefficients in \( \mathcal{R}^0 \). Take the last one as an example. Due to (5.50) from the reaction field decomposition, in the equation \( J_\mu J_7 b^G_{H} = 0 \), the quantity on the \( z \to d_\ell \) side is

\[
\begin{align*}
(k_\rho^2 J_1 + J_5) e^{ik_\ell z} \hat{G}^A,\ell + (k_\rho^2 J_1 + J_5) e^{-ik_\ell z} \hat{G}^A,\ell \\
= (k_\rho^2 J_1 + J_5) \left( \hat{G}^r_A,\ell + \delta_\ell e^1(\delta e z') \frac{-e^{-ik_\ell z}}{2\omega k_\ell} (J_1 + J_2) \right) e^{ik_\ell z} \\
+ (k_\rho^2 J_1 + J_5) \left( \hat{G}^r_A,\ell + \delta_\ell e^1(\delta e < z') \frac{-e^{ik_\ell z'}}{2\omega k_\ell} (J_1 + J_2) \right) e^{-ik_\ell z},
\end{align*}
\]  

which is seen to be written using elements of \( \mathcal{R}^0 \) as coefficients. The same result applies to the \( z \to d_\ell \) side in layer \( \ell + 1 \). So the jump equation itself at \( z = d_\ell \) will also only involves elements of \( \mathcal{R}^0 \) as coefficients.

For the upward/downward outgoing radiation conditions [11], it is sufficient to describe them in the frequency domain as the decay conditions of \( \hat{G}_E \) and \( \hat{G}_H \) as \( z \to \pm\infty \), so that waves never come from \( z = \pm\infty \). Such conditions are sufficient to uniquely determine the Green’s function (c.f. Theorem 2.5), so we don’t bother to raise more complicated statements.

In the top layer, by (5.54), the downwards propagation components of \( \hat{G}_E \) and \( \hat{G}_H \) must be zero, since its asymptotic behavior is determined by the \( e^{-i k_0 z} \) factor, so

\[
-i\omega \left( J_1 + \frac{k_\rho^2}{k_0^2} J_2 + \frac{1}{k_0^2} J_5 - \frac{i k_0 z}{k_0^2} J_3 - \frac{i k_0 z}{k_0^2} J_4 \right) \hat{G}^A,0 = 0, \tag{5.59}
\]

\[
\frac{1}{\mu_0} (J_6 + J_7 + i k_0 z J_9) \hat{G}^A,0 = 0 \tag{5.60}
\]

where \( \hat{G}^A,0 = \hat{G}^r_A,0 \) when \( z > z' \) as it happens when \( z \to \infty \). When we treat (5.59) and
(5.60) as linear equations of $\hat{G}_{A,0}^{\uparrow}$, the coefficient matrix in (5.60)

$$\frac{1}{\mu_0}(J_6 + J_7 + ik_{0z}J_9) = \frac{1}{\mu_0}
\begin{bmatrix}
0 & ik_{0z} & ik_y \\
-ik_{0z} & 0 & -ik_x \\
-ik_y & ik_x & 0
\end{bmatrix}
$$

(5.61)

clearly has rank 2, allowing one degree of freedom on each column of $\hat{G}_{A,0}^{\uparrow}$. One can verify that for arbitrary $3 \times 1$ vector $v$,

$$\hat{G}_{A,0}^{\uparrow} = \hat{V}_0 \cdot v^T$$

(5.62)

is a solution to (5.60), thus solving the linear equation. Furthermore, we can also verify it is a solution to (5.59), so (5.60) can be neglected and keeping only (5.59) does not lose any mathematical condition. It is worth mentioning that the coefficient matrix of (5.59) is an element of $\mathfrak{H}^0$.

Similarly, as $z \to -\infty$ we get another equation in the bottom layer

$$-i\omega \left( J_1 + \frac{k_L^2}{k_L^2} J_2 + \frac{1}{k_L^2} J_5 + \frac{ik_L}{k_L^2} J_3 + \frac{ik_L}{k_L^2} J_4 \right) \hat{G}_{A,L} = 0,$$

(5.63)

where $\hat{G}_{A,L}^{\uparrow} = \hat{G}_{A,L}^{\downarrow}$ when $z < z'$.

Now, the interface conditions (5.56) and the radiation conditions (5.59) and (5.63) together consist a linear system of the unknown tensors $\hat{G}_{A,\ell}^{r_s}$ in the reaction field for $0 \leq \ell \leq L$ and $s \in \{\uparrow, \downarrow\}$, with coefficients in $\mathfrak{H}^0$. They are in general sufficient to restrict all the $\hat{G}_{A,\ell}^{r_s}$ terms. By Theorem 5.10, there exists a solution to this linear system with each unit of the block in $\mathfrak{H}^0$, i.e. each $\hat{G}_{A}^{\uparrow}, \hat{G}_{A}^{\downarrow} \in \mathfrak{H}^0$ piecewisely in each layer. Therefore, we conclude that the reaction field decomposition for the vector potential Green’s function in (5.49)

$$\hat{G}_A \in \mathfrak{H}^0$$

(5.64)
has the $\mathcal{R}^0$-matrix basis formulation.

5.3.2.3. Further simplification of the formulation

With the $\mathcal{R}^0$-matrix basis formulation we can simplify the interface equations (5.56) and the radiation equations (5.59) and (5.63). Suppose $\mathcal{G}_A^{rs} \in \mathcal{R}^0$ has the following basis expansion

$$\mathcal{G}_A^{rs} = \sum_{l=1}^{5} a_l^{\ast} J_l, \quad \ast \in \{\uparrow, \downarrow\}. \quad (5.65)$$

Following (5.49), we define

$$a_l = \delta_{jl} a_l^f + e^{ik_z z} a_l^\uparrow + e^{-ik_z z} a_l^\downarrow, \quad (5.66)$$

where $a_l^f$ are the $\mathcal{R}^0$-matrix basis coefficients of the free space potential tensor (5.48)

$$\sum_{l=1}^{5} a_l^f J_l = \mathcal{G}_A^f = \frac{-1}{2\omega k_{z^l}} e^{ik_{z^l}(z-z')}(J_1 + J_2). \quad (5.67)$$

Then, we have obtained the $\mathcal{R}^0$-matrix basis expression for

$$\mathcal{G}_A = \sum_{l=1}^{5} a_l J_l$$

using a reaction field decomposition (5.66) in each $a_l$. It is straightforward that each coefficient $a_l$ satisfies a Helmholtz equation

$$\partial_{zz} a_l + k_z^2 a_l = 0 \quad (5.68)$$

piecewisely in each layer, provided $z \neq z'$.

However, the potential tensor $\mathcal{G}_A$ is still not uniquely determined provided having a $\mathcal{R}^0$-matrix basis representation. For instance, for any functions $f_1, f_2 \in C^2(\mathbb{R}^3)$ satisfying the Helmholtz equation $\nabla^2 f_j + k^2 f_j = 0, \quad j = 1, 2$, the potential tensor $\mathcal{G}_A + \partial_j \hat{f}_1 J_2 + \hat{f}_2 J_1$
\( \hat{f}_1 \mathbf{J}_3 + \partial_z \hat{f}_2 \mathbf{J}_4 + \hat{f}_2 \mathbf{J}_5 \) can be used as an alternative choice for \( \mathbf{G}_A \). In order to eliminate these uncertainties in the coefficients, we will use less functions to represent \( \mathbf{G}_A \) and its reaction field components instead. Define functions \( b_1, b_2 \) and \( b_3 \) by linear transforms of \( a_l \):

\[
\begin{align*}
    b_1 &= a_1, \\
    b_2 &= \frac{1}{\mu} (a_2 - \partial_z a_3), \\
    b_3 &= \frac{1}{\mu} \left( \partial_z a_1 + k^2_{\rho} a_4 - k^2_{\rho} \partial_z a_5 \right).
\end{align*}
\]

Each \( b_l \) inherits the property of being the piecewise solution to the Helmholtz equation

\[
\partial_z b_l + k^2 b_l = 0, \quad z \neq z'.
\]

By substituting into representations of \( \mathbf{G}_E \) and \( \mathbf{G}_H \) (see (5.51)) using the matrix basis coefficients, i.e.

\[
\mathbf{G}_E = -i\omega \left( \mathbf{I} + \frac{\hat{\nabla} \hat{\nabla}^T}{k^2} \right) \sum_{l=1}^{5} a_l \mathbf{J}_l, \quad \mathbf{G}_H = \frac{1}{\mu} \hat{\nabla} \times \sum_{l=1}^{5} a_l \mathbf{J}_l,
\]

we get representations of \( \mathbf{G}_E \) and \( \mathbf{G}_H \) using \( b_1, b_2 \) and \( b_3 \)

\[
\begin{align*}
    \mathbf{G}_E &= -\frac{i\omega}{k^2} \left( k^2 b_1 \mathbf{J}_1 + \mu k^2_{\rho} b_2 \mathbf{J}_2 + \mu \partial_z b_2 \mathbf{J}_3 + \mu b_3 \mathbf{J}_4 + \left( \frac{k^2_{\rho}}{k^2_{\rho}} b_1 + \frac{\mu}{k^2_{\rho}} \partial_z b_3 \right) \mathbf{J}_5 \right), \\
    \mathbf{G}_H &= \frac{1}{\mu} \left( b_1 \mathbf{J}_6 + \mu b_2 \mathbf{J}_7 + \left( \frac{1}{k^2_{\rho}} \partial_z b_1 - \frac{\mu}{k^2_{\rho}} b_3 \right) \mathbf{J}_8 - \partial_z b_1 \mathbf{J}_9 \right),
\end{align*}
\]

i.e. using \( b_1, b_2 \) and \( b_3 \) is sufficient to represent the desired LMDGs of the electric field and of the magnetic field.

To derive an efficient method evaluating the \( b_l \) functions in layered media, \( l = 1, 2, 3 \), we begin with their reaction component decompositions. Corresponding to (5.66), we can
expand each $b_i$ as

$$b_i = b_i(k_\rho, z, z') = \delta_{\ell, \ell'} b_i^f(k_\rho, z, z') + e^{ik_z z'} b_i^{r*}(k_\rho, z', z), \quad (5.72)$$

where

$$b_1^f = a_1^f, \quad b_2^f = \frac{1}{\mu} \left( a_2^f - \partial_z a_3^f \right), \quad b_3^f = \frac{1}{\mu} \left( \partial_z a_1^f + k_\rho^2 a_4^f - k_\rho^2 \partial_z a_5^f \right),$$

$$b_1^{r*} = a_1^{r*}, \quad b_2^{r*} = \frac{1}{\mu} \left( a_2^{r*} - \tau^{r*} i k_z a_3^{r*} \right), \quad b_3^{r*} = \frac{1}{\mu} \left( \tau^{r*} i k_z a_1^{r*} + k_\rho^2 a_4^{r*} - k_\rho^2 \tau^{r*} i k_z a_5^{r*} \right), \quad (5.73)$$

$\ast \in \{\uparrow, \downarrow\}, \quad \tau^\uparrow = 1, \quad \tau^\downarrow = -1$. Specifically, since $G^f_A = -g^f / (i\omega)I$ was chosen as the free-space component, it’s clear that

$$a_1^f = a_2^f = -\frac{1}{i\omega} \hat{g}^f, \quad a_3^f = a_4^f = a_5^f = 0, \quad (5.74)$$

where $\hat{g}^f = ie^{ik_\rho z - z'}/(2k_\rho z)$. Therefore

$$b_1^f = -\frac{1}{i\omega} \hat{g}^f, \quad b_2^f = -\frac{1}{i\omega \mu \nu} \hat{g}^f, \quad b_3^f = -\frac{1}{i\omega} \frac{1}{\mu \nu} \partial_z \hat{g}^f = -\partial_z b_2^f. \quad (5.75)$$

In order to determine $b_1, b_2$ and $b_3$, we consider the interface equations and the radiation equations. One can easily verify the interface equations (5.42), when reinterpreted in the frequency domain as in (5.56), are equivalent to the following after comparing the matrix basis coefficients. For example, using

$$\mathbf{J}_1 \cdot \hat{\mathbf{G}}_E = -i\omega \left( b_1 \mathbf{J}_1 + \omega^{-2} \varepsilon^{-1} \partial_z b_2 \mathbf{J}_3 + (k_\rho^{-2} b_1 + \omega^{-2} \varepsilon^{-1} k_\rho^{-2} \partial_z b_3) \mathbf{J}_5 \right)$$

the continuity equations

$$[-i\omega b_1] = 0, \quad [-i\omega^{-1} \varepsilon^{-1} \partial_z b_2] = 0, \quad [-i\omega k_\rho^{-2} b_1 - i\omega^{-1} k_\rho^{-2} \varepsilon^{-1}] = 0$$

155
are revealed. After removing constant factors from these equations, a complete list by items is given below:

\[
\begin{align*}
[n \times \hat{G}_E] = 0 & \iff [J_1 \cdot \hat{G}_E] = 0 \iff [b_1] = 0, \left[\frac{1}{\varepsilon} \partial_z b_2\right] = 0, \left[\frac{1}{\varepsilon} \partial_z b_3\right] = 0; \\
[\varepsilon n \cdot \hat{G}_E] = 0 & \iff [J_2 \cdot \varepsilon \hat{G}_E] = 0 \iff [b_2] = 0, [b_3] = 0; \\
[n \times \hat{G}_H] = 0 & \iff [J_9 \cdot \hat{G}_H] = 0 \iff [b_2] = 0, [b_3] = 0, \left[\frac{1}{\mu} \partial_z b_1\right] = 0; \\
[\mu n \cdot \hat{G}_H] = 0 & \iff [J_7 \cdot \mu \hat{G}_H] = 0 \iff [b_1] = 0.
\end{align*}
\] (5.76)

The radiation equations (5.59) and (5.63) are reduced to

\[ b_{r,0}^r = 0, \quad b_{r,L}^r = 0 \] (5.77)

in the top and the bottom layer, respectively, i.e. waves coming from \( z = \pm \infty \) are prohibited in the reaction field decomposition. For each \( l = 1, 2, 3 \), the above contains a total of \( 2L + 2 \) linear equations of \( b_{r,0}^r \) and \( b_{r,L}^r \) from \( L + 1 \) layers. These linear equations are solvable, from the knowledge of the acoustic wave equation in layered media:

- \( -i \omega b_1 \) is exactly the reflection/transmission coefficient in the frequency domain of the LMDG of the Helmholtz equation, with piecewise constant material parameters \( 1/\varepsilon \).

Thus, we can solve \( b_1 \) in the frequency domain like solving the known scalar layered Helmholtz problem (c.f. Algorithm 3.1, [75]).

- Similarly, \( -i \omega \mu_{r,0} b_2 \) is exactly the one with piecewise constant parameters \( 1/\mu \).

- The linear system regarding \( b_2^{\ell, r} \) has exactly the same coefficients as \( b_2^{\ell, r} \) for the unknowns, so it’s solvable since \( b_2^{\ell, r} \) are uniquely determined by the physical problem. Moreover,

\[-\partial_z b_2 = -\partial_z \delta_{\ell, \ell} b_2^f - e^{ik_z z} \partial_z b_2^{f, r} - e^{-ik_z z} \partial_z b_2^{f, \ell}, \] (5.78)
which satisfies every equation that $b_3$ should satisfy, so by uniqueness,

$$b_3 = -\partial_\nu b_2, \text{ i.e. } b_3^* = -\partial_\nu b_2^*.$$  \hfill (5.79)

**Remark 5.12.** The $b_1$ and $b_2$ functions are corresponding to the TE mode component and the TM mode component in the $E_z-H_z$ formulation [48, 17], respectively.

**Remark 5.13.** To characterize $\hat{\mathbf{G}}_E$ and $\hat{\mathbf{G}}_H$ we don’t need the intermediate, undetermined tensor potential $\hat{\mathbf{G}}_A$ anymore. We derived the formulation for $\hat{\mathbf{G}}_A$ because some integral equation methods solve vector potential $\mathbf{A}$, e.g., in [56].

**Remark 5.14.** In some situations such as the half-space problem with the impedance boundary condition on the infinite boundary, the interface conditions are not exactly in the form of (5.42), but the result of the matrix basis formulation still holds with the same derivation.

