Analysis and extensions of the Multi-Layer Born method

May 3, 2024

Abstract

This work is concerned with the three-dimensional numerical simulation of scalar wave propagation. We work in a setting where backscattering can be neglected, which then allows us to only calculate forward propagation. Our main goal is to propose solutions to a critical issue in the recently introduced Multi-Layer Born method, issue due to the discarding of evanescent modes. We derive two possible schemes, and perform a rigorous mathematical analysis of the numerical errors to show one method is more accurate. This analysis is also helpful for choosing optimally the discretization parameters. In addition, we propose high order versions of the Multi-Layer Born method that offer a lower computational cost for a given tolerance.

1 Introduction

Simulating wave propagation in strongly heterogeneous media over large distances compared to the wavelength of the electromagnetic field is an extremely challenging task. In the scalar case, this amounts to numerically solving the three-dimensional Helmholtz equation, for which a very fine grid is necessary in order to achieve a reasonable accuracy. This results in large linear systems whose resolution comes at a considerable computational cost. In the context of imaging and inverse problems, where multiple systems have to be solved to, e.g., reconstruct the refractive index or image scatterers, a full simulation of wave propagation is often prohibitive and one has to resort to approximations. A classical simplification stems from the directional nature of the propagating wavefield, as in the case of optical imaging, allowing for neglecting backscattering and taking into account only forward propagation. This is how the so-called paraxial wave equation is derived, and is simply a time-dependent two-dimensional Schrödinger equation where the time variable is replaced by the variable along the direction of propagation. The full 3D problem is then reduced to a 2D evolution problem that is now numerically tractable. The main issue with such a model is the loss of accuracy away from the axis of propagation, and several approaches have been developped to resolve this shortcoming. One of those is the Multi-Layer Born (MLB) method introduced in [1]. The main idea is to use an integral formulation of the Helmholtz equation in which backscattering is neglected, and to slice the heterogeneous medium along the axis of propagation into layers with a width small compared to either the wavelength or the typical oscillation length of the background, which ever is shorter. This condition on the layers' width is necessary to achieve a decent accuracy. This results in a very simple numerical scheme whose computational cost is inversely proportional to the layers' width, and which consists in evaluating iteratively 2D convolutions of the form, for $z > 0$,

$$f(x) = \int_{\mathbb{R}^2} G(x-y,z)U(y)dy, \qquad x \in \mathbb{R}^2, \tag{1}$$

where G is the Green's function and z the direction of propagation, (below $\hat{G}$ denotes the partial Fourier transform of G with respect to the transverse variable $x = (x_1, x_2) \in \mathbb{R}^2$ with $|x| = \sqrt{x_1^2 + x_2^2}$),

$$G(x,z) = \frac{e^{ik_0\sqrt{|x|^2+z^2}}}{4\pi\sqrt{|x|^2+z^2}}, \qquad \hat{G}(\xi,z) = \frac{i}{8\pi^2}\frac{e^{i\sqrt{k_0^2-|\xi|^2}|z|}}{\sqrt{k_0^2-|\xi|^2}}, \tag{2}$$

and where U is the product of the local field in a given layer and the perturbation of the refractive index due to the heterogeneous medium (U is denoted by VU further, for V the perturbation and U the field). In [1], f is computed using the fast Fourier transform (FFT) algorithm based on the following form of f in the Fourier space:

$$f(x) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} e^{i\xi\cdot x}\hat{G}(\xi)\hat{U}(\xi)d\xi. \tag{3}$$

This requires the FFT of U, and then an inverse FFT to get f.

The problem is actually not as simple as it looks since $\hat{G}$ is singular at $|\xi| = k_0$, and it is well-known that Fourier-based techniques do not offer high accuracy with singular functions [3]. The original MLB method avoids the issue of the singularity by introducing a cut-off preventing $|k|$ to get too close to k_0. From a physical perspective, this means that only propagating modes are calculated and that the evanescent modes are all discarded. While ignoring the evanescent field would lead to an accurate approximation in a classical far-field scattering problem where propagation occurs in free space, here this unfortunately introduces large errors because the field is always in a near-field regime. Indeed, since the width of the layer has to be small compared to the wavelength for an accurate approximation of the integral equation, scattering is in a near-field regime along z and evanescent modes cannot be neglected. This statement has to be nuanced when the beam enters the scattering medium since it is still well focused, but as it propagates, the beam widens and the transfer from propagating to evanescent modes increases, which in turn enhances backconversion from evanescent to propagating modes. We show theoretically in this work, see Appendix E, that, even in a transverse far-field regime where the norm of x is large compared to the wavelength and z is of the order of the wavelength, the contribution of the evanescent wave is as large as that of the propagating wave provided $\hat{U}(\xi)$ has a non-zero contribution around $|\xi| = k_0$. This is not surprising since, again, we are in a near-field regime along the axis of propagation.

This previous discussion suggests that evanescent modes have to be included in the simulation for accurate results, which in turn requires careful handling of the calculation of f. We propose in this article two methods, both based on the FFT for minimal computational cost. One consists in working with the real-space representation of f

and by computing the discrete Fourier transforms (DFT) of G and U. The resulting DFT of G is then an effective regularization of the exact $\hat{G}$ that includes the evanescent modes. Another natural approach consists in "fixing" the original MLB formulation based on (3) by regularizing directly $\hat{G}$ and considering a complex-valued wavenumber. We perform a rigorous error analysis that shows that the first method has a better accuracy than the second one for a comparable computational cost. Another benefit of the error analysis is to provide us with an optimal way to choose the discretization parameters.

This strategy solves the evanescent wave problem. A remaining issue is the large computational cost when the layers' width is small as more iterations along z have to be performed. We then propose higher order schemes meant to alleviate the cost based on classical high order quadrature formulas for numerical integration. We derive a second order method exploiting the midpoint rule, and a fourth order method based on Milne quadrature.

The paper is structured as follows: we define the setting and our two methods for the calculation of (1) in Section 2 and state our error estimates. We present the high order methods in Section 3. Our technical results are given in Appendix: in Appendix A, we recall how to derive the MLB scheme from the Helmholtz equation; we state our two convergence theorems in Appendix B, and detail their proofs in Appendices C and D. We finally substantiate in Appendix E the claim made in introduction that propagating and evanescent modes have similar amplitudes in a transverse far-field regime.

Acknowledments. This work is funded by the Chang-Zuckerberg initiative and OP is supported by NSF grant DMS-2006416.

2 Numerical methods

Setting. The starting point is the three-dimensional Helmholtz equation equipped with Sommerfeld radiation conditions,

$$\Delta U + k^2(x,z)U = S, \qquad \text{with} \qquad \partial_r(U - U_\mathrm{i}) = ik_0(U-U_\mathrm{i}) + O(1/r^2) \qquad \text{as} \quad r \to \infty,$$

where $x = (x_1, x_2) \in \mathbb{R}^2$, $r = \sqrt{|x|^2 + z^2}$, and the wavenumber reads $k^2(x,z) = (\omega/c)^2 n(x,z)^2$, for ω the angular frequency, c the speed of light, and n the medium refractive index. The incoming field U_i satisfies

$$\Delta U_\mathrm{i} + k_0^2 U_\mathrm{i} = S.$$

The constant background index is n_0 with $k_0 = (\omega/c)n_0$, and let the scattering potential be defined by $V(x,z) = (\omega/c)^2(n(x,z)^2 - n_0^2)$. We suppose that V is supported in a bounded domain Ω. Above, S is a source located outside of Ω. Let the scattered field be defined by $U_\mathrm{s} = U - U_\mathrm{i}$. Calculations given in Appendix A show, when backscattering is neglected, that U_s can be approximated by, when $\Delta z > 0$,

$$\begin{aligned} U_\mathrm{s}(x, z+\Delta z) \;\simeq\; & \int_{\mathbb{R}^2}\int_z^{z+\Delta z} G(x-x', z+\Delta z - z')\,[V(U_\mathrm{i}+U_\mathrm{s})]\,(x',z')dx'dz' \\ & + P_{\Delta z}U_\mathrm{s}(x,z), \end{aligned} \tag{4}$$

where $P_{\Delta z}$ is the angular propagator defined by

$$P_{\Delta z}\varphi(x,z) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} e^{i\sqrt{k_0^2-|\xi|^2}|\Delta z|} e^{i\xi\cdot x}\hat{\varphi}(\xi,z)d\xi. \tag{5}$$

Above, the complex square root has positive imaginary part, and we chose the convention $\hat{\varphi}(\xi) = \int_{\mathbb{R}^2} e^{-i\xi\cdot x}\varphi(x)dx$ for the Fourier transform.

Equation 4 is the starting point of MLB and extensions. In the standard MLB, the integral over z is simply approximated by the method of rectangles, resulting in the explicit scheme

$$U_{\mathrm{s}}(x, z+\Delta z) \simeq \Delta z \int_{\mathbb{R}^2} G(x-x', \Delta z)\left[V(U_{\mathrm{i}}+U_{\mathrm{s}})\right](x', z)dx' + P_{\Delta z}U_{\mathrm{s}}(x,z), \tag{6}$$

which provides us with an iterative formula. The key to the method is then to find an efficient and accurate way to compute the convolution with the Green's function.

We propose two methods for this in the next two sections, and perform a rigorous error analysis.

Method 1. We begin with the f defined in (1), in which U now plays the role of the term $V(U_{\mathrm{i}}+U_{\mathrm{s}})$. For simplicity, we replace Δz by $z>0$ when going from (6) to (1), and denote from now on for z fixed $G(x)\equiv G(x,z)$ since z is a fixed parameter. The length z models the width of the layers, and has to satisfy $z \ll \ell = \min(\lambda, \ell_c)$, where $\lambda = 2\pi/k_0$ is the central wavelength and ℓ_c is the typical oscillation length of $V(U_{\mathrm{i}}+U_{\mathrm{s}})$, for the expression (6) to yield an accurate approximation of the integral in z.

For some length $L>0$, denote by $S_L = [-L/2, L/2]\times[-L/2, L/2]$ the square of side L. We choose L sufficiently large so that the support of U is included in S_L. We need first to periodize f to compute its Fourier series. For this, we restrict f to S_L and extend it periodically to a function denoted f_L. For $x\in S_L$, we have then $f_L(x) = f(x)$, and for $i=1,2$, $f_L(x\pm Le_i) = f_L(x\pm Le_1 \pm Le_2) = f(x)$ where $e_1 = (1,0)$ and $e_2=(0,1)$. Note that since the function G is even, the periodization we chose does not introduce discontinuities at the boundary of S_L and there is no need for a more involved periodization.