**Remark 5.15 (modes of the system).** A mode of the layered media is an eigenstate without stimulation from any given source, i.e. the nontrivial solution of $\hat{\mathbf{G}}_A^r$ satisfying the above interface equations and radiation equations for certain values of $k_\rho$, with each $\hat{\mathbf{G}}_A^r$ replaced by 0. It corresponds to a pole in the frequency domain (see Remark 2.3). In such a situation we can still derive the simplified formulation using terms $b_1,b_2$ and $b_3$, but $b_3$ plays an independent role and is not anymore tied with $b_2$, i.e. a 2-term formulation won’t be sufficient.

5.3.2.4. **The transverse potential and the Sommerfeld potential**

Here we take a quick review on the transverse potential and the Sommerfeld potential formulations and show how to interpret them from the matrix basis formulation. Both formulations restrict certain 5 entries of the $3 \times 3$ tensor $\hat{\mathbf{G}}_A$ to be nonzero, which uniquely determine the tensor potential. Here, we claim that the potential tensors in these formulations have the $\mathfrak{R}^0$-matrix basis representation, and can be derived using $b_1$ and $b_2$. Due to the uniqueness of $b_1$ and $b_2$, it suffices to explicitly construct them.
The transverse potential takes the form

\[
\mathbf{\hat{G}}^t_A = \begin{bmatrix}
\times & \times & \\
\times & \times & \\
\times & & \\
\end{bmatrix},
\]

(5.80)

where each \(\times\) marks a nonzero entry. We claim \(\mathbf{\hat{G}}^t_A = a_1 \mathbf{J}_1 + a_2 \mathbf{J}_2 + a_5 \mathbf{J}_5\). By (5.69),

\[
b_1 = a_1, \quad b_2 = \frac{1}{\mu} a_2, \quad b_3 = -\partial_z b_2 = \frac{1}{\mu} \left( \partial_z a_1 - k_\rho^2 \partial_z a_5 \right).
\]

(5.81)

Since \(a_i, b_i\) satisfy the Helmholtz equation (5.68) and (5.70), respectively, we have

\[
a_1 = b_1, \quad a_2 = \mu b_2, \quad a_5 = b_1 - \frac{\mu \partial_z \partial_z b_2}{k_\rho^2 k_z^2}.
\]

(5.82)

The Sommerfeld potential takes the form

\[
\mathbf{\hat{G}}^s_A = \begin{bmatrix}
\times & \\
\times & \\
\times & \times \\
\end{bmatrix}
\]

(5.83)

and we claim \(\mathbf{\hat{G}}^s_A = a_1 \mathbf{J}_1 + a_2 \mathbf{J}_2 + a_4 \mathbf{J}_4\). Again by (5.69),

\[
b_1 = a_1, \quad b_2 = \frac{1}{\mu} a_2, \quad b_3 = -\partial_z b_2 = \frac{1}{\mu} \left( \partial_z a_1 + k_\rho^2 a_4 \right),
\]

(5.84)

so

\[
a_1 = b_1, \quad a_2 = \mu b_2, \quad a_4 = -\frac{\mu \partial_z b_2 + \partial_z b_1}{k_\rho^2}.
\]

(5.85)
Remark 5.16. In the transverse potential \( \hat{G}_A \), although the coefficient \( a_5 \) has a \( k_\rho^2 \) factor in the denominator, there’s no singularity in the integrand of \( \hat{G}_A \) at \( k_\rho = 0 \) since they can be cancelled out with the entries of \( J_5 \) (by using the \((k_\rho, \alpha)\) polar coordinates). The same happens to the Sommerfeld potential \( \hat{G}_S \), but it’s not explicitly shown in the expression of \( a_4 J_4 \). Numerically, we should take some care if the values as \( k_\rho \to 0 \) are required.

5.4. Numerical validation of the Maxwell’s LMDG in a 10-layer medium

The matrix basis formulation (5.71) using the coefficients \( b_1(k_\rho, z, z') \), \( b_2(k_\rho, z, z') \) and \( b_3(k_\rho, z, z') \) in the frequency domain can be used to accurately calculate the Maxwell’s LMDG.

The free-space Green’s function can be computed using its closed analytical form. Meanwhile, for the reaction field, we can use Algorithm 3.1 to compute each reaction field component of \( b_1 \) and \( b_2 \). We use (5.79) to represent \( b_3 \), while the \( \partial_z \) and \( \partial_{z'} \) operators are converted to \( \pm ik_\ell_z \) and \( \pm ik_{\ell z'} \) factors, respectively, for each reaction field component related to its field propagation direction.

To compute the inverse Fourier transforms of (3.6) for \( \hat{G}_E \) and \( \hat{G}_H \), we will convert them into single integrals for each entry of the tensors. This can be done by first representing entries of the matrix basis in terms of polar coordinates \((k_\rho, \alpha)\), e.g.

\[
-k_x^2 = k_\rho^2 \left( -\frac{1}{2} - \frac{1}{4} e^{2i\alpha} - \frac{1}{4} e^{-2i\alpha} \right)
\]

(a complete list will be shown in Section 5.6), then reduce to single integral (3.87),

\[
I_m[f] = \int_{\mathbb{R}^2} e^{im\alpha} e^{ik_x(x-x')+ik_y(y-y')} f(k_\rho) dk_x dk_y
\]

\[
= 2\pi^m e^{im\hat{\phi}} \int_0^\infty k_\rho J_m(k_\rho \rho) f(k_\rho) dk_\rho,
\]

provided that the double integral is absolutely integrable. Here, \((\rho, \hat{\phi})\) is the polar coordinate.
pair of \((x - x', y - y')\), and the integral definition of \(m\)-th order Bessel function

\[
\int_0^{2\pi} e^{ik_\rho \cos(\alpha - \hat{\phi})} e^{im(\alpha - \hat{\phi})} d\alpha = 2\pi i^m J_m(k_\rho \rho)
\]

(5.87)

has been used in the derivation of the identity. Each integral of the form \(I_m[f]\) is computed using the Gauss–Legendre quadrature on a truncated contour in the fourth quadrant in order to get rid of singularities at wave number values.

We consider a numerical validation test for the LMDG formula by computing the electric field and the magnetic field Green’s functions of a 10-layer problem. The geometry of the layered medium is defined by horizontal interface planes \(z = d_\ell, 0 \leq \ell \leq 8\), in the descending order, as

\[
\{d_\ell\}_{\ell=0}^{8} = \{0.0, -1.0, -3.0, -7.0, -8.0, -10.0, -11.0, -13.0, -14.0\}, \quad (5.88)
\]

separating the space into layers with index \(0, 1, \cdots, 9\) from top to bottom. Suppose layers 0–9 have constant relative permittivity

\[
\{\varepsilon_\ell\}_{\ell=0}^{9} = \{1.27, 3.25, 3.41, 5.7, 1.52, 3.691, 1.2, 3.5, 2.1, 3.3\}, \quad (5.89)
\]

respectively, and constant relative permeability

\[
\{\mu_\ell\}_{\ell=0}^{9} = \{1.05, 0.95, 1.05, 3.95, 10.05, 6.22, 9.97, 3.2, 10.0, 1.0\}, \quad (5.90)
\]

respectively. The time frequency \(\omega = 1.0\). A dipole source is placed in layer no. 3 (the fourth layer) at

\[
r' = (0.0, 0.0, -4.23), \quad (5.91)
\]
<table>
<thead>
<tr>
<th>$l$</th>
<th>$e_{E,x}^l$</th>
<th>$e_{E,z}^l$</th>
<th>$e_{H,x}^l$</th>
<th>$e_{H,z}^l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.44e-15</td>
<td>1.80e-15</td>
<td>8.08e-16</td>
<td>1.57e-15</td>
</tr>
<tr>
<td>1</td>
<td>1.86e-15</td>
<td>1.14e-14</td>
<td>5.62e-15</td>
<td>9.72e-15</td>
</tr>
<tr>
<td>2</td>
<td>4.42e-13</td>
<td>3.06e-12</td>
<td>8.52e-13</td>
<td>2.14e-12</td>
</tr>
<tr>
<td>3</td>
<td>1.15e-13</td>
<td>6.33e-13</td>
<td>2.37e-13</td>
<td>5.15e-13</td>
</tr>
<tr>
<td>4</td>
<td>1.06e-14</td>
<td>2.51e-14</td>
<td>6.55e-15</td>
<td>7.01e-14</td>
</tr>
<tr>
<td>5</td>
<td>1.40e-15</td>
<td>1.05e-14</td>
<td>1.25e-15</td>
<td>1.70e-14</td>
</tr>
<tr>
<td>6</td>
<td>4.45e-16</td>
<td>2.87e-15</td>
<td>5.01e-16</td>
<td>2.13e-15</td>
</tr>
<tr>
<td>7</td>
<td>9.29e-17</td>
<td>1.34e-16</td>
<td>5.85e-17</td>
<td>4.74e-16</td>
</tr>
<tr>
<td>8</td>
<td>2.76e-17</td>
<td>1.22e-16</td>
<td>5.21e-17</td>
<td>1.86e-16</td>
</tr>
</tbody>
</table>

Table 5.1: Maximum absolute error of the continuity of $E_x$, $E_y$, $\varepsilon E_z$, $H_x$, $H_y$ and $\mu H_z$ across interface planes $z = d_\ell$, $0 \leq \ell \leq 8$ for 10201 $(x,y)$ coordinate pairs in the range of $[-5.0, 5.0] \times [-5.0, 5.0]$. The dipole source locates in layer no. 3.

orientated along the direction

$$\hat{\alpha}' = (1/2, 1/2, 1/\sqrt{2}).$$

(5.92)

Figure 5.2 shows the computed electric field in real and the imaginary parts of its components on the plane $x = 0.2$ for $-5.0 \leq y \leq 5.0$ and $-14.5 \leq z \leq 0.5$. Figure 5.3 shows the computed magnetic field in the same domain. The computed electric field and magnetic field values are clamped to the range $(-0.05, 0.05)$ for clearer illustration since the reaction field is overall smaller compared to the free-space part.

The accuracy is checked by the continuity conditions at interface planes, i.e.,

$$[E_x] = 0, \quad [E_y] = 0, \quad [\varepsilon E_z] = 0, \quad [H_x] = 0, \quad [H_y] = 0, \quad [\mu H_z] = 0$$

(5.93)

from (5.42). From each side of the interface plane $z = d_\ell$, we compute the values of $E_x$, $E_y$, $\varepsilon E_z$, $H_x$, $H_y$ and $\mu H_z$ at certain $(x,y)$ pairs within the square $[-5.0, 5.0] \times [-5.0, 5.0]$ which form a 101 $\times$ 101 uniform grid, then pick the maximum difference of the value jumps for each
Figure 5.2: Electric fields in a 10-layer medium described in Section 5.4. Fields are computed along $x = 0.2$ for $-5.0 \leq y \leq 5.0$ and $-14.5 \leq z \leq 0.5$. The contouring level is clamped to the range $(-0.05, 0.05)$ for a clearer illustration of wave pattern and avoiding the peak values near the source, which is located in the fourth layer at $r' = (0.0, 0.0, -4.23)$ with an orientation along the direction $\hat{\alpha}' = (1/2, 1/2, 1/\sqrt{2})$. 
Figure 5.3: Magnetic fields in the 10-layer medium as described in Figure 5.2.
term, respectively. Namely, define

\begin{align}
\epsilon_{E,x}^l &= \max_{0 \leq p,q \leq 100} |E_x(x_p, y_q, d_l + 0) - E_x(x_p, y_q, d_l - 0)|, \\
\epsilon_{E,y}^l &= \max_{0 \leq p,q \leq 100} |E_y(x_p, y_q, d_l + 0) - E_y(x_p, y_q, d_l - 0)|, \\
\epsilon_{E,z}^l &= \max_{0 \leq p,q \leq 100} |\varepsilon_l E_z(x_p, y_q, d_l + 0) - \varepsilon_{l+1} E_z(x_p, y_q, d_l - 0)|, \\
\epsilon_{H,x}^l &= \max_{0 \leq p,q \leq 100} |H_x(x_p, y_q, d_l + 0) - H_x(x_p, y_q, d_l - 0)|, \\
\epsilon_{H,y}^l &= \max_{0 \leq p,q \leq 100} |H_y(x_p, y_q, d_l + 0) - H_y(x_p, y_q, d_l - 0)|, \\
\epsilon_{H,z}^l &= \max_{0 \leq p,q \leq 100} |\mu_l H_z(x_p, y_q, d_l + 0) - \mu_{l+1} H_z(x_p, y_q, d_l - 0)|,
\end{align}

where \( l = 0, 1, \cdots, 9 \), \( x_p = -5.0 + 0.1p \) and \( y_q = -5.0 + 0.1q \), \( p, q = 0, 1, \cdots, 100 \). Indeed, \( \epsilon_{E,x}^l = \epsilon_{E,y}^l \) and \( \epsilon_{H,x}^l = \epsilon_{H,y}^l \) in this test problem because of symmetry. Table 5.1 shows the absolute errors \( \epsilon_{E,x}^l, \epsilon_{E,z}^l, \epsilon_{H,x}^l \) and \( \epsilon_{H,z}^l \) defined above, which are bounded by 3.1e-12. When only counting the interfaces from non-source layers, the absolute errors are bounded by 7.1e-14. Figure 5.2 and Figure 5.3 also indicate the continuity of computed \( E_x, E_y, H_x \) and \( H_y \) across interface planes, as well as the expected discontinuity of \( E_z \) and \( H_z \).

5.5. Application to the elastic wave equation in layered media

In this section we will apply the \( \mathfrak{R}^l \)-matrix based formulation to the LMDG of the elastic wave equation. The interface conditions for various phases of contacting media will be discussed. In the end we also give a brief discussion on the case when source is inside a zero-viscosity fluid layer, with a simplified vector basis formulation. In these problems, we suppose the displacement of the media is small, so that linearization of the elastic wave equations is in general applicable.
5.5.1. The dyadic Green’s function of the elastic wave equation in the free space

Suppose the homogeneous and isotropic material occupies the entire space, with density \( \rho \) and Lamé constants \( \lambda, \mu \). Define

\[
\gamma = \lambda + 2\mu. \tag{5.100}
\]

In most solid materials, both \( \lambda, \mu > 0 \). If the material is (compressible) liquid or gas, we assume the viscosity is negligible (also for the rest of this chapter; otherwise we should treat it in the same way as a solid layer), then \( \mu = 0 \), and \( \gamma = \lambda \). Elastic waves do not propagate in incompressible medium, and such a layer divides the layered media into one or two equivalent half-space structures. Note that the shear modulus \( \mu \) is different from the dynamic viscosity in fluid. Also, we assume that the time dependence is harmonic, i.e. in the form of \( \exp(i\omega t) \).

With an external force \( \mathbf{b}(r) \), the elastic wave equation in solid is a partial differential equation for the displacement vector \( \mathbf{u}(r) \)

\[
-\omega^2 \rho \mathbf{u} = \nabla \cdot \mathbf{T} + \mathbf{b}, \tag{5.101}
\]

assuming \( |\mathbf{u}| \ll 1 \), where the \( 3 \times 3 \) stress tensor \( \mathbf{T} \) is defined as

\[
\mathbf{T}_{ij} = \lambda \delta_{ij} \sum_{l=1}^{3} \frac{\partial u_l}{\partial x_l} + \mu \left( \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \tag{5.102}
\]

where \((x_1, x_2, x_3)\) is used as an alternative notation for \((x, y, z)\), and \( \mathbf{u} = (u_1, u_2, u_3) \) [15]. The direct equivalent form in terms of displacement is

\[
(\lambda + \mu)\nabla \nabla \cdot \mathbf{u} + \mu \nabla^2 \mathbf{u} + \omega^2 \rho \mathbf{u} = -\mathbf{b}. \tag{5.103}
\]

Given a source location \( r' \), the dyadic Green’s function \( \mathbf{G}(r; r') \) is a \( 3 \times 3 \) tensor satisfying the equation

\[
(\lambda + \mu)\nabla \nabla \cdot \mathbf{G} + \mu \nabla^2 \mathbf{G} + \omega^2 \rho \mathbf{G} = -\delta(r - r') \mathbf{I}. \tag{5.104}
\]
The solution is known as
\[
G = G_f(r; r') = \frac{1}{\mu} \left( I + \frac{\nabla\nabla}{k_s^2} \right) g_s^f(r; r') - \frac{1}{\gamma} \frac{\nabla\nabla}{k_c^2} g_c^f(r; r'), \tag{5.105}
\]
where
\[
k_s = \sqrt{\frac{\omega^2 \rho}{\mu}}, \quad k_c = \sqrt{\frac{\omega^2 \rho}{\gamma}} \tag{5.106}
\]
are the wave numbers of the S-wave and the P-wave, respectively, and
\[
g_s^f(r; r') = \frac{e^{ik_s|r-r'|}}{4\pi|r-r'|}, \quad g_c^f(r; r') = \frac{e^{ik_c|r-r'|}}{4\pi|r-r'|} \tag{5.107}
\]
are the free space Green’s functions of the Helmholtz equation (3.4) with wave numbers \(k_s\) and \(k_c\), respectively [15].