The Fourier series of f_L reads

$$f_L(x) = \sum_{m,n\in\mathbb{Z}} \hat{f}_{m,n}e^{ik_{m,n}\cdot x}, \qquad k_{m,n} = 2\pi(m,n)/L,$$

where the Fourier coefficients are

$$\hat{f}_{m,n} = \frac{1}{L^2}\int_{S_L} f(x)e^{-ik_{m,n}\cdot x}dx.$$

Exploiting the convolution in the definition of f yields

$$\hat{f}_{m,n} = L^2\hat{G}_{m,n}\hat{U}_{m,n},$$

for $\hat{G}_{m,n}$ and $\hat{U}_{m,n}$ the Fourier coefficients of G and U, namely

$$\hat{G}_{m,n} = \frac{1}{L^2}\int_{S_L} G(x)e^{-ik_{m,n}\cdot x}dx, \qquad \hat{U}_{m,n} = \frac{1}{L^2}\int_{S_L} U(x)e^{-ik_{m,n}\cdot x}dx.$$

What can be computed efficiently with the fast Fourier transform algorithm are the DFT of G and U, and subsequently an inverse DFT to recover an approximation of f. This is formalized as follows. Let $N \in \mathbb{N}$, chosen to be odd without lack of generality, and denote by I_N the set of integers $I_N = \{-(N-1)/2, \cdots, (N-1)/2\}$ and by $I_N^c = \mathbb{Z}^2 \backslash I_N \times I_N$ the complement of $I_N \times I_N$ in $\mathbb{Z}^2$. For $h = L/N$, the set of points $\{x_{p,q}\}_{p,q \in I_N}$, where $x_{p,q} = (ph, qh)$, is a regular discretization of S_L. The DFT of G is

$$\hat{G}^N_{m,n} = \frac{1}{N^2} \sum_{p,q \in I_N} G(x_{p,q}) e^{-ik_{m,n} \cdot x_{p,q}},$$

and a similar formula holds for the DFT of U denoted $\hat{U}^N_{m,n}$. After an inverse DFT, what we compute numerically is

$$f_{L,N}(x) = L^2 \sum_{m,n \in I_N} \hat{G}^N_{m,n} \hat{U}^N_{m,n} e^{ik_{m,n} \cdot x}.$$

The goal is to characterize the error between f and $f_{L,N}$ for $x \in S_L$. There are three sources of errors: error in the approximation of the Fourier coefficients of G by its DFT, a similar error for U, and the error due to the truncation of the Fourier series of f_L. There are two parameters we can control: L and N, which determine the size of the domain and the number of discretization points.

We derive an error estimate of $f - f_{L,N}$ in Appendices B and C. We express N as $N = N_0 N_\lambda$, where $N_0 \geq 0$, $N_\lambda = [L\lambda^{-1}]$ and $[\cdot]$ denotes integer part. We expect $N_\lambda \gg 1$ in most practical situations.

Note that we do not expect a high accuracy since G is doubly singular: it does not decay fast enough to zero to be integrable, which inevitably introduces errors when periodizing; and it is singular at the origin; the parameter z then acts as a regularization.

For u a function defined on S_L, let

$$\|u\|^2 = \frac{1}{N^2} \sum_{p,q \in I_N} |u(x_{p,q})|^2,$$

which is an approximation of the average value of $|u|^2$ on S_L. Below, the notation $a \lesssim b$ means that there exists a non-dimensional constant $C > 0$, independent of the parameters of the problem, such that $a \leq Cb$. With $\hat{U}(k_{m,n}) = L^2 \hat{U}_{m,n}$ and $\hat{U}^N(k_{m,n}) = L^2 \hat{U}^N_{m,n}$, we obtain the following result, see Appendix B,

$$\frac{\|f - f_{L,N}\|}{f_{\text{ref}}} \lesssim \frac{1}{N_0^2 N_\lambda} + \frac{\max_{m,n \in I_N} |\hat{U}(k_{m,n}) - \hat{U}^N(k_{m,n})|}{\|U\|_{L^1(S_L)}}, \tag{7}$$

where f_{ref} is some reference value for f defined in Appendix B and $\|U\|_{L^1(S_L)} = \int_{S_L} |U(x)| dx$. The first term $1/N_0^2 N_\lambda$ combines the truncation error of the Fourier modes and the DFT error on G. It is optimal since all symmetries in G have been exploited to obtain the best possible decay. The second term is the DFT error on U, which varies depending on how smooth U is. Note that the estimate is uniform in the parameters k_0, ℓ and L, which is crucial since some constants in non uniform estimates can be large, in particular $k_0 L \gg 1$.

If the perturbation V presents sharp edges, we can only expect a slow decay of the DFT error on $U = V(U_s + U_{\rm i})$ of the form

$$\frac{\max_{m,n\in I_N} |\hat{U}(k_{m,n}) - \hat{U}^N(k_{m,n})|}{\|U\|_{L^1(S_L)}} \lesssim \frac{\ell_s}{LN} + \frac{\ell_s^2}{L\ell N},$$

where ℓ_s is the diameter of the transverse support of V in one layer. When the support is large so that $\ell_s \sim L$, the error is of order $(L\ell^{-1})/N$ and results from oscillations in U. When $\ell_s \sim \ell$, then the error is like $(\ell_s L^{-1})/N$. When the edges are smooth, when U is smooth, and when the support of V is large, the accuracy is limited by the oscillations of $U_s + U_{\rm i}$ and V, and we have estimates of the form

$$\frac{\max_{m,n\in I_N} |\hat{U}(k_{m,n}) - \hat{U}^N(k_{m,n})|}{\|U\|_{L^1(S_L)}} \lesssim \left(\frac{L\ell^{-1}}{N}\right)^q,$$

for $q \geq 1$ depending on how smooth U is. When λ is of the order of ℓ, which occurs when the random fluctuations and the incoming field have comparable wavelengths, then $L\ell^{-1} \sim N_\lambda$ and the error above is of order $(1/N_0)^q$, uniformly in the parameters.

We now turn to the second method, which is a modification of the original MLB scheme.

Method 2. The starting point is to write f as an inverse Fourier transform:

$$f(x) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} \hat{G}(\xi)\hat{U}(\xi)e^{i\xi\cdot x}d\xi, \qquad \text{where} \quad \hat{G}(\xi) = \frac{i}{8\pi^2}\frac{e^{i\sqrt{k_0^2-|\xi|^2}|z|}}{\sqrt{k_0^2-|\xi|^2}},$$

and the complex square root has a positive imaginary part. The main issue in the formula above is the singularity of $\hat{G}(\xi)$ when $|\xi| = k_0$, since a regular grid of wavenumbers $\{k_{m,n}\}_{m,n\in\mathbb{Z}}$ could have points arbitrarily close to k_0 leading to large numerical errors. It is therefore necessary to regularize $\hat{G}$, and a natural way to proceed is to add an artificial absorption $\eta > 0$ resulting in a complex-valued wavenumber $k_0(\eta) = \sqrt{k_0^2 + i\eta^2}$ with positive imaginary part. We then consider the regularized function f_η, obtained by replacing $G(x)$ by $G_\eta(x) = e^{ik_0(\eta)\sqrt{|x|^2+z^2}}(4\pi\sqrt{|x|^2+z^2})^{-1}$ in the definition of f. We recall that the original MLB method only considers wavenumbers ξ such that $|\xi| < k_0$, and therefore ignores the evanescent modes. We recall that we show in Appendix E that propagating and evanescent modes have a similar amplitude in a transverse far-field regime.

The first step to obtain a computable expression is to truncate the integral over ξ: with $K = 2\pi N/L$ and $S_K = [-K/2, K/2] \times [-K/2, K/2]$, let

$$f_K(x) = \frac{1}{(2\pi)^2}\int_{S_K} \hat{G}_\eta(\xi)\hat{U}(\xi)e^{i\xi\cdot x}d\xi.$$

We choose $K > k_0$ in order to account for the evanescent field. Setting $h = L/N$ as in the previous section, let

$$C_{p,q} = \frac{1}{K^2}\int_{S_K} \hat{G}_\eta(\xi)\hat{U}(\xi)e^{i\xi\cdot x_{p,q}}d\xi, \qquad x_{p,q} = h(p,q), \qquad p,q \in I_N,$$

so that $f_K(x_{p,q}) = K^2/(2\pi)^2 C_{p,q}$. The coefficients $C_{-p,-q}$ are the Fourier coefficients of the function $\hat{G}_\eta(\xi)\hat{U}(\xi)$. Let

$$\hat{U}^N(\xi) = \frac{L^2}{N^2} \sum_{p,q\in I_N} U(x_{p,q}) e^{-i\xi\cdot x_{p,q}},$$

where $\hat{U}^N_{m,n} = \hat{U}^N(k_{m,n})/L^2$ is the DFT of U which can be computed efficiently. We replace $\hat{U}$ by its DFT in $C_{p,q}$, giving

$$\widetilde{C}^N_{p,q} = \frac{1}{K^2}\int_{S_K} \hat{G}_\eta(\xi)\hat{U}^N(\xi) e^{i\xi\cdot x_{p,q}} d\xi, \qquad x_{p,q} = h(p,q), \qquad p,q \in I_N,$$

and we set

$$\widetilde{f}_{K,N}(x_{p,q}) = \frac{K^2}{(2\pi)^2}\widetilde{C}^N_{p,q}.$$

The Fourier coefficient $\widetilde{C}^N_{p,q}$ is then approximated by its DFT given by

$$C^N_{p,q} = \frac{1}{N^2}\sum_{m,n\in I_N} \hat{G}_\eta(k_{m,n})\hat{U}^N(k_{m,n}) e^{ik_{m,n}\cdot x_{p,q}} = \frac{L^2}{N^2}\sum_{m,n\in I_N} \hat{G}_\eta(k_{m,n})\hat{U}^N_{m,n} e^{ik_{m,n}\cdot x_{p,q}},$$

where $k_{m,n} = 2\pi/h(m,n)$ as in the previous section. Collecting all results, we find finally that the function f_η is approximated at $x_{p,q}$ by

$$f_{K,N}(x_{p,q}) = \frac{K^2}{(2\pi)^2} C^N_{p,q} = \sum_{m,n\in I_N} \hat{G}_\eta(k_{m,n})\hat{U}^N_{m,n} e^{ik_{m,n}\cdot x_{p,q}},$$

which can be calculated with an inverse DFT. With this construction, there are four sources of error: the regularization error after the introduction of $k_0(\eta)$, the truncation error, and two DFT errors when computing the DFT approximating the Fourier coefficients of $\hat{G}_\eta \hat{U}^N$ and of $\hat{U}$.

There are three free parameters in this discretization (we recall that the layer width z is fixed): L, which controls the grid size $2\pi/L$ in the wavenumber space, the parameter K (or equivalently N since $K = 2\pi N/L$), characterizing the maximal wavenumber accounted for, and the regularization parameter η. The theoretical analysis given in Appendix B shows it is beneficial to choose $\eta = k_0(k_0L)^{-\frac{1}{3+\varepsilon}}$, for an $\varepsilon > 0$ arbitrarily small, and $N = MN_\lambda$. The parameter M is such that the function U does no contain wavenumbers greater than Mk_0 (or equivalently that the contribution of wavenumbers larger than Mk_0 is negligible). Theorem B.2 shows it is enough to consider $M \lesssim (\lambda/\ell)N_z$ since larger wavenumbers are exponentially damped.

In order to compare with method 1, we set $N_0 = M$, so that methods 1 and 2 have the same number of discretization points. We recall that L is chosen so that the support of U is enclosed in S_L. Under assumptions on U that are verified in our setting of interest, we derive in Appendix B the estimate

$$\frac{\|f - f_{K,N}\|}{f_{\text{ref}}} \lesssim \frac{1}{N_\lambda^{\frac{2}{3+\varepsilon}}} + \frac{\max_{k\in S_K} |\hat{U}(k) - \hat{U}^N(k)|}{\|U\|_{L^1(S_L)}}, \tag{8}$$

where $\varepsilon > 0$ is arbitrarily small but not zero. We remark first that the DFT error on U is similar to that of method 1, which is expected. The other errors are overall of order $1/N_\lambda^{2/(3+\varepsilon)}$, which is nearly optimal (we expect the optimal rate to be $1/N_\lambda^{2/3} \log N_\lambda$, which can likely be obtained at the price of a more involved analysis). That rate has to be compared with $1/(N_\lambda M^2)$ of method 1, that is significantly better. In particular, the performance of method 1 vs method 2 increases as M gets larger and U contains higher wavenumbers. Note that the computational cost of method 1 is essentially the same as that of method 2 since it only requires in addition the calculation of the DFT of G, which can be performed at the beginning of the iterations once and for all. The conclusion is therefore that method 1 offers a better accuracy than method 2 for a comparable cost.