In fluid, i.e. in liquid or gas, assuming the external force \(b\) is conservative (for the rest of this chapter as well), there’s only the P-wave that propagates in the media, and the acoustic wave equation with respect to the displacement \(u\) is
\[
\begin{align*}
\lambda \nabla\nabla \cdot u + \omega^2 \rho u &= -b, \tag{5.108} \\
\nabla \times u &= 0 \tag{5.109}
\end{align*}
\]
where the first equation repeats (5.103), and the second one is introduced because of zero viscosity. We also assume the displacement \(|u| \ll 1\) and the density \(\rho\) is approximately constant.

Next, we consider the Euler’s equation
\[
-\omega^2 \rho u \approx \rho \left( \frac{Dv}{Dt} \right)^\wedge = -\nabla p + b \tag{5.110}
\]
where \(v\) is the velocity, \(Dv/Dt \approx \partial v/\partial t\) is the material derivative since \(|u| \ll 1\), and \(^\wedge\) represents the Fourier transform of time. For the pressure \(p\), from (5.108) and (5.110) we
can solve
\[ p = -\lambda \nabla \cdot \mathbf{u} + p_0 \]  
(5.111)

where \( p_0 \) is a constant. For the sake of convenience, we set \( p_0 = 0 \) without loss of generality.

Take the divergence of the first equation of (5.108),
\[ \nabla^2 p + \frac{\omega^2 \rho}{\lambda} p = \frac{1}{\lambda} \nabla \cdot \mathbf{b}. \]  
(5.112)

The dyadic Green’s function of the P-wave propagation is often proposed in terms of the pressure \( p \), namely
\[ \nabla^2 g_p + \frac{\omega^2 \rho}{\lambda} g_p = -\delta(\mathbf{r} - \mathbf{r}') \]  
(5.113)

with solution \( g_p = g_p^f(\mathbf{r}; \mathbf{r}') = g_c^f(\mathbf{r}; \mathbf{r}') \). In order to later derive the Green’s function in adjacent solid layers, we will also use the corresponding Green’s function \( g_u^f \) with respect to the displacement \( \mathbf{u} \). \( g_u^f \) should satisfy the same equation as (5.111)
\[ g_p = -\lambda \nabla \cdot \mathbf{g_u} + p_0, \]

and the Euler’s equation
\[ \lambda \nabla \nabla \cdot \mathbf{g_u} + \omega^2 \rho \mathbf{g_u} = 0, \quad \mathbf{r} \neq \mathbf{r}', \]

so
\[ \mathbf{g_u}^f = \frac{1}{\omega^2 \rho} \nabla g_p^f. \]  
(5.114)

5.5.2. The elastic wave equation in layered media

In layered media, suppose the medium in each layer is homogeneous with density \( \rho_\ell \) and Lamé constants \( \lambda_\ell \) and \( \mu_\ell \), \( \ell = 0, \cdots, L \), respectively. Define \( \gamma = \lambda + 2\mu \). The wave numbers
of the S-wave and the P-wave are given by

\[ k_s = \sqrt{\frac{\omega^2 \rho}{\mu}}, \quad k_c = \sqrt{\frac{\omega^2 \rho}{\gamma}} \]  

(5.115)

as well as in the free space. In fluid media, \( k_s \) is not defined because \( \mu = 0 \) and S-wave does not propagate.

5.5.2.1. The interface conditions of the elastic wave equation

Before discussing the Green’s functions, we take a detailed study on the interface conditions. Recall the equation (5.101) in terms of the stress tensor \( \mathbf{T} \), if the external force \( \mathbf{b} \) and the displacement \( \mathbf{u} \) are not singular near the interface, then by divergence theorem on a flat cylinder crossing the boundary with infinitesimal thickness, we get the interface condition

\[ \mathbf{n} \cdot \mathbf{T} \mathbf{K} = \mathbf{0}. \]  

(5.116)

With \( \mathbf{n} = \mathbf{e}_3 \) in our problem setting, it suffices to consider the entries

\[ \mathcal{T}_{31} = \mu \left( \frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \right), \]

\[ \mathcal{T}_{32} = \mu \left( \frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3} \right), \]

\[ \mathcal{T}_{33} = \gamma \frac{\partial u_3}{\partial x_3} + \lambda \left( \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} \right) \]  

(5.117)

for the continuity equations. The identity (5.116) also implies some additional regular conditions as listed below [15].

- Across the solid-solid interface, \( \partial u_3 / \partial x_3 \) must be regular, so \( \llbracket u_3 \rrbracket = 0 \). Then from the continuity of \( \mathcal{T}_{31} \) we further get \( \llbracket u_1 \rrbracket = 0 \), and similarly \( \llbracket u_2 \rrbracket = 0 \).
- Across the solid-fluid interface, we have \( \llbracket u_3 \rrbracket = 0 \) for the same reason. Since in the fluid side \( \mu = 0 \) so that \( \mathcal{T}_{31} = \mathcal{T}_{32} = 0 \), no additional condition is required.
• Across the fluid-fluid interface, the conditions on $\mathcal{T}_{31}$ and $\mathcal{T}_{32}$ are no more necessary, therefore only $[\mathcal{T}_{33}] = 0$ and $[u_3] = 0$ are required. Note that in this case $\mathcal{T}_{33} = \lambda \nabla \cdot \mathbf{u} = -p$, and $u_3 = (\omega^2 \rho)^{-1} \partial p / \partial x_3$.

• In the half-space problem where the media has an interface against the vacuum where no acoustic wave propagates and $\lambda = \mu = 0$, the zero-traction conditions $[\mathcal{T}_{3l}] = 0$ are required, $l = 1, 2, 3$, while the displacement on the boundary is set free.

In summary, the interface equations are listed below for each case:

• Across a solid-solid interface,

$$[\mathcal{T}_{3l}] = 0, \quad [u_l] = 0, \quad l = 1, 2, 3. \quad (5.118)$$

• Across a solid-fluid interface,

$$[u_3] = 0, \quad [\mathcal{T}_{3l}] = 0, \quad l = 1, 2, 3, \quad (5.119)$$

where from the fluid side $\mathcal{T}_{31} = \mathcal{T}_{32} = 0$ automatically holds.

• Across a fluid-fluid interface,

$$[u_3] = 0, \quad [\mathcal{T}_{33}] = 0, \quad (5.120)$$

or equivalently, in terms of the pressure $p$,

$$\left[ \begin{array}{c} \frac{1}{\rho} \partial p / \partial z \\ \partial p / \partial z \end{array} \right] = 0, \quad [p] = 0. \quad (5.121)$$

• On a solid-vacuum interface,

$$[\mathcal{T}_{3l}] = 0, \quad l = 1, 2, 3. \quad (5.122)$$
• On a fluid-vacuum interface,

\[ [T_{33}] = 0. \]  \hspace{1cm} (5.123)

Note that we have assumed zero displacement in the vacuum so as to keep the same formulation for later discussion. To flexibly handle various interface circumstances, we separate these equations into four groups:

(a) \([T_{33}] = 0; \) \hspace{1cm} (b) \([u_3] = 0; \) \hspace{1cm} (c) \([T_{31}] = [T_{32}] = 0; \) \hspace{1cm} (d) \([u_1] = [u_2] = 0. \)

For each of the above cases we simply join the groups as needed.

5.5.3. The dyadic Green’s function in layered media with source in solid

We begin with the case when all the layers are solid, and the case with fluid or vacuum layers present will follow. Suppose a source \( r' = (x', y', z') \) locates in layer \( \ell' \), and the target \( r = (x, y, z) \) is in layer \( \ell \). Take the 2-D Fourier transform \((x - x', y - y')\) to \((k_x, k_y)\) as defined in (3.6). Suppose \((k_\rho, \alpha)\) are the polar coordinates of \((k_x, k_y)\). Similar to (5.44) for the Maxwell’s equations, we define the reaction field Green’s function

\[ G^r = G - \delta_{\ell', \ell} G^f, \]  \hspace{1cm} (5.124)

then \( G^r \) satisfies a homogeneous elastic wave equation

\[ (\lambda + \mu) \nabla \nabla \cdot G^r + \mu \nabla^2 G^r + \omega^2 \rho G^r = 0 \]  \hspace{1cm} (5.125)

within each layer, which, in the frequency domain, becomes a second-order ordinary differential equation of \( z \)

\[ (\mu J_1 + \gamma J_2) \partial_{zz} \hat{G}^r + (\lambda + \mu)(J_3 + J_4) \partial_z \hat{G}^r + (\mu k_{sz}^2 (J_1 + J_2) + (\lambda + \mu) J_5) \hat{G}^r = 0, \]  \hspace{1cm} (5.126)
\[ k_{sz} = \sqrt{k_s^2 - k_p^2}, \quad k_{cz} = \sqrt{k_c^2 - k_p^2}. \] (5.127)

The differential equation (5.126) can be solved by stacking \( \hat{G}^r \) and \( \partial_z \hat{G}^r \) as unknowns and treating as a first-order ordinary differential equation, whose coefficient matrix has single eigenvalues \( \pm ik_{cz} \) and repeated eigenvalues \( \pm ik_{sz} \). By further calculating the eigenvectors, a general solution is given by

\[ \hat{G}^r = \sum_{\tau \in \{\uparrow, \downarrow\}} \left( (\tau^*ik_{sz}J_1 + J_4)e^{\tau^*ik_{sz}z} + (\tau^*ik_{cz}J_2 + J_3)e^{\tau^*ik_{cz}z} \right) X^r, \] (5.128)

where \( \tau^\uparrow = +1, \tau^\downarrow = -1 \), \( X^r = X^r^* \) is a piecewisely constant (with respect to the variable \( z \)) \( 3 \times 3 \) tensor in each layer.

The free-space part of the dyadic Green’s function \( \hat{G}^f \) indeed takes the same form. By expanding (5.105) in the frequency domain,

\[ \hat{G}^f = \frac{1}{\omega^2 \rho} \left( k_{sz,\ell}' J_1 + k_{p,\ell}' J_2 + ik_{sz,\ell}' \tau J_3 + ik_{sz,\ell}' \tau J_4 + J_5 \right) \hat{g}_{s}^f \]
\[ - \frac{1}{\omega^2 \rho} \left( -k_{cz,\ell}' J_2 + ik_{cz,\ell}' \tau J_3 + ik_{cz,\ell}' \tau J_4 + J_5 \right) \hat{g}_{c}^f \] (5.129)

for \( z \neq z' \), where \( \tau \) indicates the sign of \( z - z' \),

\[ \hat{g}_{s}^f = \frac{ie^{ik_{sz,\ell}'|z-z'|}}{2k_{sz,\ell}'}, \quad \hat{g}_{c}^f = \frac{ie^{ik_{cz,\ell}'|z-z'|}}{2k_{cz,\ell}'}, \] (5.130)

we can verify that the same pattern is matched as in (5.128):

\[ \hat{G}^f = \sum_{\tau \in \{\uparrow, \downarrow\}} \left( (\tau^*ik_{sz,\ell}'J_1 + J_4)e^{\tau^*ik_{sz,\ell}'z} + (\tau^*ik_{cz,\ell}'J_2 + J_3)e^{\tau^*ik_{cz,\ell}'z} \right) X^f, \] (5.131)
where

\[ X^{f\uparrow} = 1_{\{z > z'\}} \left( k_{s,\ell'}^2 D_s J_1 - k_{cz,\ell'}^2 D_c J_2 - i k_{sz,\ell'} D_s J_3 - i k_{cz,\ell'} D_c J_4 + D_s J_5 \right), \]

\[ X^{f\downarrow} = 1_{\{z < z'\}} \left( k_{s,\ell'}^2 D_s J_1 - k_{cz,\ell'}^2 D_c J_2 + i k_{sz,\ell'} D_s J_3 + i k_{cz,\ell'} D_c J_4 + D_s J_5 \right), \]

provided \( z \neq z' \), here

\[ D_s = \frac{-1}{2 \omega_0^2 \rho \ell' k_{sz,\ell'}^2}, \quad D_c = \frac{-1}{2 \omega_0^2 \rho \ell' k_{cz,\ell'}^2}. \] (5.133)

In total, we get the reaction field decomposition of \( \hat{G} \) by

\[ \hat{G} = \sum_{* \in \{\uparrow, \downarrow\}} \left( (-\tau^* i k_{sz} J_1 + J_4)e^{\tau^* i k_{sz} z} + (\tau^* i k_{cz} J_2 + J_3)e^{\tau^* i k_{cz} z} \right) X^*, \] (5.134)

where

\[ X^* = \delta_{j,t} X^{f*} + X^{r*}. \] (5.135)

The unknowns \( X^{r*} \) from all \( L + 1 \) layers will be determined by the interface equations and the radiation equations specified below.

The interface conditions of the Green’s function are given by \((5.118)–(5.122)\) by cases. One can re-organize them into the representation using matrices \( J_1, \cdots, J_5 \) in the following way:

(a) from \( \mathbf{T}_{33} = 0 \) we obtain the interface equations

\[ \left[ \begin{array}{c} \gamma \frac{\partial G_{3l}}{\partial z} + \lambda \left( \frac{\partial G_{1l}}{\partial x} + \frac{\partial G_{2l}}{\partial y} \right) \end{array} \right] = 0, \quad l = 1, 2, 3 \] (5.136)

\[ \Leftrightarrow \left[ \begin{array}{c} (\lambda J_4 + \gamma \partial_z J_2)\hat{G} \end{array} \right] = 0 \] (5.137)

\[ \Leftrightarrow \left[ \begin{array}{c} \sum_{* \in \{\uparrow, \downarrow\}} (2\mu \tau^* i k_{sz} e^{\tau^* i k_{sz} z} J_4 + (-\lambda k_{\rho}^2 - \gamma k_{cz}^2) e^{\tau^* i k_{cz} z} J_2) X^* \end{array} \right] = 0; \] (5.138)
(b) from $[u_3] = 0$ we have the interface equations

$$\begin{bmatrix} G_{3l} \end{bmatrix} = 0, \quad l = 1, 2, 3$$  \hspace{1cm} (5.139)

$$\Leftrightarrow \begin{bmatrix} J_2 \hat{G} \end{bmatrix} = 0$$  \hspace{1cm} (5.140)

$$\Leftrightarrow \sum_{* \in \{↑, ↓\}} \left( e^{\tau^* i k_{sz} z} J_4 + \tau^* i k_{cz} e^{\tau^* i k_{cz} z} J_2 \right) X^* = 0;$$  \hspace{1cm} (5.141)

(c) from $[T_{31}] = [T_{32}] = 0$ we have the interface equations

$$\mu \left( \frac{\partial G_{3l}}{\partial x} + \frac{\partial G_{1l}}{\partial z} \right) = \mu \left( \frac{\partial G_{3l}}{\partial y} + \frac{\partial G_{2l}}{\partial z} \right) = 0, \quad l = 1, 2, 3$$  \hspace{1cm} (5.142)

$$\Leftrightarrow \mu (\partial_z J_1 + J_3) \hat{G} = 0$$  \hspace{1cm} (5.143)

$$\Leftrightarrow \sum_{* \in \{↑, ↓\}} \left( \mu k_{sz}^2 e^{\tau^* i k_{sz} z} J_1 + \mu e^{\tau^* i k_{sz} z} J_5 + 2 \mu \tau^* i k_{cz} e^{\tau^* i k_{cz} z} J_3 \right) X^* = 0;$$  \hspace{1cm} (5.144)

(d) from $[u_1] = [u_2] = 0$ we have the interface equations

$$\begin{bmatrix} G_{1l} \end{bmatrix} = [G_{2l}] = 0, \quad l = 1, 2, 3$$  \hspace{1cm} (5.145)

$$\Leftrightarrow \begin{bmatrix} J_1 \hat{G} \end{bmatrix} = 0$$  \hspace{1cm} (5.146)

$$\Leftrightarrow \sum_{* \in \{↑, ↓\}} \left( -\tau^* i k_{sz} e^{\tau^* i k_{sz} z} J_1 + e^{\tau^* i k_{cz} z} J_3 \right) X^* = 0.$$  \hspace{1cm} (5.147)

The radiation conditions are similarly treated like in the Maxwell’s equations. The limits as $z \to \pm \infty$ are reduced to

$$\left( (ik_{sz,0} J_1 + J_4) e^{-ik_{sz,0} z} + (-ik_{cz,0} J_2 + J_3) e^{-ik_{cz,0} z} \right) X^\downarrow_0 \to 0$$ \hspace{1cm} (5.148)

as $z \to \infty$ in the top layer, which can be simplified as $X^\uparrow_0 = 0$, and

$$\left( (-ik_{sz, L} J_1 + J_4) e^{ik_{sz, L} z} + (ik_{cz, L} J_2 + J_3) e^{ik_{cz, L} z} \right) X^\downarrow_0 \to 0$$ \hspace{1cm} (5.149)

173
as \( z \to -\infty \) in the bottom layer, which can be simplified as \( X^r_L = 0 \), i.e. propagating directions from \( z = \pm \infty \) are prohibited.

Since the above equations are linear equations of \( X^r \) for all layers with coefficients in \( \mathbb{F}_0 \), and are sufficient to uniquely determine the Green’s function, i.e. the unknowns \( X^r \) within each layer, by Theorem 5.10, each

\[
X^r_t \in \mathcal{R}^0
\]

has an \( \mathcal{R}^0 \)-matrix basis formulation.

Again, we will try to reinterpret the interface equations and the radiation equations so that they’re written in terms of the \( \mathcal{R}^0 \)-matrix basis coefficients. Suppose in each layer \( X^r_t \) has the matrix basis expansion

\[
X^r_t = \sum_{l=1}^{5} x^r_{l,t} J_l,
\]

then, from (5.135), we have

\[
X = \sum_{l=1}^{5} x_{l} J_l,
\]

where

\[
x^\uparrow_l = \delta_{\ell,\ell'} 1_{\{z>z'\}} x^f_{l} + x^r_{l},
\]

\[
x^\downarrow_l = \delta_{\ell,\ell'} 1_{\{z<z'\}} x^f_{l} + x^r_{l}
\]

are the reaction field decomposition of \( x^*_l \). Using (5.134) and the \( \mathcal{R}^0 \)-matrix basis representation of \( X^* \), the radiation equations are simply

\[
x^\uparrow_{l,0} = x^\uparrow_{l,L} = 0, \quad l = 1, \cdots, 5.
\]

For the interface equations, the groups (a)–(d) are listed below.
(a) \( [T_{33}] = 0 \) is equivalent to (5.136) and now interpreted as

\[
[T_1 \mathbf{J}_2 + T_2 \mathbf{J}_4] = 0 \iff [T_1] = [T_2] = 0,
\]

where

\[
T_1 = \sum_{* \in \{\uparrow, \downarrow\}} -2\mu \tau^*ik_{sz}k^2_\rho e^{\tau^*ik_{cz}z}x^*_3 + (-\lambda k^2_\rho - \gamma k^2_{cz})e^{\tau^*ik_{cz}z}x^*_2,
\]

\[
T_2 = \sum_{* \in \{\uparrow, \downarrow\}} 2\mu \tau^*ik_{sz}e^{\tau^*ik_{sz}z}(x^*_1 - k^2_\rho x^*_5) + (-\lambda k^2_\rho - \gamma k^2_{cz})e^{\tau^*ik_{cz}z}x^*_4.
\]

(b) \( [u_3] = 0 \) is equivalent to (5.139) and now interpreted as

\[
[T_3 \mathbf{J}_2 + T_4 \mathbf{J}_4] = 0 \iff [T_3] = [T_4] = 0,
\]

where

\[
T_3 = \sum_{* \in \{\uparrow, \downarrow\}} -k^2_\rho e^{\tau^*ik_{sz}z}x^*_3 + \tau^*ik_{cz}e^{\tau^*ik_{cz}z}x^*_2,
\]

\[
T_4 = \sum_{* \in \{\uparrow, \downarrow\}} e^{\tau^*ik_{sz}z}(x^*_1 - k^2_\rho x^*_5) + \tau^*ik_{cz}e^{\tau^*ik_{cz}z}x^*_4.
\]

(c) \( [T_{31}] = [T_{32}] = 0 \) is equivalent to (5.142) and now interpreted as

\[
[T_5 \mathbf{J}_1 + T_6 \mathbf{J}_3 + T_7 \mathbf{J}_5] = 0 \iff [T_5] = [T_6] = [T_7] = 0,
\]
where

\[ T_5 = \sum_{* \in \{\uparrow, \downarrow\}} \mu k_{sz}^2 e^{\tau^* i k_{sz} z} x_1^*, \]  

\[ T_6 = \sum_{* \in \{\uparrow, \downarrow\}} \mu (k_{sz}^2 - k_{\rho}^2) e^{\tau^* i k_{sz} z} x_3^* + 2 \mu \tau^* i k_{cz} e^{\tau^* i k_{cz} z} x_2^*, \]  

\[ T_7 = \sum_{* \in \{\uparrow, \downarrow\}} \mu e^{\tau^* i k_{sz} z} x_1^* + \mu (k_{sz}^2 - k_{\rho}^2) e^{\tau^* i k_{sz} z} x_5^* + 2 \mu \tau^* i k_{cz} e^{\tau^* i k_{cz} z} x_4^*. \]  

(d) \([u_1] = [u_2] = 0\) is equivalent to (5.145) and now interpreted as

\[ [T_8 J_1 + T_9 J_3 + T_{10} J_5] = 0 \iff [T_8] = [T_9] = [T_{10}] = 0, \]  

where

\[ T_8 = \sum_{* \in \{\uparrow, \downarrow\}} -\tau^* i k_{sz} e^{\tau^* i k_{sz} z} x_1^*, \]  

\[ T_9 = \sum_{* \in \{\uparrow, \downarrow\}} -\tau^* i k_{sz} e^{\tau^* i k_{sz} z} x_3^* + e^{\tau^* i k_{cz} z} x_2^*, \]  

\[ T_{10} = \sum_{* \in \{\uparrow, \downarrow\}} -\tau^* i k_{sz} e^{\tau^* i k_{sz} z} x_5^* + e^{\tau^* i k_{cz} z} x_4^*. \]  

The interface equations can be divided into two groups of continuity equations to solve, depending on the involvement of the unknowns: \(x_1^*, x_4^*\) and \(x_5^*\) are involved in the continuity equations of \(T_2, T_4, T_5, T_7, T_8\) and \(T_{10}\), while \(x_2^*\) and \(x_3^*\) appear in the other group with \(T_1, T_3, T_6\) and \(T_9\).

**Remark 5.17 (S-wave and P-wave partition in the \(R^0\)-matrix basis formulation).** When using \(x_1^*, \ldots, x_5^*\) to represent \(T_1, \ldots, T_{10}\), the S-wave parts which have \(z\)-dependence \(e^{\pm i k_{sz} z}\) are only related with \(x_1^*, x_3^*\) and \(x_5^*\), and the P-wave parts are only related with \(x_2^*\) and \(x_4^*\).

In the matrix basis representation of \(X^*, x_1, x_3\) and \(x_5\) take the first and the second row of the matrix together with the basis matrices \(J_1, J_3\) and \(J_5\), and \(x_2, x_4\) take the last row of
\( X^* \) together with the basis matrices \( J_2 \) and \( J_4 \). Therefore, the rows of \( X^* \) naturally separates the elastic wave into the S-wave part and the P-wave part.

5.5.3.1. Case when there’s a vacuum half space

Now consider the case when a vacuum half space is added to the problem. Without loss of generality, suppose the top layer is replaced by the vacuum. For convenience we define \( X^r = 0 \) in the vacuum side. The interface equations at \( z = d_0 - 0 \) are the zero-traction conditions

\[
T_{31} = T_{32} = 0, \quad T_{33} = 0, \quad (5.169)
\]

which correspond to (5.142) and (5.136), respectively, except from the fact that wave propagation does not exist in the vacuum side. Hence the \( R^0 \)-matrix basis representation of (5.142) and (5.136) can be used to enforce these conditions, resulting in the following equations at \( z = d_0 - 0 \):

\[
\sum_{s \in \{\uparrow, \downarrow\}} (\mu k_s^2 e^{\tau^r i k_s z} J_1 + \mu e^{\tau^r i k_s z} J_5 + 2\mu \tau^r i k_c e^{\tau^r i k_c z} J_3) X^s = 0, \quad (5.170)
\]

\[
\sum_{s \in \{\uparrow, \downarrow\}} (2\mu \tau^r i k_s e^{\tau^r i k_s z} J_4 + (-\lambda k^2 - \gamma k_c^2) e^{\tau^r i k_c z} J_2) X^s = 0. \quad (5.171)
\]

The radiation condition (5.148) from \( z = +\infty \) is no longer needed. Again we can collect the constrain equations of \( X^r_t \) for \( 1 \leq t \leq L \), and the same theories can be applied. The case for two half-space vacuum layers is similar.

5.5.3.2. Case when there’s a fluid layer

In the case when one or more fluid layer is involved, our goal is to justify that \( \hat{G} \) still has the matrix basis formulation within each layer.

Since we suppose the source is from a solid layer, in any fluid layer \( \hat{G} = \hat{G}^r \). Because the
S-wave does not propagate in fluid media, the formulation (5.134) can not be directly used. Therefore, we must take a step back to the acoustic wave equation in the fluid

\[
\lambda \nabla \nabla \cdot \mathbf{G}^r + \omega^2 \rho \mathbf{G}^r = 0, \quad (5.172)
\]

\[
\nabla \times \mathbf{G}^r = 0. \quad (5.173)
\]

In the frequency domain, the general solution of these equations is given by

\[
\tilde{\mathbf{G}}^r = \begin{bmatrix} i k_x \\ i k_y \\ i k_{cz} \end{bmatrix} \begin{bmatrix} \psi_1^i \\ \psi_2^i \\ \psi_3^i \end{bmatrix}^T e^{ik_{cz}z} + \begin{bmatrix} i k_x \\ i k_y \\ -i k_{cz} \end{bmatrix} \begin{bmatrix} \psi_1^i \\ \psi_2^i \\ \psi_3^i \end{bmatrix}^T e^{-ik_{cz}z}, \quad (5.174)
\]

where each \( \psi_i^* \) does not depend on \( z \). Here we take an alternative form by just separating out the \( z \) variable

\[
\tilde{\mathbf{G}}^r = \tilde{\mathbf{G}}^{r\uparrow} e^{ik_{cz}z} + \tilde{\mathbf{G}}^{r\downarrow} e^{-ik_{cz}z}, \quad (5.175)
\]

and treat \( \tilde{\mathbf{G}}^{r*} \) as unknowns (which are independent of \( z \)). In the frequency domain, the first equation of (5.172) is reinterpreted as

\[
\lambda \left( -k_{cz}^2 \mathbf{J}_2 + i k_{cz} \mathbf{J}_3 + i k_{cz} \mathbf{J}_4 + \mathbf{J}_5 \right) \tilde{\mathbf{G}}^{r\uparrow} e^{ik_{cz}z} + \lambda \left( -k_{cz}^2 \mathbf{J}_2 - i k_{cz} \mathbf{J}_3 - i k_{cz} \mathbf{J}_4 + \mathbf{J}_5 \right) \tilde{\mathbf{G}}^{r\downarrow} e^{-ik_{cz}z} = 0, \quad (5.176)
\]

and the second equation is reinterpreted as \((\mathbf{J}_6 + \mathbf{J}_7 - \partial_z \mathbf{J}_9)\tilde{\mathbf{G}}^r = 0\), which is equivalent to the pair of equations

\[
(k_{\rho}^2 \mathbf{J}_2 + i k_{cz} \mathbf{J}_4) \tilde{\mathbf{G}}^{r\uparrow} e^{ik_{cz}z} + (k_{\rho}^2 \mathbf{J}_2 - i k_{cz} \mathbf{J}_4) \tilde{\mathbf{G}}^{r\downarrow} e^{-ik_{cz}z} = \mathbf{J}_6(\mathbf{J}_6 + \mathbf{J}_7 - \partial_z \mathbf{J}_9) \tilde{\mathbf{G}}^r = 0, \quad (5.177)
\]

\[
(k_{\rho}^2 \mathbf{J}_1 + \mathbf{J}_5)(\tilde{\mathbf{G}}^{r\uparrow} e^{ik_{cz}z} + \tilde{\mathbf{G}}^{r\downarrow} e^{-ik_{cz}z}) = \mathbf{J}_7(\mathbf{J}_6 + \mathbf{J}_7 - \partial_z \mathbf{J}_9) \tilde{\mathbf{G}}^r = 0,
\]

which are written using only \( \mathbf{J}_1, \cdots, \mathbf{J}_5 \). The radiation condition, if applied, is simply \( \tilde{\mathbf{G}}^{r*} = \)
0 for the corresponding prohibited propagation direction. The interface equations are a subset of (5.136), (5.139), (5.142) and (5.145). For each of these equations, in the solid side we adopt the representation using unknowns \( X^r * \), and in the fluid side we use the representation with unknowns \( \tilde{G}^r * \). The resulting series of constrain equations is a linear system of unknowns from each layer with coefficients in \( \mathbb{R}^0 \), so Theorem 5.10 can be applied. Together with the uniqueness of the Green’s function, we again conclude that each \( X^r * \) from solid layers and each \( \tilde{G}^r * \) from fluid layers have a \( \mathbb{R}^0 \)-matrix basis representation.

**Remark 5.18.** We can still separate the unknown \( \mathbb{R}^0 \)-matrix basis coefficients into groups exactly like the pure solid case, i.e. the equations of \( T_2, T_4, T_5, T_7, T_8 \) and \( T_{10} \) consist one group and the rest consist the other one.

5.5.4. The dyadic Green’s function in layered media with source in fluid

When the source is in a fluid layer, the Green’s function \( g_p \) with respect to the pressure is a scalar function, and the corresponding \( g_u \) with respect to the displacement is a vector function, hence no \( 3 \times 3 \) tensor seems involved. However, we can still simplify the formulation of \( g_u \) by using a vector version derived from the matrix basis (5.8).

5.5.4.1. The vector basis

Define \( 3 \times 1 \) vectors

\[
\begin{align*}
\mathbf{j}_2 &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \\
\mathbf{j}_3 &= \begin{bmatrix} i k_x \\ i k_y \\ 0 \end{bmatrix}, \\
\mathbf{j}_7 &= \begin{bmatrix} i k_y \\ -i k_x \\ 0 \end{bmatrix},
\end{align*}
\]

which are the third column of \( J_2, J_3 \) and \( J_7 \), respectively. The vectors \( \mathbf{j}_2, \mathbf{j}_3 \) and \( \mathbf{j}_7 \) are exactly all the non-trivial third columns of the matrix basis \( J_1, \cdots, J_9 \). Hence the product between
a $J_i$ matrix and a $j_j$ vector

$$J_i \cdot j_j = J_i \cdot J_j e_3 = (J_i J_j) e_3$$

(5.179)

is still a linear combination of $j_2, j_3$ and $j_7$ (with coefficients $\pm 1, \pm k^2$).

Similar to the linear spaces of block matrices defined in (5.12), for any $p, q \in \mathbb{N}$ and any subfield $K \subset F$, define

$$r_{p \times q}(K) = \{K_2 \otimes j_2 + K_3 \otimes j_3 : K_2, K_3 \in K^{p \times q}\},$$

$$i_{p \times q}(K) = \{K_7 \otimes j_7 : K_7 \in K^{p \times q}\},$$

$$m_{p \times q}(K) = r_{p \times q}(K) \oplus i_{p \times q}(K).$$

(5.180)

Note that $i_{p \times q}(F) = F^{3p \times q}$ because $j_2, j_3$ and $j_7$ are linearly independent. The solution filtering Theorem 5.10 is accordingly adjusted as follows.