From now on, we only consider method 1 for calculation of the convolution in (1). In the case of the first order MLB (6), the approximation error of the z integral is of order $\Delta z(\Delta z \ell^{-1})$ when U is differentiable w.r.t. z, and therefore, after propagating over a distance L_z with $L_z \Delta z^{-1}$ layers, the error is of order $N_z^{-1} = [z\ell^{-1}]$. This can be improved by deriving high order versions of MLB. In a globally fourth order discretization of the z integral, the error is of order N_z^{-4}. High order schemes are derived in the next section.

The previous discussion applies to the case where V is smooth in the z variable, which occurs for instance when V models smooth random perturbations. When V models an inclusion such as a sphere with a different refractive index than the background, then V presents sharp transitions and the quadrature error in one layer is of order $\Delta z \Delta V \ell_V$, where ΔV quantifies the jump in refractive index and ℓ_V is the length of the boundary of the transverse support of V in the layer. After passing through all layers, the error is $\Delta z \Delta V$ times the surface of the scattering object. In that situation, there is no interest in moving to high order schemes since the error due to the jump singularity will dominate.

3 High order MLB

The starting point is expression (4). Different schemes are obtained depending on the order of the quadrature rule used to approximate the integral w.r.t. z. The simplest one is the first order MLB derived below. The calculation of the angular propagator $P_{\Delta z}$ is also necessary, and is equivalent to computing the convolution (1) with G replaced by $\partial_z G$. It is natural to use the Fourier representation of the propagator since the kernel $e^{i\sqrt{k_0^2-|\xi|^2}|\Delta z|}$, contrary to that of G, does not exhibit the singularity at k_0. Now, a closer look shows that the Fourier transform of the scattered field $U_{\rm s}$ to be propagated by $P_{\Delta z}$ and originating from the convolution in expression (6), is proportional to $\hat{G}$, which has the singularity. Hence, the evaluation of $P_{\Delta z} U_{\rm s}$ is essentially similar to that of the convolution in (6) since now $\hat{U}_{\rm s}$ is singular, and we expect a comparable accurary as in the calculation of (1).

Note that the methods of the previous section used to estimate the errors can be adapted to the calculation of $P_{\Delta z} U_{\rm s}$, and in addition to the DFT error due to $U_{\rm s}$, method 1 has an error of order $N_z/(M^2 N_\lambda)$, while that of method 2 is now $1/N_\lambda^{3/2}$. Method 1 becomes worse because of the increased singularity around zero in $\partial_z G$, while in the Fourier space, method 2 gets better since $e^{i\sqrt{k_0^2-|\xi|^2}|\Delta z|}$ has one more derivative than $\hat{G}$ w.r.t. ξ. In practical situations, we expect N_λ to be significantly larger than M and N_z, which makes method 2 more accurate and is our preferred choice here. As already mentioned, the dominant error term in the calculation of $P_{\Delta z}U_{\rm s}$ is due to the DFT error on $U_{\rm s}$, which the one obtained in the previous section with method 1.

We start with a modified version of first order MLB.

First order MLB. The integral w.r.t. z. in (4) is approximated using the rectangle rule, leading to

$$\Delta z \int_{\mathbb{R}^2} G(x-x',\Delta z)\left[V(U_{\rm i}+U_{\rm s})\right](x',\Delta z)dx' + O\left(\Delta z \frac{\Delta z}{\ell}\right),$$

where we recall that $\ell = \min(\lambda, \ell_c)$ and we use the Landau notation $O(\cdot)$. Denoting by $\star$ the convolution in the x variable (evaluated numerically using method 1), and setting $G_z(x) = G(x,z)$, the first order scheme reads for $n \geq 0$,

$$U_{\rm s}^{n+1}(x) \quad = \quad \Delta z G_{\Delta z} \star \left[V^n(U_{\rm i}^n + U_{\rm s}^n)\right] + P_{\Delta z}U_{\rm s}^n(x),$$

with initialization $U_s^0 = 0$. Above, we have set $V^n = V(z_n, x)$, $U_{\rm i}^n = U_{\rm i}(z_n, x)$, where z_n denotes the entry position of the n-th layer, and $U_{\rm s}^{n+1}(x)$ is an approximation of the scattered field in this n-th layer. The term $P_{\Delta z}U_{\rm s}^n$ is computed using method 2 without the regularization η since it is not necessary here.

Second order MLB. We now use a second order quadrature rule for the integral. We have to be careful in doing so since the standard trapezoidal rule involves the end points 0 and Δz and therefore the function $G(x,0)$, which is singular at $|x| = 0$. We then use the midpoint rule to approximate the integral, which is of order two and does not require the endpoints. We have then

$$\begin{aligned}&\int_0^{\Delta z} G(x-x',\Delta z - z')\left[V(U_{\rm i}+U_{\rm s})\right](x', z'+z)dz' \\ &= \Delta z\, G\left(x-x', \frac{\Delta z}{2}\right)\left[V(U_{\rm i}+U_{\rm s})\right]\left(x', z+\frac{\Delta z}{2}\right) + O\left(\Delta z\left(\frac{\Delta z}{\ell}\right)^2\right),\end{aligned}$$

and we use first order MLB to get $U_{\rm s}\left(x', z_n + \frac{\Delta z}{2}\right)$:

$$U_{\rm s}\left(x, z_n + \frac{\Delta z}{2}\right) \quad = \quad \frac{\Delta z}{2}\, G_{\Delta z/2} \star \left[V^n(U_{\rm i}^n + U_{\rm s}^n)\right] + P_{\Delta z/2}U_{\rm s}^n + O\left(\Delta z \frac{\Delta z}{\ell}\right).$$

The second order scheme then reads, for $n \geq 0$,

$$U_{\rm s}^{n+1} \quad = \quad \Delta z\, G_{\Delta z/2} \star \left[V^{n+\frac{1}{2}}(U_{\rm i}^{n+\frac{1}{2}} + U_{\rm s}^{n+\frac{1}{2}})\right] + P_{\Delta z}U_s^n$$

where $U_{\rm s}^0 = 0$ and

$$U_{\rm s}^{n+\frac{1}{2}} \quad = \quad \frac{\Delta z}{2}\, G_{\Delta z/2} \star \left[V^n(U_{\rm i}^n + U_{\rm s}^n)\right] + P_{\Delta z/2}U_{\rm s}^n, \qquad n \geq 0.$$

One iteration of second order MLB involves two additional convolutions to compute compared to first order MLB, for a total of four convolutions.

Fourth order MLB. As in the second order case, we cannot use a Simpson type formula of order four since it involves the end points. We instead employ Milne quadrature rule, which is a globally fourth order open Newton-Cotes type method to approximate the integral. The Simpson rule would have lead to the standard Runge-Kutta 4 algorithm (RK4), while the Milne rule yields a modified version of RK4.

We first rewrite the scattered field as

$$U_s\left(z+\Delta z\right) = U_1\left(z+\Delta z\right) + U_2\left(z+\Delta z\right),$$

where

$$\begin{aligned} U_1\left(z+\Delta z\right) &= P_{\Delta z} U_s(z) \\ U_2\left(z+\Delta z\right) &= \int_0^{\Delta z}\int_{\mathbb{R}^2} G(x-x',\Delta z - z')\left[V(U_{\rm i}+U_1+U_2)\right](x',z+z')dz'dx'. \end{aligned}$$

The Milne quadrature then yields

$$\begin{aligned} U_2\left(z+\Delta z\right) &= \frac{2\Delta z}{3} G_{3\Delta z/4}\star\left[V(U_{\rm i}+U_1+U_2)\right](\Delta z/4) \\ &\quad -\frac{\Delta z}{3} G_{\Delta z/2}\star\left[V(U_{\rm i}+U_1+U_2)\right](\Delta z/2) \\ &\quad +\frac{2\Delta z}{3} G_{\Delta z/4}\star\left[V(U_{\rm i}+U_1+U_2)\right](3\Delta z/4) + O\left(\Delta z\left(\frac{\Delta z}{\ell}\right)^4\right). \end{aligned}$$

There are two points to address in the formula above. First, U_2 is not known at the locations $\Delta z/4$, $\Delta z/2$ and $3\Delta z/4$. We then use explicit approximations of U_2 at these points in the spirit of RK4 that preserve the overall order of accuracy. Second, U_1 needs to be computed at $\Delta z/4$, $\Delta z/2$ and $3\Delta z/4$, which is done as in the second order case by using the propagator $P_{\Delta z}$. Let

$$H^n(t,X) \;=\; G_{\Delta z(1-t)}\star\left[V^{n+t}(U_{\rm i}^{n+t}+U_1^{n+t}+X)\right] \qquad n\geq 0.$$

The fourth order scheme then reads, for $n \geq 0$:

$$\begin{aligned} U_s^{n+1} &= U_1^{n+1}+U_2^{n+1} \\ U_1^{n+t} &= P_{t\Delta z}U_{\rm s}^n, \qquad U_1^0=0 \\ U_2^{n+1} &= \frac{\Delta z}{3}(2h_1^n - h_2^n + 2h_3^n), \end{aligned}$$

where

$$\begin{aligned} h_1^n &= H^n\left(\frac{1}{4},X_1^n\right), \qquad X_1^n = \frac{\Delta z}{4}H^n(0,0) \\ h_2^n &= H^n\left(\frac{1}{2},X_2^n\right), \qquad X_2^n = X_1^n + \frac{\Delta z}{4}h_1^n \\ h_3^n &= H^n\left(\frac{3}{4},X_3^n\right), \qquad X_3^n = X_2^n + \frac{\Delta z}{4}h_2^n. \end{aligned}$$

Computing one h_j^n requires 2 convolutions, so that computing U_2^{n+1} involves 6 convolutions. In total, U_{s}^{n+1} requires 7 convolutions, that is 5 more than the first order MLB.

The higher order schemes, whether second or fourth order, require $V(U_{\mathrm{i}} + U_s)$ be smooth since the quadrature error depends on the z derivative of $G \star V(U_{\mathrm{i}} + U_s)$ (third order derivative for second order MLB and fifth order for fourth order MLB). Indeed, a key observation is, while G is not itself regular w.r.t. to z, that a z derivative can be exchanged for an x derivative due to the particular form of G. More precisely, we write, for $U = V(U_{\mathrm{i}} + U_s)$,

$$\int_{\mathbb{R}^2} G(x', z - z')(U(x - x', z') - U(x, z'))dx' + \int_{\mathbb{R}^2} G(x', z - z')dx'U(x, z'), \tag{9}$$

and realize that, when $z - z' > 0$,

$$\int_{\mathbb{R}^2} G(x', z - z')dx' = \frac{i}{8\pi^2 k_0} e^{ik_0(z-z')}$$

which is smooth w.r.t. z. Then,

$$U(x - x', z') - U(x, z') \simeq -x' \cdot \nabla_x U(x, z'),$$

and the x' compensates the singularity arising when differentiating G w.r.t. z. Hence, a z derivative on G can be exchanged for an x derivative on U. The same idea applies to high order derivatives.

4 Conclusion

We proposed in this work two remedies for the evanescent mode issue of the original MLB. We showed with a rigorous error analysis that the one based on a direct computation of the DFT of the Green's function provides us with a natural regularization and yields better error estimates than the one based on regularizing by adding an artifical absorption. The mathematical analysis provided us in addition with an optimal way to choose the discretization parameters.

We have moreover derived high order numerical schemes based on high order quadrature rules for numerical integration. These methods theoretically offer a lower computation cost for a given accuracy than lower order schemes, and they are applicable when the background perturbations are smooth, as for instance in wave propagation in random smooth media. A numerical validation of these algorithms will be the object of a future work.