**Theorem 5.19 (Vector solution filtering).** Suppose $p, q, r \in \mathbb{N}$, the block matrices $\bar{A} \in R_{p \times r}(F_0)$, $\bar{X} \in m_{r \times q}(F)$ and $\bar{B} \in R_{p \times q}(F_0)$ satisfy $\bar{A} \cdot \bar{X} = \bar{B}$. Then, there exists a “filtered” block matrix $\bar{X}_0 \in R_{r \times q}(F_0)$ such that $\bar{A} \cdot \bar{X}_0 = \bar{B}$.

**Proof.** For any $u, v \in \mathbb{N}$, the mapping

$$P_{u \times v} : K_2 \otimes j_2 + K_3 \otimes j_3 + K_7 \otimes j_7 \in m_{u \times v}(K) \mapsto K_2 \otimes J_2 + K_3 \otimes J_3 + K_7 \otimes J_7 \in M_{u \times v}(K)$$

(5.181)

is bijective. It is obvious that

$$\bar{A} \cdot P_{r \times q}(\bar{X}) = P_{p \times q}(\bar{B})$$

(5.182)

and that $P_{p \times q}(\bar{B}) \in R_{p \times q}(F_0)$. By Theorem 5.10, there exists $\bar{X}_M \in R_{r \times q}(F_0)$ such that

$$\bar{A} \cdot \bar{X}_M = P_{p \times q}(\bar{B}).$$

(5.183)
Then, by removing the first and the second column of each $3 \times 3$ block from $\bar{X}_M$ we get the desired result.

\[ \Box \]

5.5.4.2. Derivation of the vector basis formulation

We begin with briefly repeating the reaction field decomposition. The free space Green’s function $g_u^f$ in fluid with respect to the displacement $u$ was defined in (5.114). In the frequency domain,

\[
\hat{g}_u^f = \frac{1}{\omega^2 \rho'} \begin{bmatrix} i k_x \\ i k_y \\ \partial_z \end{bmatrix} = \frac{1}{\omega^2 \rho'} \begin{bmatrix} i k_x \\ i k_y \\ \partial_z \end{bmatrix} \frac{i e^{ik_{cz,t'} |z-z'|}}{2 k_{cz,t'}}
\]

(5.184)

\[
= e^{ik_{cz,t'z}} g_u^{f\uparrow} + e^{-ik_{cz,t'z}} g_u^{f\downarrow},
\]

\[
= (ik_{cz,t'}j_2 + j_3)e^{ik_{cz,t'z}} g_u^{f\uparrow} + (-ik_{cz,t'}j_2 + j_3)e^{-ik_{cz,t'z}} g_u^{f\downarrow},
\]

where

\[
g_u^{f\ast} = (\tau^*ik_{cz,t'}j_2 + j_3)g_u^{f\ast}, \quad g_u^{f\uparrow} = 1_{(z>z')} \frac{i e^{-ik_{cz,t'z'}}}{2 \omega^2 \rho' k_{cz,t'}}, \quad g_u^{f\downarrow} = 1_{(z<z')} \frac{i e^{ik_{cz,t'z'}}}{2 \omega^2 \rho' k_{cz,t'}}.
\]

(5.185)

The reaction field of the Green’s function is defined by

\[
g_u^r = g_u - \delta_{t,t'} g_u^f
\]

(5.186)

in each layer. In a fluid layer, the reaction field Green’s function $g_u^r$ satisfies the acoustic wave equations

\[
\lambda \nabla \nabla \cdot g_u^r + \omega^2 \rho g_u^r = 0,
\]

\[
\nabla \times g_u^r = 0,
\]

(5.187)

181
as in (5.172). When treated as ordinary differential equations in $z$ in the frequency domain, the general solutions to these equations are

$$
\hat{g}_u^r = (i k c_z j_2 + j_3) e^{i k c_z z} \psi^\uparrow + (-i k c_z j_2 + j_3) e^{-i k c_z z} \psi^\downarrow,
$$

(5.188)

where the scalars $\psi^\ast$ do not depend on $z$. We take an alternative form

$$
\hat{g}_u^r = e^{i k c_z z} \hat{g}_u^r \uparrow + e^{-i k c_z z} \hat{g}_u^r \downarrow
$$

(5.189)

where $\hat{g}_u^{r\uparrow} = g_u^{r\uparrow}(\tau^* i k c_z j_2 + j_3)$, satisfying the equations

$$
\lambda (-k_c^2 J_2 + i k c_z J_3 + i k c_z J_4 + J_5) \hat{g}_u^r \uparrow e^{i k c_z z} + \lambda (-k_c^2 J_2 - i k c_z J_3 - i k c_z J_4 + J_5) \hat{g}_u^r \downarrow e^{-i k c_z z} = \vec{0},
$$

(5.190)

and

$$
(h^2 \rho J_2 + i k c_z J_4) \hat{g}_u^r \uparrow e^{i k c_z z} + (h^2 \rho J_2 - i k c_z J_4) \hat{g}_u^r \downarrow e^{-i k c_z z} = J_0 (J_6 + J_7 - \partial_z J_9) \hat{g}_u^r = \vec{0},
$$

(5.191)

which are the vector versions of (5.176) and (5.177), respectively. Then, let

$$
\hat{g}_u^* = \delta_{\ell,\ell'} \hat{g}_u^{r\ast} + \hat{g}_u^{r\ast}.
$$

One can verify that $\hat{g}_u^*$ satisfies the same equations as the above of $\hat{g}_u^{r\ast}$ within each layer provided $z \neq z'$. In a solid layer, the reaction field Green’s function $g_u^r = g_u$ satisfies the homogeneous elastic wave equation

$$
(\lambda + \mu) \nabla \nabla \cdot g_u^r + \mu \nabla^2 g_u^r + \omega^2 \rho g_u^r = \vec{0}
$$

(5.192)
as in (5.125). In the frequency domain we again have the general solution

\[ \hat{g}_u = \sum_{* \in \{\uparrow, \downarrow\}} \left( (-\tau^* i k_{sz} J_1 + J_4) e^{\tau^* i k_{sz} z} + (\tau^* i k_{cz} J_2 + J_3) e^{\tau^* i k_{cz} z} \right) \mathbf{x}^{r*}, \tag{5.193} \]

where \( \mathbf{x}^{r*} \) are \( 3 \times 1 \) vectors within each layer independent from \( z \). This repeats the result of (5.128).

The interface equations and the radiation equations are interpreted as the vector versions of the tensor case we’ve discussed in Section 5.5.3, just by replacing the unknowns from tensors to vectors, and replacing the free-space contribution from \( \hat{g}_f^* \) to \( \hat{g}_u^* \). Therefore, by Theorem 5.19, the solution to the unknown vectors, which is known uniquely determined by the constrain equations (since the coefficients of unknowns are shared with the tensor version), has the vector basis representation, i.e. each \( \hat{g}_u^* \in \mathfrak{r}_{1 \times 1}(\mathbb{R}) \) and each \( \mathbf{x}^{r*} \in \mathfrak{r}_{1 \times 1}(\mathbb{R}) \).

Finally, we use the vector basis to reinterpret the interface equations. Suppose in the fluid layers

\[ \hat{g}_u^* = g_{u,2}^* \mathbf{j}_2 + g_{u,3}^* \mathbf{j}_3. \tag{5.194} \]

Since the displacement is curl-free, the general solution form (5.188) implies

\[ g_{u,2}^* = \tau^* i k_{cz} g_{u,3}^*. \tag{5.195} \]

Hence we have the reaction field decomposition in fluid layers

\[ \hat{g}_u = (i k_{cz} \mathbf{j}_2 + \mathbf{j}_3) e^{i k_{cz} z} g_u^* + (-i k_{cz} \mathbf{j}_2 + \mathbf{j}_3) e^{-i k_{cz} z} g_u^*, \tag{5.196} \]

where

\[ g_u^* = \delta_{\ell,\ell'} g_u^{f*} + g_u^{r*}. \tag{5.197} \]
Then, let us suppose in the solid layers

\[ x^{r*} = x_2^*j_2 + x_3^*j_3, \quad (5.198) \]

and the general solution in solid layers (5.193) is simplified as

\[ \dot{g}_u = \ddot{g}_u = \sum_{* \in \{\uparrow, \downarrow\}} \left( (-\tau^*ik_{sz}s_j^3 - k_{p}\bar{j}_2)x_3^*e^{\tau^*ik_{sz}s} + (\tau^*ik_{cz}j_2 + j_3)x_2^*e^{\tau^*ik_{cz}s} \right), \quad (5.199) \]

i.e. rows of \( x^{r*} \) again reveals the wave decomposition in the solid media, with variable \( x_2^* \) for the S-wave and \( x_3^* \) for the P-wave. For the simplification of the interface equations and the radiation equations with the vector basis \( j_2 \) and \( j_3 \) and their coefficients, it suffices to repeat the matrix forms used in (5.136), (5.139), (5.142) and (5.145).

(a) \( [T_{33}] = 0 \) is equivalent to

\[ [(\lambda J_4 + \gamma \partial_z J_2)\dot{g}_u] = \vec{0}. \quad (5.200) \]

In solid it’s expanded as the terms in the brackets of (5.155) multiplied by \( j_2 \)

\[ (\lambda J_4 + \gamma \partial_z J_2)\dot{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} -2\mu \tau^*ik_{sz}s_j^2e^{\tau^*ik_{sz}s}x_3^* + (-\lambda k_{p}^2 - \gamma k_{cz}^2)e^{\tau^*ik_{cz}s}x_2^* \right)j_2, \quad (5.201) \]

while in fluid

\[ (\lambda J_4 + \gamma \partial_z J_2)\dot{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} -\omega^2\rho e^{\tau^*ik_{cz}s}g_u^* \right)j_2. \quad (5.202) \]

(b) \( [u_3] = 0 \) is equivalent to

\[ [J_2\dot{g}_u] = \vec{0}. \quad (5.203) \]
In solid the terms inside the brackets are expanded as

\[ J_2 \hat{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} -k^2_\rho e^{\tau^*i k_{cz} z} x_3^* + \tau^*i k_{cz} e^{\tau^*i k_{cz} z} x_2^* \right) j_2 \] (5.204)

and in fluid

\[ J_2 \hat{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} \tau^*i k_{cz} e^{\tau^*i k_{cz} z} g_u^* \right) j_2. \] (5.205)

(c) \([T_{31}] = [T_{32}] = 0\) is equivalent to

\[ [\mu (\partial_z J_1 + J_3) \hat{g}_u] = \vec{0} \] (5.206)

which does not influence the fluid, and the solid side should satisfy

\[ \mu (\partial_z J_1 + J_3) \hat{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} \mu (k_{sz}^2 - k^2_\rho) e^{\tau^*i k_{sz} z} x_3^* + 2\mu \tau^*i k_{cz} e^{\tau^*i k_{cz} z} x_2^* \right) j_3 = \vec{0}. \] (5.207)

(d) \([u_1] = [u_2] = 0\) is equivalent to

\[ [J_1 \hat{g}_u] = \vec{0}. \] (5.208)

In solid we have

\[ J_1 \hat{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} -\tau^*i k_{sz} e^{\tau^*i k_{sz} z} x_3^* + e^{\tau^*i k_{cz} z} x_2^* \right) j_3 \] (5.209)

and in fluid,

\[ J_1 \hat{g}_u = \left( \sum_{* \in \{\uparrow, \downarrow\}} e^{\tau^*i k_{cz} z} g_u^* \right) j_3. \] (5.210)

The above equations are exactly the variants of \(T_1, T_3, T_6\) and \(T_9\) defined before. The 6-term group of unknowns previously in the solid-source case is gone.
5.6. A FMM framework using the matrix basis formulation

Here we briefly illustrate the required modification from the algorithm in Chapter 3 for applying FMM to the Maxwell’s equations and the elastic wave equation in layered media.

First, we refer to Section 5.3 and Section 5.5 for evaluation of the LMDG in the frequency domain. The LMDG consists of the free-space part and the reaction field, and the reaction field is written either in a $3 \times 3$ tensor with matrix basis $J_1, \cdots, J_9$, or in a 3-dimensional vector with vector basis $j_2, j_3, j_7$. Let $j(k_x, k_y)$ be any entry of the basis functions, and $f(k_\rho)$ be the corresponding coefficient, which is a function of $k_\rho$ with exponential dependence on $z$ and $z'$. In terms of the reaction field decomposition, by using local coordinates, we have

$$f(k_\rho) = \sum_{*,* \in \{\uparrow, \downarrow\}} e^{ik_\ell_\tau^\ast (z-d_\tau^\ast)} e^{ik_\ell^\ast_\tau^\ast (z'-d'^\ast_\tau)} f^{**}(k_\rho).$$  \hspace{1cm} (5.211)

To derive the far-field expansions, we use $f^{**}(k_\rho)j(k_x, k_y)$ as the reflection/transmission coefficient like the $\sigma^{**}_{\ell\ell}(k_\rho)$ functions in (3.17). The ME, LE and M2L are proposed like in (3.67), (3.73) and (3.76), respectively. The difference is the way to reduce to single integrals, where we reinterpret the basis factor $j(k_x, k_y)$ using the polar angles $(k_\rho, \alpha)$, and the exponential dependence of $\alpha$ will be reduced in the conversion to single integrals in (3.87). An example has been shown in (5.86). The following list covers all the cases for the angular part of $j(k_x, k_y)$:

\begin{align*}
\frac{ik_x}{k_\rho} &= \frac{i}{2} e^{i\alpha} + \frac{i}{2} e^{-i\alpha}, \\
\frac{ik_y}{k_\rho} &= \frac{1}{2} e^{i\alpha} - \frac{1}{2} e^{-i\alpha}, \\
\frac{-k_x^2}{k_\rho^2} &= -\frac{1}{2} - \frac{1}{4} e^{2i\alpha} - \frac{1}{4} e^{-2i\alpha}, \\
\frac{-k_x k_y}{k_\rho^2} &= \frac{i}{4} e^{2i\alpha} - \frac{i}{4} e^{-2i\alpha}, \\
\frac{-k_y^2}{k_\rho^2} &= -\frac{1}{2} + \frac{1}{4} e^{2i\alpha} + \frac{1}{4} e^{-2i\alpha}.
\end{align*}

\hspace{1cm} (5.212)
5.7. Conclusion

In this chapter, a matrix basis formulation is proposed for handling the LMDG of the Maxwell’s equations and the elastic wave equation, where the Green’s functions are usually in $3 \times 3$ tensors. The formulation is then used to simplify the representation and derivation of the Green’s functions in both cases with specific physical interface conditions. In particular, the interface conditions for the electromagnetic waves is reduced to decoupled conditions for the matrix basis coefficients. A numerical test of the electric field and the magnetic field in a 10-layer medium verifies the matrix basis formulation. For the elastic wave equation, various phases of contacting media have been studied. As a corollary, a vector basis formulation is introduced for the LMDG of the elastic wave equation with sources in a fluid layer. In the end, we propose a guideline for extending the Helmholtz FMM in 3-D layered media (see Chapter 3) to these equations.
Chapter 6

FBSDE based deep neural network algorithms for high-dimensional quasilinear parabolic PDEs

The content in this chapter has been submitted for publication in collaboration with Wei Cai under the title:


6.1. Introduction

The relationship between stochastic processes and the solution of partial differential equations represents one of the high achievements of probability theory in potential theory research [26], represented by the celebrated Feynman–Kac formula in linear parabolic and elliptic PDEs as a result of the Kolmogorov backward equation for the generator of the stochastic process for the former [61] and Dynkin formula for the latter [60]. The recent work by Pardoux–Peng [27] has extended the concept of classic linear Feynman–Kac formula to a nonlinear version, which connects the solution of a quasilinear parabolic PDE to a coupled pair of forward and backward stochastic processes. This extraordinary development has made much impact in the mathematical finance in option pricing [30].