An aspect that is not covered in this article is the stability of the numerical scheme, namely what is the condition of the slab width Δz for the iterates of the field to remain bounded. Runge-Kutta type methods for ordinary differential equations are conditionnally stable, i.e. Δz has be small enough for the solution to not blow up, but in general the stability criteria are not too strict. We expect the same to hold here since our schemes are variations of Runge-Kutta methods.

A Derivation of the MLB scheme

The starting point is the three-dimensional Helmholtz equation and and we use notations introduced in Section 2. Since the Helmholtz equation can be recast as

$$\Delta(U - U_{\mathrm{i}}) + k_0^2(U - U_{\mathrm{i}}) + VU = 0,$$

we can express U with the integral equation

$$U(x, z) = U_{\mathrm{i}}(x, z) + \int_{\mathbb{R}^2}\int_{\mathbb{R}} G(x - x, z - z')(VU)(x', z')dx'dz',$$

where G is the Green's function defined in (2). It is more convenient to write the integral equation only in terms of the scattered field $U_{\mathrm{s}} = U - U_{\mathrm{i}}$, i.e.

$$U_s(x, z) = \int_{\mathbb{R}^2}\int_{\mathbb{R}} G(x - x', z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz'.$$

We suppose the incoming wave is propagating along the positive z axis and then separate the forward wave and backscattering as follows:

$$U_s(x, z) = \int_{\mathbb{R}^2}\int_{-\infty}^{z} G(x - x', z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz' + B_{sc}(x, z),$$

where

$$B_{sc}(x, z) = \int_{z}^{\infty} G(x - x', z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz'.$$

The goal now is to write the field at $z + \Delta z$ in terms of the one at z. We begin with

$$\begin{aligned} U_s(x, z + \Delta z) &= \int_{\mathbb{R}^2}\int_{-\infty}^{z+\Delta z} G(x - x', z + \Delta z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz' \\ &\quad + B_{sc}(x, z + \Delta z) \\ &= \int_{\mathbb{R}^2}\int_{z}^{z+\Delta z} G(x - x', z + \Delta z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'_{\perp}dz' \\ &\quad + \int_{\mathbb{R}^2}\int_{-\infty}^{z} G(x - x', z + \Delta z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz' \\ &\quad + B_{sc}(x, z + \Delta z). \end{aligned}$$

In order to get a recursive formula, we use the angular propagator $P_{\Delta z}$ defined in (5) and observe that

$$P_{\Delta z}G(x - x', z - z') = G(x - x', |z - z'| + |\Delta z|).$$

When $z > z'$ and $\Delta z > 0$, this gives $G(x - x', z - z' + \Delta z) = P_{\Delta z}G(x - x', z - z')$, and it follows that

$$\begin{aligned} U_s(x, z + \Delta z) &= \int_{\mathbb{R}^2}\int_{z}^{z+\Delta z} G(x - x', z + \Delta z - z')\left[V(U_{\mathrm{i}} + U_s)\right](x', z')dx'dz' \\ &\quad + P_{\Delta z}U_{\mathrm{s}}(x, z) - P_{\Delta z}B_{sc}(x, z) + B_{sc}(x, z + \Delta z). \end{aligned}$$

At this stage there is no approximation and the integral equation above is exact. A computationally tractable equation is obtained by exploiting the directionality of the incoming field and by neglecting backscattering, giving the expression

$$\begin{aligned} U_s(x, z + \Delta z) &\simeq \int_{\mathbb{R}^2} \int_z^{z+\Delta z} G(x - x', z + \Delta z - z') \left[V(U_{\mathrm{i}} + U_s)\right](x', z') dx' dz' \\ &\quad + P_{\Delta z} U_s(x, z), \end{aligned}$$

which is the starting point of MLB and extensions.

B Convergence results

We use the notations of Section 2. Consider the non-dimensional constants

$$\begin{aligned} c_1 &= (k_0 z)^{-1} + k_0 L \\ c_2 &= L k_0^{-1} z^{-2} + k_0^2 L^2. \end{aligned}$$

The next theorem, proved in Section C, concerns the convergence of method 1. We purposedly keep track of the dependency of various constants on important parameters of the problem such as k_0, L and z since these constants are large in relevant regimes.

Theorem B.1 *We have the estimate*

$$\|f - f_{L,N}\| \lesssim E_{\mathit{trunc+DFT\text{-}G}} + E_{\mathit{DFT\text{-}U}}.$$

where, for all $\varepsilon \in (0, 1)$*,*

$$\begin{aligned} E_{\mathit{trunc+DFT\text{-}G}} &= k_0 \big(L^2 \|U\| + L \|U\|_{L^2(S_L)}\big) \left(\frac{c_1}{N^2} + \frac{c_1^{1-\varepsilon} c_2^{\varepsilon}}{N^{2(1+\varepsilon)}} \right) \\ E_{\mathit{DFT\text{-}U}} &= L \|G\|_{L^2(S_L)} \max_{m,n \in I_N} |\hat{U}_{m,n} - \hat{U}^N_{m,n}|. \end{aligned}$$

The term $E_{\text{trunc+DFT-G}}$ combines the error due to the truncation of the Fourier series of f and the error due to the approximation of the Fourier coefficients of G by the DFT. The second term, $E_{\text{DFT-U}}$, is the error due to the approximation of the Fourier coefficients of U by the DFT. We observe that $E_{\text{trunc+DFT-G}}$ decays quadratically in N, and therefore the error decreases as can be expected as the grid get finer. The quadratic decay is optimal as all symmetries of G were exploited in the estimate.

The constants c_1 and c_2 are actually large since $k_0 L \gg 1$ and $z \ll \lambda$, as we expect oscillations over multiple wavelengths across the transverse domain. The parameters N and L have then to be chosen accordingly. Let first $N_z = [\ell z^{-1}]$ and $N_\lambda = [L \lambda^{-1}]$, where $[\cdot]$ denotes integer part. We have $N_z \gg 1$ and $N_\lambda \gg 1$. We have already assumed that L is greater than the diameter of the support of U. Supposing for instance that $N_z = 10$ for reasonable accuracy, we expect in practical situations that $N_\lambda \geq 10$, namely that the transverse support of U is at least ten wavelengths wide, so that $(k_0 z)^{-1} \lesssim k_0 L$. The dominating term in c_1 is then $k_0 L$.

Hence, $c_1 \lesssim N_\lambda$, and owing to the fact that $Lk_0^{-1}z^{-2} = Lk_0(k_0 z)^{-2}$, we have $c_2 \lesssim N_\lambda^3$. Setting $N = N_\lambda N_0$ and defining as reference

$$f_{\mathrm{ref}} = \max\Big(k_0\big(L^2\|U\| + L\|U\|_{L^2(S_L)}\big), L^{-1}(\|G\|_{L^2(S_L)} + \|G\|_{L^2(S_K)})\|U\|_{L^1(S_L)}\Big),$$

we obtain (7). This choice for the reference is motivated by two facts: (i) the terms appearing in f_{ref} are precisely the ones obtained in the estimates of Theorems 1 and 2 below, and (ii) the square of the second term is a direct estimate of the average value of $|f|^2$ over S_L.

In the scenario where $N_z \gg N_\lambda$, then the rate $1/(N_0^2 N_\lambda)$ has to be changed to $N_z/(N_0^2 N_\lambda^2)$. This occurs when the support of U is of the order of the wavelength or smaller, and models a localized scatterer. This is not the practical context we have in mind here since we are mostly interested in extended scattering media.

Regarding method 2, we will need the following quantity to state our main result: let

$$k_U = \max\left(\frac{\max_{j=1,2}\|\partial_{\xi_j}\hat{U}^N\|_{L^1(S_K)}}{\max_{\xi\in S_K}|\hat{U}^N(\xi)|}, \frac{\max_{i,j=1,2}\|\partial^2_{\xi_i\xi_j}\hat{U}^N\|_{L^1(S_K)}}{L\max_{\xi\in S_K}|\hat{U}^N(\xi)|}\right.$$
$$\left.\frac{\max_{i,j=1,2}\|\partial^2_{\xi_i^2\xi_j}\hat{U}^N\|_{L^1(S_K)}}{L^2\max_{\xi\in S_K}|\hat{U}^N(\xi)|}\right), \qquad (10)$$

which is a length quantifying how steep are the derivatives of $\hat{U}^N(\xi)$ after integration over S_K. We make the following assumptions that are not necessary but simplify expressions in the next theorem: $(k_0 L)^{-1} \lesssim \eta_0$ where $\eta^2 = k_0^2\eta_0$, and $\log(k_0 z) \le \eta_0^{-1/2}$. We have then the

Theorem B.2 *The following estimate holds:*

$$\|f - f_{K,N}\| \lesssim \mathcal{E}_{reg} + \mathcal{E}_{trunc} + \mathcal{E}_{DFT} + \mathcal{E}_{DFT\text{-}U},$$

where, for any $\varepsilon \in (0,1)$,

$$\begin{aligned}
\mathcal{E}_{reg} &= \|U\|_{L^1(S_L)} k_0 \eta_0 \\
\mathcal{E}_{trunc} &= \max_{|\xi|\ge K}|\hat{U}(\xi)| z^{-1} e^{-z\sqrt{K^2-k_0^2}} \\
\mathcal{E}_{DFT} &= \|U\|_{L^1(S_L)} L^{-1}(\alpha_U + \eta_0^{-1/2-\varepsilon}) \\
\mathcal{E}_{DFT\text{-}U} &= L^{-1}\|\hat{G}\|_{L^2(S_K)} \max_{\xi\in S_K}|\hat{U}(\xi) - \hat{U}^N(\xi)|
\end{aligned}$$

and where $\alpha_U = k_U k_0^{-1}\eta_0^{-1/2}$.

As expected, the regularization error $\mathcal{E}_{\mathrm{reg}}$ goes to zero as $\eta_0 \to 0$, and $\mathcal{E}_{\mathrm{trunc}} \to 0$ as the truncation parameter $K \to +\infty$. The parameter controlling the DFT error is L, as the cell size in wavenumbers is $2\pi/L$. Observe the term $\eta_0^{-1/2-\varepsilon}$ which is reminiscent from the singular nature of $\hat{G}$. These estimates are nearly optimal, we believe the term $\eta_0^{-1/2-\varepsilon}$ can be transformed into $\eta_0^{-1/2}|\log\eta_0|$ at the price of even more involved calculations.

We use now the previous theorem to set the parameters η_0, K and L. Ignore for the moment α_U and set it to zero. We then choose η_0 such that $\mathcal{E}_{\rm reg}$ and $\mathcal{E}_{\rm DFT}$ are about the same order, that is we set $k_0 L\eta_0 = \eta_0^{-1/2-\varepsilon}$, which is $\eta_0 = (k_0L)^{-\frac{2}{3+2\varepsilon}} \ll 1$. Note that with such a η_0, the assumptions $(k_0L)^{-1} \lesssim \eta_0$ and $\log(k_0 z) \leq \eta_0^{-1/2}$ are satisfied.

Concerning $\mathcal{E}_{\rm trunc}$, we assume that $\hat{U}$ does not have contributions for wavenumbers greater than some Mk_0 (or equivalently that contributions for $|\xi| \geq Mk_0$ are negligible), for some non-dimensional number M. We then set $K = Mk_0$, and writing K as $K = 2\pi N/L$, this amounts to choose N of the order of Mk_0L, that is MN_λ with previous notation. In that case, $\mathcal{E}_{\rm trunc}$ vanishes or is negligible compared to the other errors. We then obtain the estimate (8). This shows that the error $\|f - f_{K,N}\|$ is at best as (8) when $\alpha_U \neq 0$. Note that the exponential term in $\mathcal{E}_{\rm trunc}$ damps wavenumbers that are of the order of Δz^{-1}, so that it is enough to consider values of M that are less than $(\lambda/\ell)N_z$ since larger wavenumbers are exponentially damped.

We can actually estimate α_U in our cases of interest. We first replace $\hat{U}^N$ by $\hat{U}$ since the former is an approximation of the latter and we expect the associated α_U to be comparable. We treat $U_{\rm i}$ and U_s separately.

In practice $U_{\rm i}$ is often a plane wave, and as a consequence the Fourier transform of $VU_{\rm i}$ is obtained by shifting that of V. When V models an heterogeneous medium, it is usually numerically obtained as an inverse DFT, that is

$$V(x) = \varphi(x/L_\varphi) \sum_{m,n\in I_N} R(k_{m,n}) e^{ik_{m,n}\cdot x},$$

for some enveloppe function φ defining the support of V with diameter L_φ and a function R characterizing its the spatial frequency content. Its Fourier transform reads

$$\hat{V}(\xi) = L_\varphi^2 \sum_{m,n\in I_N} R(k_{m,n})\hat{\varphi}(L_\varphi(\xi - k_{m,n})).$$

By construction $L_\varphi \leq L$, and a short calculation shows that k_U is of the order of L_φ^{-1}. In that case, α_U is of the order of $L_\varphi^{-1}k_0^{-1}\eta_0^{-1}$. Supposing the incoming field oscillates multiple times over the support of V, we have $k_0/L_\varphi \lesssim 1$ and $\alpha_U \lesssim n_0^{-1/2}$. This means that (8) is actually sharp.

The situation is slightly different for the term VU_s. Let us consider expression (4) and ignore the angular propagator part since it is more regular than the first term. We then write $U_s = G_z \star W$ for some W. The leading part in $\widehat{VU_s}$ is then given by the convolution of $\hat{V}$ and $\hat{G}\hat{W}$. We have $\|\partial_{\xi_j}\hat{V}\star\hat{U}_s\|_{L^1(S_K)} \sim L_\varphi\|\partial_{\xi_j}\hat{V}\|_{L^1(S_K)}\|\hat{G}_z\hat{W}\|_{L^1(S_K)} \sim L_\varphi\|\hat{G}_z\hat{W}\|_{L^1(S_K)}$. On the other hand, since $\max_{\xi\in S_K}|\hat{V}(\xi)| \sim L_\varphi^2$, we have $\max_{\xi\in S_K}|\hat{V}\star(\hat{G}_zW)| \sim L_\varphi^2\|\hat{G}_z\hat{W}\|_{L^1(S_K)}$. There are similar relations for higher order derivatives, and in the end k_U/k_0 is of the order of $(k_0L_\varphi)^{-1} \lesssim 1$. We have therefore $\alpha_U \lesssim n_0^{-1/2}$ as above.

We proceed now to the proof of Theorem B.1.

C Proof of Theorem B.1

The first step is quantify the decay of the Fourier coefficients of G.

Step 1: Estimates on $\hat{G}_{m,n}$. We have the following lemma:

Lemma C.1 *Let*

$$\begin{aligned} C_0 &= L^{-1}, \qquad C_1 = k_0 + L^{-1} + z^{-1} + k_0^2 L \\ C_2 &= k_0 + L^{-1} + z^{-1} + Lz^{-2} + k_0^3 L^2. \end{aligned}$$

Then, for all $m, n \in \mathbb{Z}$,

$$|\hat{G}_{m,n}| \lesssim \Big(C_0^{-1} + C_1^{-1}(m^2+n^2) + C_2^{-1}|n|^{3/2}|m|^{3/2} \Big)^{-1}.$$

The estimate above is optimal since we have exploited all symmetries in G and there are boundary terms left that cannot be cancelled.

Proof. We recall that

$$G(x) = \frac{e^{ik_0\sqrt{|x|^2+z^2}}}{4\pi\sqrt{|x|^2+z^2}}, \qquad \hat{G}_{m,n} = \frac{1}{L^2}\int_{S_L} G(x) e^{-ik_{m,n}\cdot x} dx.$$

Denoting by B_L the ball of radius L centered at zero, we have first the following trivial bound

$$|\hat{G}_{m,n}| \;\leq\; \frac{1}{L^2}\int_{B_L} |G(x)| dx \leq \frac{2\pi}{L^2}\int_0^L \frac{r dr}{4\pi r} \lesssim L^{-1}. \tag{11}$$

We estimate now the partial derivatives of G. Let $r = \sqrt{|x|^2+z^2} = \sqrt{x_1^2+x_2^2+z^2}$. Then

$$\partial_{x_1} G(x) = ik_0 x_1 \frac{e^{ik_0 r}}{4\pi r^2} - x_1 \frac{e^{ik_0 r}}{4\pi r^3} = x_1 \big(ik_0 - 1/r\big) \frac{e^{ik_0 r}}{4\pi r^2}$$

and

$$\begin{aligned} \partial^2_{x_1^2} G(x) &= \big(ik_0 - 1/r\big)\frac{e^{ik_0 r}}{4\pi r^2} + x_1^2 \frac{e^{ik_0 r}}{4\pi r^5} + x_1\big(ik_0 - 1/r\big)\Big[ik_0 x_1 \frac{e^{ik_0 r}}{4\pi r^3} - x_1 \frac{e^{ik_0 r}}{2\pi r^4}\Big] \\ &= \big(ik_0 - 1/r\big)\frac{e^{ik_0 r}}{4\pi r^2} + x_1^2\big(ik_0 - 1/r\big)^2 \frac{e^{ik_0 r}}{4\pi r^3} + x_1^2 \frac{e^{ik_0 r}}{4\pi r^5}. \end{aligned}$$

Hence,

$$|\partial_{x_1} G(x)| \lesssim k_0/r + 1/r^2, \qquad |\partial^2_{x_1} G(x)| \lesssim r^{-3} + k_0^2/r,$$

and similar calculations give

$$|\partial^2_{x_1 x_2} G(x)| \lesssim r^{-3} + k_0^2/r, \qquad |\partial^3_{x_1^2 x_2} G(x)| + |\partial^3_{x_1 x_2^2} G(x)| \lesssim r^{-4} + k_0^3/r. \tag{12}$$

Using the previous estimates, we can now control $\hat{G}_{m,n}$. We have first, after an integration by parts, since G is even,

$$\hat{G}_{m,n} = \frac{1}{2i\pi Lm}\int_{S_L} \partial_{x_1} G(x) e^{-ik_{m,n}\cdot x} dx.$$

Another integration by parts gives

$$\hat{G}_{m,n} = \frac{1}{(2\pi)^2 m^2} B - \frac{1}{(2\pi)^2 m^2} \int_{S_L} \partial^2_{x_1} G(x) e^{-ik_{m,n}\cdot x} dx$$

where the boundary term can be controlled by

$$|B| \lesssim k_0 + L^{-1}.$$

Hence,

$$|\hat{G}_{m,n}| \lesssim |m|^{-2}\big(k_0 + L^{-1} + z^{-1} + k_0^2 L\big) = \frac{C_1}{m^2}. \tag{13}$$

The same calculation as above gives $|\hat{G}_{m,n}| \lesssim n^{-2} C_1$. We treat now the mixed derivatives. We use for this the fact that $\partial_{x_1} G(x)$ is even with respect to x_2 and obtain

$$\hat{G}_{m,n} = \frac{1}{(2i\pi)^2 mn} \int_{S_L} \partial^2_{x_1 x_2} G(x) e^{-ik_{m,n}\cdot x} dx.$$

An integration by parts with respect to x_1 gives

$$\hat{G}_{m,n} = \frac{L}{i(2\pi)^3 m^2 n} B' + \frac{L}{i(2\pi)^3 m^2 n} \int_{S_L} \partial^3_{x_1^2 x_2} G(x) e^{-ik_{m,n}\cdot x} dx \tag{14}$$

where the boundary term B' can be controlled by, using (12),

$$|B'| \lesssim k_0^2 L + L^{-1}.$$

Using (12) again, the second term in (14) is bounded by

$$\begin{aligned} \frac{L}{(2\pi)^3 m^2 |n|} \left(k_0^3 L + \int_0^L \frac{r dr}{(r^2+z^2)^2} \right) &\leq \frac{L}{(2\pi)^3 m^2 |n|} \left(k_0^3 L + z^{-2} \int_0^\infty \frac{r dr}{(r^2+1)^2} \right) \\ &\lesssim \frac{1}{m^2 |n|} \left(k_0^3 L^2 + L z^{-2} \right). \end{aligned}$$

An integration by parts with respect to x_2 instead yield a similar result with $m^2 n$ replaced by mn^2. Collecting the latter estimate and the one on B', gives, using that $2k_0^2 L \leq k_0 + k_0^3 L^2$,

$$(m^2|n| + n^2|m|)|\hat{G}_{m,n}| \lesssim \big(k_0 + L^{-1} + Lz^{-2} + k_0^3 L^2\big) = C_2.$$

Since $m^2|n| + n^2|m| = |m||n|(|m|+|n|) \geq |m|^{3/2}|n|^{3/2}$, this ends the proof together with (11) and (13). □

We continue with the truncation error

Step 2: truncation error. Let $\widetilde{f}_{L,N}(x)$ be the function obtained by truncating the Fourier series of f_L:

$$\widetilde{f}_{L,N}(x) = \sum_{m,n \in I_N} \hat{f}_{m,n} e^{ik_{m,n}\cdot x}.$$

Using the fact that

$$\sum_{p,q\in I_N} e^{ik_{m,n}\cdot x_{p,q}} = \frac{\sin(\pi m)\sin(\pi n)}{\sin(\pi m/N)\sin(\pi n/N)},$$

we find

$$\|f_L - \widetilde{f}_{L,N}\|^2 = \frac{1}{N^2}\sum_{p,q\in I_N} |(f_L - \widetilde{f}_{L,N})(x_{p,q})|^2 = \sum_{(m,n)\in I_N^c}\ \sum_{(\ell,\ell')\in J_N(m,n)} \hat{f}_{m,n}\hat{f}^*_{m+\ell N,n+\ell' N}$$

where $J_N(m,n) = \{\ell,\ell'\in\mathbb{Z} : (m+\ell N, n+\ell N)\in I_N^c\}$ and $I_N^c = \mathbb{Z}^2\backslash I_N\times I_N$. The Cauchy-Schwarz inequality applied once gives

$$\sum_{(\ell,\ell')\in J_N(m,n)} |\hat{f}_{m+\ell N,n+\ell' N}| \leq L^2 M_N \left(\sum_{(\ell,\ell')\in J_N(m,n)} |\hat{U}_{m+\ell N,n+\ell' N}|^2\right)^{1/2}$$

and another time yields

$$\|f_L - \widetilde{f}_{L,N}\|^2 \ \ \leq\ \ L^4 M_N \max_{(m,n)\in I_N^c} |\hat{G}_{m,n}| \sum_{m,n\in\mathbb{Z}} |\hat{U}_{m,n}|^2.$$

Above, we have introduced

$$M_N = \max_{(m,n)\in I_N^c}\left(\sum_{(\ell,\ell')\in J_N(m,n)} |\hat{G}_{m+\ell N,n+\ell' N}|^2\right)^{1/2}.$$

The estimate $|\hat{G}_{m,n}| \lesssim C_1(m^2+n^2)$ of Lemma C.1 gives $\max_{(m,n)\in I_N^c} |\hat{G}_{m,n}| \lesssim C_1 N^{-2}$. To study M_N, we write $m = m' + pN$, $n = n' + p'N$, with $m', n' \in I_N$ and $p, p' \in \mathbb{Z}$ so that $\hat{G}_{m+\ell N,n+\ell' N}$ becomes $\hat{G}_{m'+(p+\ell)N,n'+(p'+\ell')N}$. The term in the sum over (ℓ,ℓ') such that $\ell = -p$ and $\ell' = -p'$ reduces to $\hat{G}_{m',n'}$. Since $(\ell,\ell')\in J_N(m,n)$, this implies that at least $|m'|$ or $|n'|$ is greater than $(N-1)/2$ and therefore $|\hat{G}_{m,n}| \lesssim C_1 N^{-2}$. When $\ell = -p$ or $\ell' = -p'$, and when $\ell \neq -p$ and $\ell' \neq -p'$, we proceed similarly and obtain a term bounded again by $C_1 N^{-2}$. Consequently, $M_N \lesssim C_1 N^{-2}$.