Meanwhile, in the field of scientific computing, this connection between SDEs and quasilinear PDEs has inspired new approaches of solving high dimensional parabolic partial differential equations (PDEs) which are ubiquitous in material sciences such as the Allen–Cahn equations for phase transitions, and quantum mechanics such as Schrodinger equations as well as option pricing such as the Black–Scholes and Hamilton–Jacobi–Bellman equations.
For PDEs in high dimensions, the main challenge of the traditional numerical methods, such as finite element, finite difference and spectral methods, is the curse of dimensionality, namely, the number of unknowns in the discretized systems for the PDEs grows exponentially in terms of the dimension of the problem. Recently, machine learning approaches based on the deep neural network have taken advantage of the Pardoux–Peng’s theory for forward and backward stochastic differential equations (FBSDEs) and PDEs. The solution to the PDEs can be learned by sampling the paths of involved stochastic processes, which are discretized in time by the classic Euler–Maruyama scheme [47]. The first such an attempt was done in the work of [38], where neural network was used as an approximator to the gradient of the PDEs solutions, while the PDE’s solution follows the dynamics of the FBSDEs, and the learning was carried out by imposing the terminal condition provided by the parabolic PDEs. Another approach [63] is to approximate the PDE’s solution itself by a deep neural network, which also provides the gradient of the solution as required by the FBSDEs, the learning is then carried out by minimizing the difference between the solution given by the discretized SDEs and that given by the DNN at all discretization time stations.

In this chapter, an improved learning scheme will be proposed based on a similar approach in [63] but with more mathematical consistence and reasoning for the learning processes, to ensure the numerical methods’ mathematical consistency and improved convergence for the solutions to PDEs.

The rest of the chapter is organized as follows. In Section 6.2, we will review the Pardoux–Peng theory, which establishes the relation between FBSDEs and quasilinear parabolic PDEs, with an emphasis on the relation between the classic Feynman–Kac formula and the nonlinear version represented by the Pardoux–Peng theory. Section 6.3 will first review the algorithm proposed in [63], and then two new improved methods will be proposed. In Section 6.4 we generalize the algorithms proposed in Section 6.3 to compute the committor function in the high-dimensional space, which satisfies a time-independent PDE. Section 6.5 will present numerical results of the new schemes for solving a 100-dimensional Black–Scholes–Barenblatt
equation, as well as the committor function in a model problem. Finally, a conclusion will be given in the ending section.

6.2. Pardoux–Peng theory on FBSDEs and quasilinear parabolic PDEs

In this chapter, we consider the scalar solution \( u(t, x), t \in [0, T], x \in \mathbb{R}^d \) for the following \( d \)-dimensional parabolic PDE

\[
\partial_t u + \frac{1}{2} \text{Tr}[\sigma \sigma^T \nabla \nabla u] + \mu \cdot \nabla u = \phi, \tag{6.1}
\]

with a terminal condition

\[
u(T, x) = g(x), \tag{6.2}
\]

where \( \sigma = \sigma(t, x, u), \phi = \phi(t, x, u, \nabla u), \mu = \mu(t, x, u, \nabla u) \) are functions with ranges in with dimensions \( d \times d, 1 \) and \( d \), respectively. We are interested in finding the initial value \( u(0, x_0) \) given \( x_0 \in \mathbb{R}^d \). Therefore, in some sense our problem is similar to a time reverse problem for a time reversed version of (6.1) with an initial data at \( t = 0 \).

Following Pardoux–Peng in [27], under certain regularity conditions, the forward-backward SDE reformulation gives a nonlinear implicit Feynman–Kac formula for the solution of the parabolic PDE (6.1). The FBSDEs are proposed as follows. Let \( W_t = (W^1_t, \cdots, W^d_t) \) where each \( W^j_t \) is a standard Brownian motion. Let \( \{ \mathcal{F}_t : 0 \leq t \leq T \} \) be its natural filtration on the time interval \( [0, T] \). Then, we have the equations of stochastic processes \( X_t, Y_t \) and \( Z_t \) in \( d, 1 \) and \( d \) dimensions that are adaptive to the filtration \( \{ \mathcal{F}_t : 0 \leq t \leq T \} \), respectively,

\[
dX_t = \mu(t, X_t, Y_t, Z_t)dt + \sigma(t, X_t, Y_t)dw_t, \tag{6.3}
\]

\[
X_0 = x_0,
\]

\[
dY_t = \phi(t, X_t, Y_t, Z_t)dt + Z^T_t \sigma(t, X_t, Y_t)dW_t, \tag{6.4}
\]

\[
Y_T = g(X_T).
\]
If $\mu$ and $\sigma$ do not explicitly depend on $Y_t$ or $Z_t$, the FBSDEs are decoupled.

We can easily show that processes defined by
\[ Y_t = u(t, X_t), \quad Z_t = \nabla u(t, X_t) \] (6.5)
in fact satisfy the above equations (6.3) and (6.4). By using the Itô’s formula \[60\] and the forward SDE of $X_t$, we have
\[
dY_t = du(t, X_t) = \partial_t u dt + \nabla u \cdot dX_t + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \partial_{ij} u [X^i, X^j]_t \]
\[ = \partial_t u dt + \nabla u \cdot (\mu dt + \sigma \cdot dW_t) + \frac{1}{2} \text{Tr}[\sigma \sigma^T \nabla \nabla u] dt \]
\[ = \left( \partial_t u + \nabla u \cdot \mu + \frac{1}{2} \text{Tr}[\sigma \sigma^T \nabla u] \right) dt + Z_t^T \sigma dW_t, \]
which gives the PDE (6.1) by comparing (6.6) with the backward SDE (6.4) for $Y_t$.

The determination of the third stochastic process $Z_t$ from the two SDEs in (6.3) and (6.4) makes use of the martingale representation theory \[46\]. Consider the following special case of the backward SDE (6.4) as an example:
\[
Y_t + \int_t^T f(s, X_s) ds + \int_t^T Z_s \cdot dW_s = g(X_T), \quad 0 \leq t \leq T, \] (6.7)
i.e. $\mu(t, x, u, \nabla u) = f(t, x)$, and $\sigma(t, x, u) = I_{d \times d}$ is the identity matrix. By taking the conditional expectation with respect to $\mathcal{F}_t$, we have
\[ Y_t = \mathbb{E} [Y_t | \mathcal{F}_t] = \mathbb{E} \left[ g(X_T) - \int_t^T f(s, X_s) ds \bigg| \mathcal{F}_t \right], \quad 0 \leq t \leq T. \] (6.8)
Next, define the following martingale
\[ L_t = \mathbb{E} \left[ g(X_T) - \int_t^T f(s, X_s) ds \bigg| \mathcal{F}_t \right], \quad 0 \leq t \leq T, \] (6.9)
where $L_0 = Y_0$. By the martingale representation theorem [46], there exists a process $Z^*_t$ such that

$$L_t = Y_0 + \int_0^t Z^*_s \cdot dW_s, \quad 0 \leq t \leq T.$$  \hfill (6.10)

The stochastic process $Z^*_t$ is unique in the sense that

$$\int_0^T \|Z^*_t - Z^*_s\|^2 dt = 0, \quad \text{a.s.}$$  \hfill (6.11)

if $Z^*_t$ satisfies the same condition (6.10) as $Z^*_t$ (cf. [46]).

Meanwhile, we can show that $Z_t = Z^*_t$ solves the backward SDE (6.7),

\begin{align*}
Y_t + \int_t^T f(s, X_s)ds + \int_t^T Z^*_s \cdot dW_s - g(X_T) \\
= \mathbb{E} \left[ g(X_T) - \int_t^T f(s, X_s)ds \bigg| \mathcal{F}_t \right] + \left( \int_0^T - \int_t^T \right) f(s, X_s)ds + (L_T - L_t) - g(X_T) \\
= \int_0^T f(s, X_s)ds + L_T - g(X_T) \\
= L_T - \mathbb{E} \left[ g(X_T) - \int_0^T f(s)ds \bigg| \mathcal{F}_T \right] \\
= 0.
\end{align*}

Connection with the classic Feynman–Kac formula is interpreted as follows. If in the parabolic PDE (6.1), $\phi$ has a linear dependence on $u$, i.e.

$$\phi(t, x, u, \nabla u) = c(t, x)u(t, x) + f(t, x),$$  \hfill (6.12)

then, the backward SDE (6.4) has an explicit solution

$$Y_t = e^{-\int_t^T c(s, X_s)ds} g(X_T) - \int_t^T e^{-\int_t^s c(r, X_r)dr} f(s, X_s)ds - \int_t^T e^{-\int_t^s c(r, X_r)dr} Z^T_{s} \sigma(s, X_s, Y_s) dW_s.$$  \hfill (6.13)
By taking the conditional expectation on both sides, we arrive at

\[ Y_t = \mathbb{E} \left[ e^{-f_t^T c(s, X_s) ds} g(X_T) - \int_t^T e^{-f_s^T c(\tau, X_\tau) d\tau} f(s, X_s) ds \right] \mathcal{F}_t. \] (6.14)

For \((t, x) \in [0, T] \times \mathbb{R}^d\), using \(X_t = x\) as the initial condition of the forward SDE (6.3) on the time interval \([t, T]\) instead of \(X_0 = x_0\), the traditional Feynman–Kac formula [60] is recovered,

\[ u(t, x) = \mathbb{E} \left[ e^{-\int_t^T c(s, X_s) ds} g(X_T) - \int_t^T e^{-\int_s^\tau c(r, X_r) dr} f(s, X_s) ds \right] X_t = x. \] (6.15)

For a general parabolic equation with a nonlinear function \(\phi(s, x, u, \nabla u)\), we have

\[ Y_t = \mathbb{E} \left[ g(X_T) - \int_t^T \phi(s, X_s, Y_s, Z_s) ds \right] \mathcal{F}_t, \]

and for given \((t, x) \in [0, T] \times \mathbb{R}^d\), the following nonlinear equation for \(u(t, x)\) is obtained

\[ u(t, x) = \mathbb{E} \left[ g(X_T) - \int_t^T \phi(s, X_s, u(s, X_s), \nabla u(s, X_s)) ds \right] X_t = x. \] (6.16)

6.3. FBSDE based neural network algorithms for quasilinear parabolic PDEs

The learning of the solution will be based on the sample paths of the FBSDEs, which are linked to the PDE solution in (6.5). Paths of the FBSDEs will be produced by a time discretization algorithm with samples of the Brownian motion \(W_t\).

Let \(0 = t_0 < \cdots < t_N = T\) be a uniform partition of \([0, T]\). On each interval \([t_n, t_{n+1}]\), define time and Brownian motion increments as

\[ \Delta t_n = t_{n+1} - t_n, \quad \Delta W_n = W_{t_{n+1}} - W_{t_n}. \] (6.17)

Denoting \(X_{t_n}, Y_{t_n}\) and \(Z_{t_n}\) by \(X_n, Y_n\) and \(Z_n\), respectively, and applying the Euler–Maruyama
scheme to the FBSDEs (6.3) and (6.4), respectively, we have

\[
X_{n+1} \approx X_n + \mu(t_n, X_n, Y_n, Z_n) \Delta t_n + \sigma(t_n, X_n, Y_n) \Delta W_n,
\]

(6.18)

\[
Y_{n+1} \approx Y_n + \phi(t_n, X_n, Y_n, Z_n) \Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n) \Delta W_n.
\]

(6.19)

Due to the relationship with the parabolic PDE, the solution to the parabolic PDE provides an alternative representation for \( Y_{n+1} \) and \( Z_{n+1} \),

\[
Y_{n+1} = u(t_{n+1}, X_{n+1}),
\]

(6.20)

\[
Z_{n+1} = \nabla u(t_{n+1}, X_{n+1}).
\]

(6.21)

In this chapter, fully connected networks of \( L \) hidden layers will be used, which are given in the form

\[
f_\theta(x) = W^{[L-1]} \sigma \circ (\cdots (W^{[1]} \sigma \circ (W^{[0]}(x) + b^{[0]}) + b^{[1]}) \cdots) + b^{[L-1]},
\]

(6.22)

where \( W^{[1]}, \ldots, W^{[L-1]} \) and \( b^{[1]}, \ldots, b^{[L-1]} \) are the weight matrices and bias unknowns, respectively, denoted collectively by \( \theta \), to be optimized via the training, \( \sigma(x) \) is the activation function and \( \circ \) is the application of the activation function \( \sigma \) applied to a vector quantity component-wisely.

6.3.1. Existing FBSDE based neural network algorithms

6.3.1.1. Deep BSDE

The Deep BSDE [38] trains a network to approximate the random value \( Y_N \) at time \( t = T \), where \( X_0 = x_0 \) is the input. \( Y_0, Z_0 \) are trainable variables, where \( Y_0 \) is the targeted quantity of the algorithm. \( W_n, X_n, 0 \leq n \leq N \) can be obtained similarly as before. The algorithm can be organized as follows.
1. The initial value $X_0 = x_0$ is given. Trainable variables $Y_0$ and $Z_0$ are randomly initialized.

2. On each time interval $[t_n, t_{n+1}]$, use the Euler–Maruyama scheme to calculate $X_{n+1}$ and $Y_{n+1}$ as in (6.18) and (6.19). Then, train a fully connected feedforward network

$$f^{(n+1)}_{\theta}(\cdot) \approx \nabla u(t_{n+1}, \cdot)$$

(6.23)

where $f^{(n+1)}_{\theta}(\cdot)$ is a fully connected neural network with $H$ hidden layers of the form given in (6.22). Activation functions including ReLU, Tanh, Sigmoid, etc. can been used.

3. Connect all quantities (subnetworks $f^{(n)}_{\theta}(\cdot)$, etc.) at $\{t_n\}_{n=1}^N$ to form a network that outputs $Y_N$, which is expected to be an approximation of $u(t_N, X_N)$.

4. The loss function is then defined by a Monte Carlo approximation of

$$E\|Y_N - g(X_N)\|^2.$$  

(6.24)

The Deep BSDE has been shown to give convergent numerical results for various high dimensional parabolic equations [38]. A posteriori estimate suggests strong convergence of half order [39].

**Remark 6.1.** The Deep BSDE method from [38] trains the network for the specific initial data $X_0 = x_0$ and yields only an approximation to the PDE solution $Y_0 = u(0, x_0)$. Therefore, once the desired initial data is changed, a new training may have to be carried out. Also, the total size of $N$ individual sub-networks used to approximate $Z_n = \nabla u(t_n, X_n)$, $n = 1, \cdots, N-1$ will grow linearly in terms of the number of time discretization steps $N$, resulting in large amount of training parameters if higher accuracy of the PDE solution is desired.
6.3.1.2. FBSNNs, Scheme 1

The FBSNNs [63] trains a network \( u_\theta(t, x) \) that directly approximates the solution to the PDE (6.1) in some region in the \((t, x)\) space. The network has a fixed size of number of hidden layers and neurons per layer. The algorithm can be organized as follows.

1. The initial value \( X_0 = x_0 \) is given. Evaluate \( Y_0 \) and \( Z_0 \) using the network

\[
Y_0 = u_\theta(t_0, X_0), \quad Z_0 = \nabla u_\theta(t_0, X_0).
\]

The gradient above is calculated by an automatic differentiation.

2. On each time interval \([t_n, t_{n+1}]\), use the Euler–Maruyama scheme (6.18) to calculate \( X_{n+1} \), and use the network for \( Y_{n+1} \) and \( Z_{n+1} \), i.e.

\[
X_{n+1} = X_n + \mu(t_n, X_n, Y_n, Z_n) \Delta t_n + \sigma(t_n, X_n, Y_n) \Delta W_n,
\]
\[
Y_{n+1} = u_\theta(t_{n+1}, X_{n+1}),
\]
\[
Z_{n+1} = \nabla u_\theta(t_{n+1}, X_{n+1}).
\]

On the other hand, calculate a reference value \( Y^*_n \) using the Euler–Maruyama scheme (6.19)

\[
Y^*_{n+1} = Y_n + \phi(t_n, X_n, Y_n, Z_n) \Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n) \Delta W_n.
\]

3. The loss function is taken as a Monte Carlo approximation of

\[
\mathbb{E} \left[ \sum_{n=1}^N \|Y_n - Y^*_n\|^2 + \|Y_N - g(X_N)\|^2 + \|Z_N - \nabla g(X_N)\|^2 \right].
\]

In this chapter, we will name the above numerical method Scheme 1. In order to compare the training results using different values of \( N \), the loss function for Scheme
1 is modified as

\[
L_1[u_\theta; x_0] = \frac{1}{M} \left[ \sum_\omega \frac{1}{N} \sum_{n=1}^{N} \| Y_n - Y^*_n \|^2 + \beta_1 \| Y_N - g(X_N) \|^2 + \beta_2 \| Z_N - \nabla g(X_N) \|^2 \right]
\]

(6.29)

where \( M \) serves as the batch size of the training and \( \omega \) denotes any instance of sampling of the discretized Brownian motion \( W_n, 0 \leq n \leq N - 1 \), and \( \beta_1, \beta_2 \) are the penalty parameters for the terminal conditions. The averaging factor \( 1/N \) is introduced for consistency consideration as the reduction of the loss function as \( N \) increases, when applied to the exact solution, is expected.