Since Parseval's inequality yields

$$\sum_{m,n\in\mathbb{Z}} |\hat{U}_{m,n}|^2 = \frac{1}{L^2}\int_{S_L} |U(x)|^2 dx = \frac{1}{L^2}\|U\|^2_{L^2(S_L)},$$

we finally find

$$\|f_L - \widetilde{f}_{L,N}\| \lesssim L\|U\|_{L^2(S_L)} C_1/N^2. \tag{15}$$

This gives the truncation error. The next step is to use once more the discrete Parseval equality to arrive at

$$\|\widetilde{f}_{L,N} - f_{L,N}\|^2 = L^4 \sum_{m,n\in I_N} |\hat{G}_{m,n}\hat{U}_{m,n} - \hat{G}^N_{m,n}\hat{U}^N_{m,n}|^2.$$

The triangle inequality then gives

$$\begin{aligned} L^{-2}\|\widetilde{f}_{L,N}-f_{L,N}\| \;\; &\leq \;\; \left(\sum_{m,n\in I_N} |\hat{G}_{m,n}(\hat{U}_{m,n}-\hat{U}^N_{m,n})|^2\right)^{1/2} \\ &\qquad + \left(\sum_{m,n\in I_N} |\hat{U}^N_{m,n}(\hat{G}_{m,n}-\hat{G}^N_{m,n})|^2\right)^{1/2}. \end{aligned}$$

Parserval's inequality yields

$$\sum_{m,n\in I_N} |\hat{G}_{m,n}|^2 \leq \sum_{m,n\in \mathbb{Z}} |\hat{G}_{m,n}|^2 = \frac{1}{L^2}\int_{S_L} |G(x)|^2 dx = \frac{1}{L^2}\|G\|^2_{L^2(S_L)}$$

and we have as well

$$\sum_{m,n\in I_N} |\hat{U}^N_{m,n}|^2 = \frac{1}{N^2}\sum_{p,q\in I_N} |U(x_{p,q})|^2 = \|U\|^2.$$

Hence

$$\|\widetilde{f}_{L,N}-f_{L,N}\| \lesssim L^2\left(L^{-1}\|G\|_{L^2(S_L)} \max_{m,n\in I_N} |\hat{U}_{m,n}-\hat{U}^N_{m,n}| + \|U\| \max_{m,n\in I_N} |\hat{G}_{m,n}-\hat{G}^N_{m,n}|\right). \tag{16}$$

Step 3: the DFT error $\hat{G}_{m,n}-\hat{G}^N_{m,n}$. We use the following aliasing formula (adapted to 2D), see [2] Section 8.4, where the sum runs over p and q in $\mathbb{Z}$,

$$\hat{G}^N_{m,n}-\hat{G}_{m,n} = \sum_{(p,q)\neq(0,0)} \hat{G}_{m+pN,n+qN}, \qquad m,n\in I_N.$$

We break the sum into 3 terms:

$$\hat{G}^N_{m,n}-\hat{G}_{m,n} = \sum_{q\neq 0}\hat{G}_{m,n+qN} + \sum_{p\neq 0}\hat{G}_{m+pN,n} + \sum_{p\neq 0}\sum_{q\neq 0}\hat{G}_{m+pN,n+qN}. \tag{17}$$

Using Lemma C.1, the first term is bounded by

$$C_1\sum_{q\neq 0}\frac{1}{(n+qN)^2} = \frac{C_1}{N^2}\sum_{q\neq 0}\frac{1}{(\frac{n}{N}+q)^2} \lesssim \frac{C_1}{N^2},$$

where we used that $|n/N|<1/2$ since $n\in I_N$. We obtain the same result for the second term in (17). With Lemma C.1, we control the third term in (17) by, for any $\varepsilon\in(0,1)$,

$$C_1^{1-\varepsilon}C_2^{\varepsilon}\sum_{p\neq 0}\sum_{q\neq 0}\frac{1}{(n+qN)^{1+\varepsilon}(m+pN)^{1+\varepsilon}} \lesssim \frac{C_1^{1-\varepsilon}C_2^{\varepsilon}}{N^{2(1+\varepsilon)}}.$$

We introduced the parameter ε in order to minimize the power of the (large) constant C_2. As a conclusion, we have, for all $m,n\in I_N$:

$$|\hat{G}_{m,n}-\hat{G}^N_{m,n}| \lesssim \frac{C_1}{N^2} + \frac{C_1^{1-\varepsilon}C_2^{\varepsilon}}{N^{2(1+\varepsilon)}}, \qquad \forall\varepsilon\in(0,1). \tag{18}$$

Step 4: Conclusion. The triangle inequality gives

$$\|f_L - f_{L,N}\| \leq \|f_L - \widetilde{f}_{L,N}\| + \|f_{L,N} - \widetilde{f}_{L,N}\|.$$

Using (15)-(16)-(18) then yields

$$\|f_{L,N}-\widetilde{f}_{L,N}\| \lesssim L\|G\|_{L^2(S_L)} \max_{m,n\in I_N} |\hat{U}_{m,n}-\hat{U}^N_{m,n}|+L^2(\|U\|+L^{-1}\|U\|_{L^2(S_L)})\left(\frac{C_1}{N^2}+\frac{C_1^{1-\varepsilon}C_2^{\varepsilon}}{N^{2(1+\varepsilon)}}\right),$$

which ends the proof by noticing that $C_1 \lesssim k_0 c_1$ and $C_2 \lesssim k_0 c_2$ for c_1 and c_2 defined in Appendix B.

D Proof of Theorem B.2

We start with the error $f_\eta - f_{K,N}$ and split it into three terms using the triangle inequality:

$$\|f_\eta - f_{K,N}\| \quad \leq \quad \|f_\eta - f_K\| + \|f_K - \widetilde{f}_{K,N}\| + \|f_{K,N} - \widetilde{f}_{K,N}\|$$

where, using Parseval's equality in the second line,

$$\begin{aligned}
\|f_\eta - f_K\| &\leq \max_{p,q\in I_N} |(f - f_K)(x_{p,q})| \\
\|f_K - \widetilde{f}_{K,N}\|^2 &\leq N^{-2}\sum_{p,q\in\mathbb{Z}} |(f_K - \widetilde{f}_{K,N})(x_{p,q})|^2 = \frac{1}{L^2}\int_{S_K} |\hat{G}_\eta(\xi)|^2|\hat{U}(\xi) - \hat{U}^N(\xi)|^2 d\xi \\
&\leq L^{-2}\|\hat{G}_\eta\|^2_{L^2(S_K)} \max_{\xi\in S_K} |\hat{U}(\xi) - \hat{U}^N(\xi)|^2,
\end{aligned}$$

and

$$\|f_{K,N} - \widetilde{f}_{K,N}\| \quad \leq \quad \frac{K^2}{(2\pi)^2} \max_{p,q\in I_N} |C^N_{p,q} - \widetilde{C}^N_{p,q}|.$$

Step 1: the truncation error $f_\eta - f_K$. First, owing to (20), and since the square root function is of Hölder regularity 1/2, we have, for $|\xi| \geq k_0$,

$$\left|\Im\sqrt{k_0^2(\eta) - |\xi|^2} - \sqrt{|\xi|^2 - k_0^2}\right| \leq \eta/\sqrt{2}.$$

Hence,

$$\begin{aligned}
|(f_\eta - f_K)(x_{p,q})| &\lesssim \max_{|\xi|\geq K} |\hat{U}(\xi)| \int_K^\infty \frac{e^{-z\Im\sqrt{k_0^2(\eta)-r^2}}}{\sqrt{r^2 - k_0^2}} r dr \\
&\lesssim \max_{|\xi|\geq K} |\hat{U}(\xi)| \int_K^\infty \frac{e^{-z\sqrt{r^2-k_0^2}}}{\sqrt{r^2 - k_0^2}} r dr \\
&\lesssim \max_{|\xi|\geq K} |\hat{U}(\xi)| \int_{\sqrt{K^2-k_0^2}}^\infty e^{-zu} du \\
&\lesssim \max_{|\xi|\geq K} |\hat{U}(\xi)| \frac{e^{-z\sqrt{K^2-k_0^2}}}{z}
\end{aligned}$$

This is small either when $\hat{U}$ has negligible contributions when $|\xi| \geq K$, or when $K^2 \geq k_0^2 + \frac{a^2}{z^2}$, for some a sufficiently large.

Step 2: the DFT error $C^N_{p,q} - \widetilde{C}^N_{p,q}$. We can use the aliasing formula since the Green's function $\hat{G}_\eta$ is smooth thanks to the regularization. The main technicality in the analysis is to obtain (nearly) optimal estimates in the parameter η.

Let

$$\ell_U = \max \left(\left(\frac{\max_{i,j=1,2} \max_{\xi \in S_K} |\partial^3_{\xi_j^2 \xi_i} \hat{U}^N(\xi)|}{\max_{\xi \in S_K} |\hat{U}^N(\xi)|} \right)^{1/3}, \left(\frac{\max_{i,j=1,2} \max_{\xi \in S_K} |\partial^2_{\xi_i \xi_j} \hat{U}^N(\xi)|}{\max_{\xi \in S_K} |\hat{U}^N(\xi)|} \right)^{1/2} \right.$$
$$\left. \frac{\max_{j=1,2} \max_{\xi \in S_K} |\partial_{\xi_j} \hat{U}^N(\xi)|}{\max_{\xi \in S_K} |\hat{U}^N(\xi)|} \right). \tag{19}$$

We have the following lemma:

Lemma D.1 *Let $\eta_0 = \eta^2/k_0^2$, $h = L/N$, and for ℓ_U and k_U defined in* (19)-(10), *let $\alpha_U = k_U k_0^{-1} \eta_0^{-1/2}$ and $\beta_U = \ell_U + k_U k_0^{-2} \eta_0^{-3/2} + L k_U k_0^{-1} \eta_0^{-1/2}$, $\gamma_U = \ell_U^2 + k_U k_0^{-3} \eta_0^{-5/2} + L^2 k_U k_0^{-2} \eta_0^{-3/2}$. Define in addition, for any $\varepsilon \in (0,1)$,*

$$\begin{aligned} D_0 &= \alpha_U + \eta_0^{-1/2-\varepsilon}, \qquad D_1 = \beta_U + k_0^{-1} \eta_0^{-3/2-\varepsilon} \\ D_2 &= \gamma_U + k_0^{-2} \eta_0^{-5/2-\varepsilon}. \end{aligned}$$

Then, for $p, q \in \mathbb{Z}$,

$$K^2 |\widetilde{C}^N_{p,q}| \leq \max_{\xi \in S_K} |\hat{U}(\xi)| \Big(k_0^{-2} z^{-1} + D_0^{-1}(|ph| + |qh|) + D_1^{-1}(|ph|^2 + |qh|^2 + |ph||qh|)$$
$$+ D_2^{-1} |ph|^{3/2} |qh|^{3/2} \Big)^{-1}.$$

Proof. The proof is tedious but not difficult, and consists in estimating integrals of the partial derivatives of $\hat{G}_\eta \hat{U}^N$. We start with estimating $\hat{G}_\eta$.