**Remark 6.2.** The FBSNNs algorithm proposed in [63] relies on a loss function involving the difference between sequences \( \{ Y_n \} \) and \( \{ Y^*_n \} \), which carry the information inside the time interval \((0, T)\). While the discrete stochastic process \( \{ Y_n \} \) can be expected to approach a continuous stochastic process as defined in the backward SDE (6.4), the question whether the discrete sequence of random variables \( \{ Y^*_n \} \) will converge to the same stochastic process is not clear. As a result, the rate and extent for the difference between \( \{ Y_n \} \) and \( \{ Y^*_n \} \), thus the loss function, approaching to zero is not certain. Our numerical test will provide some evidence for this concern.

### 6.3.2. FBSDE based deep neural network algorithms for PDEs

In this section, we propose improved algorithms for the FBSDEs based deep neural networks similar to the approach in [63], but are mathematically consistent in the definition of the loss function and the discretization of both forward and backward SDEs related to the PDE solutions. Specifically, the loss will be made of the difference of two discrete stochastic processes, which will approach the same process given by the backward SDEs if the overall scheme converges.
6.3.2.1. FBSDE based algorithms - Scheme 2

Based on Remark 6.2 from Section 6.3.1.2, we would like to design a new scheme whose loss function is expected to show the strong convergence rate of the Euler–Maruyama scheme for the discretization of the FBSDEs. A key factor will be to make the loss function as the pathwise differences between two stochastic processes, which will converge to the same continuous adapted diffusion process if the time discretization of FBSDEs and DNN approximations converge.

**Scheme 2.** Train a DNN $u_\theta(t, x)$ to approximate the solution $u(t, x)$ of the parabolic PDE (6.1).

1. Given $X_0 = x_0$ and let $Y_0 = u_\theta(t_0, X_0)$, $Z_0 = \nabla u_\theta(t_0, X_0)$.

2. On each time interval $[t_n, t_{n+1}]$, calculate $X_{n+1}$ and $Y_{n+1}$ using the Euler–Maruyama scheme (6.18) and (6.19), respectively, and calculate $Z_{n+1}$ using the network, i.e.

$$
X_{n+1} = X_n + \mu(t_n, X_n, Y_n, Z_n) \Delta t_n + \sigma(t_n, X_n, Y_n) \Delta W_n,
$$

$$
Y_{n+1} = Y_n + \phi(t_n, X_n, Y_n, Z_n) \Delta t_n + Z_n^T \sigma(t_n, X_n, Y_n) \Delta W_n,
$$

$$
Z_{n+1} = \nabla u_\theta(t_{n+1}, X_{n+1}).
$$

Next, calculate a reference quantity by the DNN representation of the PDE solution (6.20),

$$
Y^*_{n+1} = u_\theta(t_{n+1}, X_{n+1}).
$$

3. For a batch size $M$ with $\omega$ denoting any of the $M$ sample paths, the loss function is given as

$$
L_2[u_\theta; x_0] = \frac{1}{M} \sum_\omega \left[ \frac{1}{N} \sum_{n=1}^N \|Y_n - Y^*_n\|^2 + \beta_1 \|Y^*_N - g(X_N)\|^2 + \beta_2 \|Z_N - \nabla g(X_N)\|^2 \right],
$$

(6.32)
where $\beta_1, \beta_2$ are the penalty parameters of the terminal condition.

The reference quantity $Y_N^\star$ is used in the terminal term in the loss function $L_2[u_\theta; x_0]$, because here it is a straightforward output of the neural network $u_\theta$.

6.3.2.2. FBSDE based algorithms - Scheme 3

In the Scheme 2 above, the discrete process (6.31) is defined through the composite function using the DNN representation of the PDE solution $u_\theta(t, x)$. An alternative way is given below where both discrete processes are obtained from an Euler–Maruyama discretization of the SDEs.

Scheme 3. Train a DNN $u_\theta(t, x)$ to approximate the solution $u(t, x)$ of the parabolic PDE (6.1).

1. Given the initial values $X_{0}^{(1)} = X_{0}^{(2)} = x_0$ and we compute

$$Y_{0}^{(1)} = Y_{0}^{(2)} = u_\theta(t_0, x_0), \quad Z_{0}^{(1)} = Z_{0}^{(2)} = \nabla u_\theta(t_0, x_0)$$

from the network $u_\theta(t, x)$.

2. On each time interval $[t_n, t_{n+1}]$, calculate $X_{n+1}^{(1)}, Y_{n+1}^{(1)}$ and $Z_{n+1}^{(1)}$ as in (6.26) of Scheme 1, then $X_{n+1}^{(2)}, Y_{n+1}^{(2)}$ and $Z_{n+1}^{(2)}$ as in (6.30) of Scheme 2, i.e.

$$X_{n+1}^{(1)} = X_{n}^{(1)} + \mu(t_n, X_{n}^{(1)}, Y_{n}^{(1)}, Z_{n}^{(1)})\Delta t_n + \sigma(t_n, X_{n}^{(1)}, Y_{n}^{(1)})\Delta W_n,$$

$$Y_{n+1}^{(1)} = u_\theta(t_{n+1}, X_{n+1}^{(1)}),$$

$$Z_{n+1}^{(1)} = \nabla u_\theta(t_{n+1}, X_{n+1}^{(1)}),$$

$$X_{n+1}^{(2)} = X_{n}^{(2)} + \mu(t_n, X_{n}^{(2)}, Y_{n}^{(2)}, Z_{n}^{(2)})\Delta t_n + \sigma(t_n, X_{n}^{(1)}, Y_{n}^{(1)})\Delta W_n,$$

$$Y_{n+1}^{(2)} = Y_{n}^{(2)} + \phi(t_n, X_{n}^{(2)}, Y_{n}^{(2)}, Z_{n}^{(2)})\Delta t_n + (Z_{n}^{(2)})^T \sigma(t_n, X_{n}^{(2)}, Y_{n}^{(2)})\Delta W_n,$$

$$Z_{n+1}^{(2)} = \nabla u_\theta(t_{n+1}, X_{n+1}^{(2)}).$$

3. For a batch size $M$ with $\omega$ denoting any of the $M$ sample paths, the loss function is
defined by

\[
L_3[u_\theta; x_0] = \frac{1}{M} \sum_\omega \left[ \frac{1}{N} \sum_{n=1}^N \left\| Y_n^{(1)} - Y_n^{(2)} \right\|^2 + \beta_1 \left\| Y_N^{(1)} - g(X_N^{(1)}) \right\|^2 + \beta_2 \left\| Z_N^{(1)} - \nabla g(X_N^{(1)}) \right\|^2 \right],
\]

(6.36)

where \( \beta_1, \beta_2 \) are the penalty parameters of the terminal condition.

6.4. Application in committor functions of the transition path theory

In the transition path theory [28, 29, 53], the transition between two states is described by a high dimensional stochastic process \( X_t \in \Omega \subset \mathbb{R}^d \), satisfying

\[
dX_t = -\nabla U(X_t) dt + \sigma dW_t,
\]

(6.37)

where \( U : \mathbb{R}^d \rightarrow \mathbb{R} \) is a smooth potential function, \( \sigma = \sqrt{2\beta^{-1}}, \beta > 0 \) is the inverse of temperature, and \( W_t \) is a \( d \)-dimensional stochastic process whose entries are standard Brownian motions. For two states \( A \) and \( B \) corresponding to simply connected domains with smooth boundaries in the state space, the committor function \( q(x) \) is the key to many statistical properties of the transition, defined as

\[
q(x) = \mathbb{P} [\tau_B < \tau_A | X_0 = x],
\]

(6.38)

describing the probability of hitting \( B \) prior to \( A \) with the stochastic process in (6.37) starting at \( x \), where \( \tau_D \) refers to the first hitting time of any domain \( D \). The committor function satisfies the Fokker–Planck equation

\[
\mathcal{L}[q] = \frac{1}{2} \text{Tr}[\sigma^2 \nabla^2 q] - \nabla U \cdot \nabla q = 0, \quad q|_{\partial A} = 0, \quad q|_{\partial B} = 1,
\]

(6.39)
which can be treated as a time-independent version of the parabolic PDE (6.1)

\[ \partial_t q + L[q] = 0, \quad \partial_t q = 0. \] (6.40)

When the potential \( U(x) \) is confining, such that the stationary distribution

\[ \mu(x) = \frac{1}{Z} e^{-\beta U(x)}, \quad Z = \int_{\Omega \setminus (A \cup B)} e^{-\beta U(x)} dx \]

exists and the stochastic process is ergodic [62], the Monte Carlo approach can be applied to compute the committor function for given initial value \( x \), as

\[ q(x) = \mathbb{E} \left[ X_{\tau_{A \cup B}} \middle| X_0 = x \right]. \] (6.41)

However, one can also extend the Pardoux–Peng theory to represent the solution to (6.40). The detailed discussion can be found in [24]. The DNN approach will allow us to access the approximated solution in a region rather than from one chosen point as the initial value. A numerical example is given in the following section.

The FBSDE based DNN algorithm is modified to remove the time dependency, but still aims at solving (6.40) in a time-dependent form. The scheme is shown as follows.

**Scheme 4.** Train a DNN \( u_\theta(x) \) to approximate the time-independent solution to (6.40).

1. Pick a terminal time \( T > 0 \) and number of intervals \( N \) for the “time” discretization on \([0, T] \). We will then follow the notations used above in Scheme 1–Scheme 3 for the time discretization.

2. Given the initial value \( X_0 = x_0 \), compute \( Y_0 = Y_0^* = u_\theta(x_0) \), and \( Z_0 = \nabla u_\theta(x_0) \).

   Initialize the loss function for each sample path \( \omega \) as \( L(\omega) = 0 \).

3. On each time interval \([t_n, t_{n+1}] \), compute

\[ X_{n+1} = X_n - \nabla U(X_n) \Delta t_n + \sigma \Delta W_n. \] (6.42)
If $X_{n+1} \not\in \Omega \setminus (A \cup B)$, then the sampling ends, and we compute the terminal loss

\[ L(\omega) \leftarrow L(\omega) + \beta_t 1_{X_{n+1} \in A} \|Y_n^*\|^2 + \beta_t 1_{X_{n+1} \in B} \|Y_n^* - 1\|^2, \tag{6.43} \]

where $\beta_t$ is a weighting factor. Otherwise, we continue on the sampling by computing

\[ Y_{n+1} = Y_n + \sigma Z_n \cdot \Delta W_n, \quad Z_{n+1} = \nabla u_\theta(X_{n+1}) \tag{6.44} \]

and a reference quantity

\[ Y_{n+1}^* = u_\theta(X_{n+1}). \tag{6.45} \]

The sample path loss is accumulated

\[ L(\omega) \leftarrow L(\omega) + \frac{1}{N} \|Y_{n+1} - Y_{n+1}^*\|^2. \tag{6.46} \]

4. The total loss function

\[ L_2[u_\theta; x_0] = \frac{1}{M} \sum_\omega L(\omega) \tag{6.47} \]

is a Monte Carlo approximation of $\mathbb{E}[L(\omega)]$.

6.5. Numerical results

In this section, we will carry out several tests on Scheme 1 from [63] and the new Scheme 2 and Scheme 3, for a 100-dimensional Black–Scholes–Barenblatt equation and its variants.

6.5.1. 100-dimensional Black–Scholes–Barenblatt equation

Consider the following 100-dimensional Black–Scholes–Barenblatt (BSB) equation from [63] as the model problem: for $t \in [0, T]$ and $x \in \mathbb{R}^d$, the scalar function $u(t, x)$ satisfies

\[ u_t + \frac{1}{2} \text{Tr} [\sigma^2 \text{diag}(xx^T) \nabla \nabla u] = r(u - \nabla u \cdot x), \tag{6.48} \]

\[ u(T, x) = \|x\|^2. \]
The PDE is linked to the FBSDEs
\[
\begin{align*}
    dX_t &= \sigma \text{diag}(X_t) dW_t, \\
    X_0 &= x_0, \\
    dY_t &= r(Y_t - Z_t \cdot X_t) dt + \sigma Z_t^{T} \text{diag}(X_t) dW_t, \\
    Y_T &= g(X_T),
\end{align*}
\] (6.49)

where \( g(x) = \|x\|^2 \), and \( x_0 \in \mathbb{R}^d \) is the position where we like to get the initial value \( u(0, x_0) \). The exact solution to the PDE (6.48) is given in a closed form by
\[
    u(t, x) = e^{(r+\sigma^2)(T-t)} \|x\|^2,
\] (6.50)

so that we can test the accuracy of the DNN schemes. Parameters are given by \( d = 100, T = 1.0, \sigma = 0.4, r = 0.05 \) and
\[
    x_0 = (1, 0.5, 1, 0.5, \cdots, 1, 0.5).
\] (6.51)

We use a 6-layer fully connected feedforward neural network for \( u_\theta(t, x) \) with 5 hidden layers, each having 256 neurons. The activation function is the sine function as suggested by [63]. We train the network with the Adam optimizer with descending learning rates 1e-3, 1e-4, 1e-5, 1e-6 and 1e-7, each for 10000 steps. The batch size is \( M = 100 \).

In the loss functions (6.29), (6.32) and (6.36), the penalty parameters are chosen as \( \beta_1 = \beta_2 = 0.02 \).

Illustration of the training results in the high-dimensional space is provided along the sample paths. When the training is finished, we randomly generate 1000 sample paths for verification of the accuracy, with a finer time discretization with time steps \( \Delta t_n = 1/1000 \). For each (discretized) sample path \( \omega \) and for \( 0 \leq n \leq 1000 \), the relative error of this model
Figure 6.1: (Non-convergence) relative error of Scheme 1 for $N = 12$ (middle), 48 (bottom) and 192 (top).

Problem at $(t_n, X_n(\omega))$ (or at $(t_n, X_n^{(2)}(\omega))$ when using Scheme 3) is defined by

$$e_n(\omega) = \frac{|u_\theta(t_n, X_n(\omega)) - u(t_n, X_n(\omega))|}{|u(t_n, X_n(\omega))|}. \quad (6.52)$$

The mean and the standard deviation (SD) of each $e_n$ can also be calculated.

6.5.1.1. Scheme 1 from FBSNNs

Figure 6.1 shows the relative error of Scheme 1 [63] for $N = 12$, 48 and 192, where the mean error and the mean error plus two standard deviations of the error are presented. We can see the reduction of the errors from $N = 12$ to $N = 48$, however, the error increases from $N = 48$ to $N = 192$. This degeneracy in accuracy is an indication that as the time discretization is refined, the two quantities in the definition of loss function (6.28) do not approach the same continuous stochastic process. In fact, as defined by (6.27), $\{Y_n^*\}$ may not converge to a continuous stochastic process at all.
Figure 6.2: Relative error of Scheme 2 for $N = 12, 48, 192$ and $768$.

(a) Error Mean  
(b) Error Mean plus two SDs

Figure 6.3: Relative error of Scheme 3 for $N = 12, 48, 192$ and $768$.

(a) Error Mean  
(b) Error Mean plus two SDs

6.5.1.2. Scheme 2 and Scheme 3

Figure 6.2 and Figure 6.3 show the mean error and mean error plus two standard derivations of the error for Scheme 2 and Scheme 3 for $N = 12$, $N = 48$, $N = 192$ and $N = 768$, respectively. Both results in Figure 6.2 and Fig. 6.3 show the convergence of the new Scheme 2 and Scheme 3, respectively, in contrast to the degeneracy of the accuracy of Scheme 1 when the time discretization is refined. For both new schemes, we can see improvement of the accuracy from $N = 48$ to $N = 192$ is close to the one from $N = 12$ to $N = 48$, but the improvement of $N = 768$ over $N = 192$ is a little less. This indicates the network training might dominate the error compared to the time discretization error. In fact, the terminal
Figure 6.4: Prediction of 8 test sample paths from training results of Scheme 2 and Scheme 3, \( N = 192 \).

parts of the loss function failed to halve in the \( N = 768 \) cases compared to \( N = 192 \).