Step 2a: estimates on $\hat{G}_\eta$. We denote by $\partial_{|\xi|}$ the radial derivative w.r.t. $|\xi|$. The following explicit representation of the complex square with positive imaginary part is helpful:

$$\sqrt{x + iy} = \frac{1}{\sqrt{2}} \mathrm{sign}(y) \sqrt{\sqrt{x^2 + y^2} + x} + \frac{i}{\sqrt{2}} \sqrt{\sqrt{x^2 + y^2} - x}. \tag{20}$$

With $x = k_0^2 - |\xi|^2$ and $y = \eta^2$, we find $\Im \sqrt{k_0^2(\eta) - |\xi|^2} \geq \Im \sqrt{k_0^2 - |\xi|^2}$. We have also $|\sqrt{k_0^2(\eta) - |\xi|^2}| = ((k_0^2 - |\xi|^2)^2 + \eta^4)^{1/4} \geq |k_0^2 - |\xi|^2|^{1/2}$. Using the last two results, we then find

$$|\hat{G}_\eta(\xi)| \lesssim \frac{e^{-\Im \sqrt{k_0^2 - |\xi|^2} z}}{|\sqrt{k_0^2 - |\xi|^2}|},$$

which is integrable around the singularity at $|\xi| = k_0$. The first order radial derivative reads

$$\partial_{|\xi|} \hat{G}_\eta(\xi) \quad = \quad 2i\pi |\xi| \frac{e^{i\sqrt{k_0^2(\eta) - |\xi|^2} z}}{k_0^2(\eta) - |\xi|^2} \left(-z + \left(k_0^2(\eta) - |\xi|^2 \right)^{-1/2} \right),$$

which has non-integrable singularities and we need to use the regularization parameter η. For this, let $k_0(\eta) = \sqrt{k_0^2 + i\eta^2}$ and set $\eta = k_0\sqrt{\eta_0}$ for simplicity. It follows directly from (20) with $x = k_0^2$ and $y = \eta^2$ that $\Re k_0(\eta) \geq k_0$, and that, for $0 \leq \eta_0 \leq 1$, $\Im k_0(\eta) \geq k_0\eta_0/(2\sqrt{2})$. These previous inequalities give

$$|k_0^2(\eta) - |\xi|^2| = |k_0(\eta) - |\xi|||k_0(\eta) + |\xi|| \gtrsim \eta_0 k_0(k_0 + |\xi|).$$

For $0 \leq |\xi| \leq 2k_0$, $0 < \varepsilon < 1$, this yields a first bound

$$|\partial_{|\xi|}\hat{G}_\eta(\xi)| \lesssim \frac{|\xi|}{k_0|k_0 - |\xi||^{1-\varepsilon}(\eta_0 k_0)^\varepsilon}\left(z + \left(\eta_0 k_0^2\right)^{-1/2}\right),$$

which has now an integrable singularity at k_0. The expression can be simplified using that $z \leq \lambda \leq k_0^{-1}$ and that $\eta_0 < 1$. For $|\xi| \geq 2k_0$, we have $|\xi|^2 - k_0^2 \geq 3|\xi|^2/4$, so that

$$|\partial_{|\xi|}\hat{G}_\eta(\xi)| \lesssim \frac{e^{-\Im\sqrt{k_0^2-|\xi|^2}z}}{|\xi|}\left(z + |\xi|^{-1}\right).$$

Combining both estimates we find

$$|\partial_{|\xi|}\hat{G}_\eta(\xi)| \lesssim \begin{cases} \dfrac{|\xi|}{k_0^2|k_0 - |\xi||^{1-\varepsilon}(\eta_0 k_0)^\varepsilon \eta_0^{1/2}} & \text{for} \quad |\xi| \leq 2k_0 \\ \dfrac{e^{-\Im\sqrt{k_0^2-|\xi|^2}z}}{|\xi|}\left(z + |\xi|^{-1}\right) & \text{for} \quad |\xi| \geq 2k_0. \end{cases} \tag{21}$$

Regarding the second order radial derivative we have

$$\begin{aligned} \partial^2_{|\xi|}\hat{G}_\eta(\xi) \quad = \quad & 2i\pi\frac{e^{i\sqrt{k_0^2(\eta)-|\xi|^2}z}}{k_0^2 - |\xi|^2}\left(-z + \left(k_0^2(\eta) - |\xi|^2\right)^{-1/2}\right) \\ & 2\pi z|\xi|^2\frac{e^{i\sqrt{k_0^2(\eta)-|\xi|^2}z}}{(k_0^2(\eta) - |\xi|^2)^{3/2}}\left(-z + \left(k_0^2(\eta) - |\xi|^2\right)^{-1/2}\right) \\ & +4i\pi|\xi|^2\frac{e^{i\sqrt{k_0^2(\eta)-|\xi|^2}z}}{(k_0^2(\eta) - |\xi|^2)^2}\left(-z + \left(k_0^2(\eta) - |\xi|^2\right)^{-1/2}\right) \\ & +2i\pi|\xi|^2\frac{e^{i\sqrt{k_0^2(\eta)-|\xi|^2}z}}{(k_0^2(\eta) - |\xi|^2)^{5/2}} \end{aligned}$$

Proceeding as in the first-orde case, we find when $0 \leq |\xi| \leq 2k_0$,

$$\begin{aligned} &|\partial^2_{|\xi|^2}\hat{G}_\eta(\xi)| \lesssim \frac{e^{-\Im\sqrt{k_0^2-|\xi|^2}z}}{|k_0(\eta)^2 - |\xi|^2|}\times \\ &\left(|z| + \frac{1}{((k_0 + |\xi|)\eta_0 k_0)^{1/2}} + \frac{z^2|\xi|^2}{((k_0 + |\xi|)\eta_0 k_0)^{1/2}} + \frac{z|\xi|^2}{((k_0 + |\xi|)\eta_0 k_0)} + \frac{|\xi|^2}{((k_0 + |\xi|)\eta_0 k_0)^{3/2}}\right). \end{aligned}$$

Using that $0 \leq |\xi| \leq 2k_0$, $z \leq \lambda \leq k_0^{-1}$, and that $\eta_0 < 1$, this simplifies to

$$|\partial^2_{|\xi|^2}\hat{G}_\eta(\xi)| \lesssim \frac{1}{k_0^2|k_0 - |\xi||^{1-\varepsilon}(\eta_0 k_0)^\varepsilon \eta_0^{3/2}}.$$

When $|\xi| \geq 2k_0$, we proceed as above, and find for all $|\xi|$,

$$|\partial^2_{|\xi|}\hat{G}_\eta(\xi)| \lesssim \begin{cases} \dfrac{1}{k_0^2|k_0-|\xi||^{1-\varepsilon}(\eta_0 k_0)^\varepsilon \eta_0^{3/2}} & \text{for} \qquad |\xi| \leq 2k_0 \\ \dfrac{e^{-\Im\sqrt{k_0^2-|\xi|^2}z}}{k_0|\xi|}\left(z+|\xi|^{-1}\right) & \text{for} \qquad |\xi| \geq 2k_0. \end{cases} \tag{22}$$

For the third order derivative, the calculations are very similar and we give the result without proof: we find

$$|\partial^3_{|\xi|}\hat{G}_\eta(\xi)| \lesssim \begin{cases} \dfrac{1}{k_0^3|k_0-|\xi||^{1-\varepsilon}(\eta_0 k_0)^\varepsilon \eta_0^{5/2}} & \text{for} \qquad |\xi| \leq 2k_0 \\ \dfrac{e^{-\Im\sqrt{k_0^2-|\xi|^2}z}}{k_0^2|\xi|}\left(z+|\xi|^{-1}\right) & \text{for} \qquad |\xi| \geq 2k_0. \end{cases} \tag{23}$$

The partials of $\hat{G}_\eta$ can finally be controlled using the radial derivatives by

$$|\partial^2_{\xi_1\xi_2}\hat{G}_\eta(\xi)| \;\lesssim\; |\partial^2_{|\xi|}\hat{G}_\eta(\xi)| + \frac{1}{|\xi|}|\partial_{|\xi|}\hat{G}_\eta(\xi)| \tag{24}$$

$$|\partial^3_{\xi_i^2\xi_j}\hat{G}_\eta(\xi)| \;\lesssim\; |\partial^3_{|\xi|}\hat{G}_\eta(\xi)| + \frac{1}{|\xi|}|\partial^2_{|\xi|}\hat{G}_\eta(\xi)| + \frac{1}{|\xi|^2}|\partial_{|\xi|}\hat{G}_\eta(\xi)|. \tag{25}$$

We can now turn to the proof of Lemma D.1. We define $H(\xi) = \hat{G}_\eta \hat{U}^N$, and $x = (x_1, x_2) = h(p,q)$ for simplicity. Since $\hat{G}_\eta$ is even, since $\hat{U}^N$ is periodic, and since $\partial_{\xi_1}\hat{G}_\eta$ is even w.r.t ξ_2, we find after two integrations by parts,

$$\begin{aligned} H^F(x) &:= \int_{S_K} H(\xi)e^{i\xi\cdot x}d\xi = -\frac{1}{ix_1}\int_{S_K}\partial_{\xi_1}H(\xi)e^{i\xi\cdot x}d\xi \\ &= \frac{1}{x_1^2}\left[\partial_{\xi_1}H(\xi)e^{i\xi\cdot x}\right] - \frac{1}{x_1^2}\int_{S_K}\partial^2_{\xi_1}H(\xi)e^{i\xi\cdot x}d\xi, \end{aligned} \tag{26}$$

as well as

$$\begin{aligned} H^F(x) &= -\frac{1}{x_1x_2}\int_{S_K}\partial^2_{\xi_1\xi_2}H(\xi)e^{i\xi\cdot x}d\xi \\ &= \frac{i}{x_1^2x_2}\left[\partial^2_{\xi_1\xi_2}H(\xi)e^{i\xi\cdot x}\right] - \frac{i}{x_1^2x_2}\int_{S_K}\partial^3_{\xi_1^2\xi_2}H(\xi)e^{i\xi\cdot x}d\xi. \end{aligned}$$

Above, the terms in brackets denote the boundary terms coming from the integration by parts. We proceed now to estimating the integrals. We find, using (21), for all $\varepsilon \in (0,1)$,

$$\begin{aligned} \int_{S_K}|\partial_{|\xi|}\hat{G}_\eta(\xi)|d\xi \;&\lesssim\; \eta_0^{-1/2-\varepsilon}\int_{r\leq 2}\frac{dr}{|1-r|^{1-\varepsilon}} \\ &\quad + \int_{r\geq 2}(zk_0+r^{-1})e^{-\sqrt{r^2-1}k_0z}d\xi. \end{aligned}$$

When $r \geq 2$, we have $r^2-1 \geq 3r^2/4$, and using this in the second integral gives

$$\int_{r\geq k_0z}(1+r^{-1})e^{-r}dr \lesssim \log(k_0z).$$

Using the assumption that $\log(k_0 z) \lesssim \eta_0^{-1}$, we find

$$\int_{S_K} |\partial_{|\xi|} \hat{G}_\eta(\xi)| d\xi \lesssim \eta_0^{-1/2-\varepsilon}.$$