Figure 6.4 (a) (b) show the prediction of trained networks using Scheme 2 and Scheme 3 with \( N = 192 \) along 8 sampled test paths depicted in Figure 6.4 (c), in comparison with the exact solution, where the average error of the prediction is given in Figure 6.4 (d).

6.5.1.3. Extrapolation for higher order accuracy in the point solution

In Section 6.5.1.2 we have seen that Scheme 2 and Scheme 3 have the convergence behavior as the Euler–Maruyama scheme, so we can assume that the truncation error may have the
following asymptotic ansatz

\[ u^n_\theta - u = C_1 N^{-\frac{1}{2}} + C_2 N^{-1} + O(N^{-\frac{3}{2}}), \]  

(6.53)

where the leading term \( C_1 N^{-\frac{1}{2}} \) dominates the error when \( N \) is sufficiently large. If this holds for both \( u^n_\theta \) and \( u^{4N}_\theta \) for some constants \( C_1 \) and \( C_2 \), then we can define an extrapolated solution

\[
u^{\text{ex}}_{4N} = 2u^{4N}_\theta - u^n_\theta = u - \frac{C_2}{2} N^{-1} + O(N^{-\frac{3}{2}}) \]

(6.54)

as an improved approximation to the solution.

For the model problem (6.48), the extrapolation is valid for the approximation of \( Y_0 = u(0, x_0) \) and \( u(0, x) \) in a neighborhood near \( x_0 \), as shown by Table 6.1 and Figure 6.5. In terms of the accuracy of \( Y_0 \), by training the DNNs only with \( N = 12 \) and \( N = 48 \), the extrapolated result \( u^{48}_{\text{ex}}(0, x_0) \) has its accuracy outperforming those using \( N = 768 \) which takes more than 10 times longer time to train, when using both Scheme 2 and Scheme 3. Due to training difficulties, the improvement for using extrapolation on \( N = 768 \) is marginal, but still exists.

Note that the extrapolation approach usually may not work for the whole time interval

\[
\begin{array}{|c|c|c|c|c|}
\hline
N & u^n_\theta & u^{\text{ex}}_\theta & u^{4N}_\theta & u^{4N}_{\text{ex}} \\
\hline
12 & 2.91e-03 & 2.82e-03 & 2.82e-03 & 2.82e-03 \\
48 & 1.67e-03 & 4.29e-04 & 1.13e-03 & 5.57e-04 \\
192 & 7.58e-04 & 8.43e-04 & 5.55e-04 & 5.55e-04 \\
\hline
\end{array}
\]

Table 6.1: Relative error of \( Y_0 \) from the network approximation and extrapolation.
along the entire sample paths. For instance, the values at $t = T$ are subject to explicit fitting of the terminal condition from the loss functions (6.32) and (6.36), so we cannot expect a general constant $C_1$ in (6.53) for $u^N_\theta(T, x)$ and $u^{4N}_\theta(T, x)$. The result in Figure 6.5 shows that the extrapolation technique can be used for a time interval $0 \leq t \leq 0.1$.

6.5.1.4. Region of validity of the DNN in a neighborhood of the initial value

In this section, we will verify the validity of the networks $u_\theta(t, x)$ in a region that are larger than the one sampled during the training process. For this purpose, we randomly sample the initial value $X_0 = \tilde{x}_0$ from a cubic neighborhood of $x_0$ with halved edge length $R$, i.e.,

$$(\tilde{x}_0)_j = (x_0)_j \cdot (1 + \varepsilon_j), \quad 1 \leq j \leq d = 100,$$

(6.55)

where $\varepsilon_j$ are i.i.d. random variables with uniform distribution on $(-R, R)$. For the network trained with Scheme 2 and $N = 192$, we compare the resulting error using the same measurement with $R = 0.25$ and $R = 0.5$, while keeping one sample starting exactly from $x_0$ (for the sake of plotting), see Figure 6.6. The averaged relative error is slightly larger at $t = 0$ because during the training process these regions are less likely to be visited since we fixed the initial value for all training pathes at $X_0 = x_0$. If we look at the overall maximum for $t \in [0, T]$, we can still have an averaged relative error of 0.34% for $R = 0.25$ and 1.25% for
Figure 6.6: Training error verified with initial value $\bar{x}_0$ from a neighborhood of $x_0$, using Scheme 2, $N = 192$.

Also, it is noted that, in comparison with the non-perturbed result, the trained network fits the solution of the PDE better when $Y_t$ has a value below 80.

This result shows that the DNN we trained for $x = x_0$ is in fact can be used in a local neighbourhood around $x_0$ for the whole time interval $0 \leq t \leq T$.

6.5.2. MultiscaleDNN for the BSB equation with temporal oscillations

In a recent work [52], a multi-scale DNN was proposed, which consists of a series of parallel normal sub-networks, each of which receiving a scaled version of the input, and outputs of the sub-networks are combined to form the final output of the MscaleDNN (see
Fig. 6.7). The individual sub-networks in the MscaleDNN with a scaled input is designed to approximate a segment of frequency content of the targeted function, and the effect of the scaling is to convert a specific high frequency content to a lower frequency range so that the learning can be accomplished more quickly, which is shown by the recent work [52] on the frequency dependence of the DNN convergence.

Figure 6.7 shows the schematics of a MscaleDNN consisting of \( n \) sub-networks. Each scaled input passing through a fully-connected sub-network, which can be expressed in the formula (6.22), here again we use the sine function for the activation function, i.e.,

\[
\sigma(x) = \sin(x). \quad (6.56)
\]

Mathematically, the final output of a MscaleDNN solution is represented by the following sum of sub-networks \( f_{\theta^i} \) with network parameters denoted by \( \theta^i \) (i.e. weight matrices and bias)

\[
f(x) \sim \sum_{i=1}^{M} W^{[L]}_i f_{\theta^i}(\alpha_i \cdot x) + b^{[L]},
\]
where \( \alpha_i \) is the chosen scale vector for the \( i \)-th sub-network in Figure 6.7. For more details on the design of the MscaleDNN, refer to [52].

For the input scales, the general idea is to adopt various scaling factors for different components of the input, depending on the complexity of the PDE to be solved.

The MscaleDNN is tested with the following model problem, modified from the BSB equation above with an oscillatory factor to effectively increase the training difficulty:

\[
\frac{\partial_t u}{2} + \text{Tr}[\sigma^2 \nabla \nabla u] = \phi,
\]

\[
u(T, x) = g(x),
\]

where the dimension \( d = 100 \), \( T = 1.0 \), \( \sigma = 0.4 \) and \( r = 0.05 \) are unchanged parameters compared to (6.48),

\[
g(x) = \|x\|^2 \left(1 + \alpha \sin (\beta S_1 - \gamma T)\right),
\]

\[210\]
Figure 6.7: Illustration of a MsceLDNN [52].

\[
\phi(t, x, u, \nabla u) = r(u - \nabla u \cdot x) + \alpha e^{(r+\sigma^2)(T-t)} P(t, x), \tag{6.60}
\]

\[
P(t, x) = (r\beta S_1 S_2 - \gamma S_2 + 2\sigma^2 \beta S_3) \cos(\beta S_1 - \gamma t) - \frac{\sigma^2 \beta^2}{2} S_2^2 \sin(\beta S_1 - \gamma t), \tag{6.61}
\]

where each \( S_j = \sum_{i=1}^{d} x_i^j \), and \( \alpha, \beta \) and \( \gamma \) are parameters to be tuned. The modified PDE (6.58) has solution in closed form

\[
u(t, x) = e^{(r+\sigma^2)(T-t)} \| x \|^2 (1 + \alpha \sin(\beta S_1 - \gamma t)), \tag{6.62}
\]

and corresponds to the FBSDEs

\[
dX_t = \sigma \text{diag}(X_t) dW_t,
\]

\[
X_0 = x_0,
\]

\[
dY_t = \left( r(Y_t - Z_t \cdot X_t) + \alpha e^{(r+\sigma^2)(T-t)} P(t, X_t) \right) dt + \sigma Z_t^T \text{diag}(X_t) dW_t,
\]

\[
Y_T = g(X_T).
\]

We apply \( \alpha = 0.025, \beta = 0.25 \) and \( \gamma = 32 \) to the above equation. During the training
process, we use the same settings for the fully-connected DNN as in previous tests. For the MscaleDNN, the network is divided into 4 sub-networks, each having 5 hidden layers with 64 neurons per layer, so that sizes of the networks in the comparison are matching. The scaled inputs for the sub-networks are given by

\[(3^0 t, x), (3^1 t, x), (3^2 t, x), (3^3 t, x),\]  \hspace{1cm} (6.64)

so that a wider range of frequency of \( t \) can be captured with the MscaleDNN. When applying Scheme 2 and \( N = 48 \), the MscaleDNN halves the overall error compared to the fully-connected network.

One can also predict sample paths with better accuracy using the MscaleDNN, too, see Figure 6.9.

6.5.3. 100-dimensional committor function for a double-well potential

We refer to the following double-well potential problem from [85] to test the DNN algorithm proposed for evaluating the committor function. In this model problem, we set the
Figure 6.9: Prediction of 8 sample paths for problem with oscillation (6.58), using the MscaleDNN with Scheme 2 and $N = 192$.

Dimension $d = 100$, $\sigma = 1$ and

$$U(x) = (x_1^2 - 1)^2 + 0.3 \sum_{i=2}^{d} x_i^2,$$

(6.65)

with regions

$$A = \{ x \in \mathbb{R}^d | x_1 \leq -1 \}, \quad B = \{ x \in \mathbb{R}^d | x_1 \geq 1 \}.$$

(6.66)

The exact solution is given by

$$q(x) = \frac{\int_{-1}^{x_1} e^{2(y^2-1)^2} dy}{\int_{-1}^{1} e^{2(y^2-1)^2} dy}, \quad -1 < x_1 < 1,$$

(6.67)

which can be used to test the accuracy of Scheme 4 for the committor function.

In our numerical approximation, we set $T = 2$, $N = 200$ and $\beta_t = 0.5$. From the initial value

$$x_0 = (0.37, 0.37, \cdots, 0.37),$$

(6.68)

a fully-connected DNN with 5 hidden layers, each with 256 neurons is trained. The activation function is chosen as Tanh. The Adam optimizer is applied with descending learning rates
Figure 6.10: Training result of Scheme 4 from Section 6.4 for the committor function verified on a list of points (6.69) for the model problem in Section 6.5.3.

1e-4, 1e-5, 1e-6 and 1e-7, each for 5000 steps. The batch size $M = 64$. We pick a list of points on a straight line

$$(w_n, 0.37, 0.37, \cdots, 0.37) \in \mathbb{R}^{100}, \quad w_n = -1 + 0.05n, \quad 0 \leq n \leq 40 \quad (6.69)$$

to validate the training result. Figure 6.10 shows that the training result starting from one initial point in the high-dimensional space well represents the targeting committor function on one dimension. The absolute error for the committor function on this line as a quantity of probabilities is bounded by 0.014.

6.6. Conclusion

In this chapter, we have proposed two FBSDE based DNN algorithms for high dimensional quasilinear parabolic equations. The key component of the proposed algorithms is the loss function used, consisting of, in addition to the terminal condition of the PDE, the path-wise difference of two convergent stochastic processes from either discretized SDEs or the PDEs network solution. As the two stochastic processes converge to the same stochastic processes in the Pardoux–Peng theory, the new algorithms are able to demonstrate closely the half-order strong convergence of the underlying Euler–Maruyama scheme for the SDEs.
We also show that the extrapolation method verifies the convergence order of the DNN solutions and further enhances the resulting accuracy of the estimate on the initial value of the PDE. For PDEs with time oscillatory solutions, we demonstrated that the MscaleDNN is shown to provide an enhancement of the resulting accuracy. We also proposed an approach based on the same framework to compute the committor function which satisfies a time-independent equation.

Future research will be done to improve the convergence of the networks and the overall schemes, including MscaleDNN for PDEs with spatially oscillatory solutions.
In this chapter we give concluding remarks for this thesis, and briefly discuss topics for the future work.

7.1. Overall contributions

In this thesis, we have developed the fast multipole method for wave and change source interactions in layered media regarding various types of partial differential equations, and deep neural network based algorithms equipped with the Pardoux–Peng theory solving high-dimensional quasi-linear parabolic partial differential equations.

In Chapter 2 we presented theories of the layered media Green’s functions for the 2-D Helmholtz equation, including their representation and asymptotic behavior in the frequency domain. The separation of field propagation directions in the frequency domain produces the reaction field decomposition. For each reaction field component, we proposed the far-field expansions including the multipole expansion, the local expansion, and the translation operators between these expansions such as the multipole-to-local translation. The exponential convergence with respect to the polarization distance in layered media has been discovered and theoretically proven. A framework for the fast multipole method has been established following the convergence estimates.

In Chapter 3, we have developed the fast multipole method for the Helmholtz equation and the linearized Poisson–Boltzmann equation in 3-D layered media. Following the framework suggested by the 2-D work, we present the generalized Funk–Hecke formula for the target-source separation in the frequency domain, then proposed the far-field expansions for
the reaction field components and established the framework for the fast multipole method. A few techniques for practical implementation to reduce computational cost and to improve accuracy have been discussed. Numerical tests verify both the accuracy and the efficiency of the proposed algorithm.

In Chapter 4, we have developed the fast multipole method for the Laplace’s equation in 3-D layered media. In addition to a series of results similar to those of the Helmholtz equations and the linearized Poisson–Boltzmann equations, we also presented the proof of exponential convergence for the far-field expansions.

In Chapter 5, we have designed a matrix basis formulation to efficiently represent the dyadic Green’s functions for the Maxwell’s equations and the elastic wave equation in layered media, which are in the shape of $3 \times 3$ tensors. The matrix basis formulation separates the rotational dependence from the tensor Green’s functions in the frequency domain, while the coefficients benefit from the formulation so that evaluation is much simplified. Regarding various phases of contacting media for the elastic wave equation, mixed forms for compression waves propagation have been discussed, together with a simplified vector basis formulation representing the vector Green’s function in layered media when source particles locate in a fluid layer. A framework for the fast multipole method for the Maxwell’s equations and the elastic wave equation is discussed under the matrix basis formulation.

In Chapter 6, we have proposed deep neural network algorithms based on the forward-backward stochastic differential equations to solve high-dimensional quasi-linear parabolic partial differential equations. The algorithms train a deep neural network that represent the solution to the partial differential equation by comparing two converging stochastic processes as the time discretization gets finer. Numerical tests verify the order of convergence of the algorithms as the one in the Euler–Maruyama scheme for time discretization. An improvement using extrapolation is achieved to acquire better accuracy for the evaluation of the result at the given initial value, while the algorithm also successfully predicts in a neighborhood of the initial value in the high-dimensional space. The MscaleDNN is applied
to enhance the accuracy when the differential equation has a time oscillatory solution.

Below is a list of published journal papers and preprints related to this thesis.


7.2. Future work

There are several worthwhile and promising research directions in extending this thesis work.

Based on the series of work about the fast multipole method in layered media, the corresponding fast algorithms for the Maxwell’s equations and for the elastic wave equation in layered media are to be numerically implemented, and to be applied to boost integral equation solvers in various simulation problems. For wave interactions in layered media with high frequency, efficient approaches of the fast multipole method are to be discovered. On the theory side, an exponential convergence estimate of the numerically validated far-field expansions for the Helmholtz equation and the linearized Poisson–Boltzmann equation in 3-D layered media is a potential work. The corresponding estimates for the Maxwell’s equations and for the elastic wave equations will be a direct corollary.

Based on the work regarding the deep neural network solver for high-dimensional differential equations, a posteriori error estimate is yet to be given for the current algorithms. The same methodology from the work of this thesis can be applied to the Fokker–Planck equation as well. Another interesting topic is to apply general Lévy process to solve fractional partial differential equations under a similar framework.
BIBLIOGRAPHY


222


[48] **Kong, J. A.** Electromagnetic Fields Due To Dipole Antennas Over Stratified Anisotropic Media††Manuscript received by the Editor March 10, 1972; revised manuscript July 12, 1972. *Geophysics* 37, 6 (12 1972), 985–996.