Proceeding in the same manner and using (22)-(23), we find, for all $\varepsilon \in (0,1)$,

$$\begin{aligned}
\int_{S_K} |\xi|^{-1} |\partial_{|\xi|} \hat{G}_\eta(\xi)| d\xi &\lesssim k_0^{-1} \eta_0^{-1/2-\varepsilon} \\
\int_{S_K} |\partial^2_{|\xi|} \hat{G}_\eta(\xi)| d\xi &\lesssim k_0^{-1} \eta_0^{-3/2-\varepsilon} \\
\int_{S_K} |\xi|^{-1} |\partial^2_{|\xi|} \hat{G}_\eta(\xi)| d\xi &\lesssim k_0^{-2} \eta_0^{-3/2-\varepsilon} \\
\int_{S_K} |\partial^3_{|\xi|} \hat{G}_\eta(\xi)| d\xi &\lesssim k_0^{-2} \eta_0^{-5/2-\varepsilon},
\end{aligned}$$

which yields, from (24)-(25),

$$\int_{S_K} |\partial_{\xi_j} \hat{G}_\eta(\xi)| d\xi \lesssim \eta_0^{-1/2-\varepsilon} \tag{27}$$

$$\int_{S_K} |\partial^2_{\xi_j} \hat{G}_\eta(\xi)| + |\partial^2_{\xi_i \xi_j} \hat{G}_\eta(\xi)| d\xi \lesssim k_0^{-1} \eta_0^{-3/2-\varepsilon} \tag{28}$$

$$\int_{S_K} |\partial^3_{\xi_j^2 \xi_i} \hat{G}_\eta(\xi)| d\xi \lesssim k_0^{-2} \eta_0^{-5/2-\varepsilon}. \tag{29}$$

From (26) and (27), this gives the first estimate

$$\begin{aligned}
|x_j||H^F(x)| &\lesssim \max_{\xi \in S_K} |\hat{G}_\eta(\xi)| \int_{S_K} |\partial_{\xi_j} \hat{U}(\xi)| d\xi + \eta_0^{-1/2-\varepsilon} \max_{\xi \in S_K} |\hat{U}(\xi)| \\
&\lesssim \max_{\xi \in S_K} |\hat{U}(\xi)| (k_U k_0^{-1} \eta_0^{-1/2} + \eta_0^{-1/2-\varepsilon}),
\end{aligned}$$

since $\max_{\xi \in S_K} |\hat{G}_\eta(\xi)| \lesssim k_0^{-1} \eta_0^{-1/2}$. We recall that k_U is defined in (10). This gives

$$|x_j||H^F(x)| \lesssim \max_{\xi \in S_K} |\hat{U}(\xi)| (\alpha_U + \eta_0^{-1/2-\varepsilon}) = \max_{\xi \in S_K} |\hat{U}(\xi)| D_0,$$

where $\alpha_U = k_U k_0^{-1} \eta_0^{-1/2}$.

We move on now to the second order derivative. We find that the boundary term $\left[\partial_{\xi_1} H(\xi) e^{i\xi \cdot x}\right]$ is bounded by

$$\max_{\xi \in S_K} |\partial_{\xi_1} \hat{U}(\xi)| + \max_{\xi \in S_K} |\hat{U}(\xi)| K^{-1}. \tag{30}$$

We have

$$\begin{aligned}
\int_{S_K} |\partial^2_{\xi_j} H(\xi) e^{i\xi \cdot x}| d\xi \leq\ & \max_{\xi \in S_K} |\hat{U}(\xi)| \int_{S_K} |\partial^2_{\xi_j} \hat{G}_\eta(\xi)| d\xi \\
& + \max_{\xi \in S_K} |\partial_{\xi_j} \hat{G}_\eta(\xi)| \int_{S_K} |\partial_{\xi_j} \hat{U}(\xi)| d\xi \\
& + \max_{\xi \in S_K} |\hat{G}_\eta(\xi)| \int_{S_K} |\partial^2_{\xi_j} \hat{U}(\xi)| d\xi.
\end{aligned}$$

Using that (30), the fact that $\max_{\xi\in S_K} |\partial_{x_j}\hat{G}_\eta(\xi)| \lesssim k_0^{-2}\eta_0^{-3/2}$ and (28), this gives

$$\begin{aligned}&(|x_j|^2+|x_i||x_j|)|H^F(x)| \\ &\qquad\lesssim \max_{\xi\in S_K}|\hat{U}(\xi)|(\ell_U+K^{-1}+k_Uk_0^{-2}\eta_0^{-3/2}+Lk_Uk_0^{-1}\eta_0^{-1/2}+k_0^{-1}\eta_0^{-3/2-\varepsilon}),\end{aligned}$$

and therefore

$$(|x_j|^2+|x_i||x_j|)|H^F(x)| \ \lesssim \ \max_{\xi\in S_K}|\hat{U}(\xi)|(\beta_U+k_0^{-1}\eta_0^{-3/2-\varepsilon}) = \max_{\xi\in S_K}|\hat{U}(\xi)|D_1,$$

where $\beta_U = \ell_U + k_Uk_0^{-2}\eta_0^{-3/2} + Lk_Uk_0^{-1}\eta_0^{-1/2}$.

With similar calculations involving (29), we find with the third order derivative,

$$|x_i|^2|x_j||H^F(x)| \ \lesssim \ \max_{\xi\in S_K}|\hat{U}(\xi)|(\gamma_U+k_0^{-2}\eta_0^{-5/2-\varepsilon}) = \max_{\xi\in S_K}|\hat{U}(\xi)|D_2,$$

where $\gamma_U = \ell_U^2 + k_Uk_0^{-3}\eta_0^{-5/2} + L^2k_Uk_0^{-2}\eta_0^{-3/2}$. This ends the proof. □

We have now everything needed to quantify the DFT error.

Step 2c: conclusion. The aliasing formula gives for $p, q \in I_N$,

$$C_{p,q}^N - \widetilde{C}_{p,q}^N = \sum_{(m,n)\neq(0,0)} \widetilde{C}_{p+mN,q+nN}^N.$$

We break the sum into 3 terms:

$$C_{p,q}^N - \widetilde{C}_{p,q}^N = \sum_{n\neq 0}\widetilde{C}_{p,q+nN}^N + \sum_{m\neq 0}\widetilde{C}_{p+mN,q}^N + \sum_{m\neq 0}\sum_{n\neq 0}\widetilde{C}_{p+mN,q+nN}^N. \tag{31}$$

Using Lemma D.1, the first term is bounded by, for any $\varepsilon' \in (0,1)$, up to the term $\max_{\xi\in S_K}|\hat{U}(\xi)|/K^2$ that will be reintroduced in the end,

$$\frac{D_0^{1-\varepsilon'}D_1^{\varepsilon'}}{h^{1+\varepsilon'}}\sum_{n\neq 0}\frac{1}{(q+nN)^{1+\varepsilon'}} = \frac{D_0^{1-\varepsilon'}D_1^{\varepsilon'}}{(Nh)^{1+\varepsilon'}}\sum_{n\neq 0}\frac{1}{(\frac{q}{N}+n)^{1+\varepsilon'}} \leq \frac{D_0^{1-\varepsilon'}D_1^{\varepsilon'}}{L^{1+\varepsilon'}},$$

where we used that $q \in I_N$ so that $|q/N| \leq 1/2$. Using the subadditivity of the function x^α for $x > 0$ and $\alpha \in (0,1)$, we find

$$\begin{aligned} D_0^{1-\varepsilon'}(L^{-1}D_1)^{\varepsilon'} \ &\leq \ \big(\alpha_U^{1-\varepsilon'} + \eta_0^{-(1/2+\varepsilon)(1-\varepsilon')}\big)\big((L^{-1}\beta_U)^{\varepsilon'} + (k_0L)^{-\varepsilon'}\eta_0^{-(3/2+\varepsilon)\varepsilon'}\big) \\ &\lesssim \ (\alpha_U)^{1-\varepsilon'}(k_0\ell_U^2)^{\varepsilon'} + k_0^{-\varepsilon'}\eta_0^{-1/2-\varepsilon-\varepsilon'}(1+\eta^{3\varepsilon'/2+\varepsilon\varepsilon'}(k_0\ell_U)^{2\varepsilon'})). \end{aligned}$$

To simplifty the expression above, we remark that $\ell_U \lesssim L$. It is indeed clear that ℓ_U is bounded by the support of U, which is itself bounded by L by construction. Using the assumption $(k_0L)^{-1} \lesssim \eta$, we have $L^{-1}\beta_U \lesssim \alpha_U$. Then,

$$\begin{aligned} D_0^{1-\varepsilon'}(L^{-1}D_1)^{\varepsilon'} \ &\leq \ \big(\alpha_U^{1-\varepsilon'} + \eta_0^{-(1/2+\varepsilon)(1-\varepsilon')}\big)\big((\alpha_U)^{\varepsilon'} + (k_0L)^{-\varepsilon'}\eta_0^{-(3/2+\varepsilon)\varepsilon'}\big) \\ &\lesssim \ \alpha_U + \eta_0^{-1/2-\varepsilon}. \end{aligned}$$

We obtain the same result for the second term in (31). For the third term, we control it by

$$\frac{D_1^{1-\varepsilon'}D_2^{\varepsilon'}}{h^{2+\varepsilon'}}\sum_{m\neq 0}\sum_{n\neq 0}\frac{1}{(p+nN)^{1+\varepsilon'}(q+mN)^{1+\varepsilon'}} \leq \frac{D_1^{1-\varepsilon'}D_2^{\varepsilon'}}{L^{2+\varepsilon'}},$$

where we find

$$\begin{aligned}(L^{-1}D_1)^{1-\varepsilon'}(L^{-1}D_2)^{\varepsilon'} \quad &\leq \quad \big((L^{-1}\beta_U)^{1-\varepsilon'}+(k_0L)^{-(1-\varepsilon')}\eta_0^{-(3/2+\varepsilon)(1-\varepsilon')}\big)\\ &\qquad \times\big((L^{-2}\gamma_U)^{\varepsilon'}+(k_0L)^{-2\varepsilon'}\eta_0^{-(5/2+\varepsilon)\varepsilon'}\big).\end{aligned}$$

As before, using that $(k_0L)^{-1} \lesssim \eta$ and $\ell_U \lesssim L$, we find $L^{-2}\gamma_U \lesssim \alpha_U$. The above estimate then reduces to

$$(L^{-1}D_1)^{1-\varepsilon'}(L^{-1}D_2)^{\varepsilon'} \quad \lesssim \quad \alpha_U+\eta_0^{-1/2-\varepsilon}.$$

Hence, going back to (31) and collecting previous estimates, we finally find

$$|C_{p,q}^N-\widetilde{C}_{p,q}^N| \lesssim K^{-2}L^{-1}\max_{\xi\in S_K}|\hat{U}(\xi)|\big(\alpha_U+\eta_0^{-1/2-\varepsilon}\big).$$

which allows us to estimate the DFT error. It remains now to treat the regularization error to end the proof of Theorem B.2.

Step 3: the regularization error. We write $k_0(\eta) = k_0\sqrt{1+\eta_0} = k_r+ik_i$, and express $e^{ik_0(\eta)r}-e^{ik_0r}$ as

$$e^{ik_0(\eta)r}-e^{ik_0r} = e^{ik_rr}e^{-k_ir}-e^{ik_0r}e^{-k_ir}+e^{ik_0r}e^{-k_ir}-e^{ik_0r}.$$

This yields the estimate

$$\left|e^{ik_0(\eta)r}-e^{ik_0r}\right| \leq r|k_r-k_0|+rk_i.$$

Using (20), we find

$$\frac{1}{\sqrt{2}}\sqrt{\sqrt{1+y^2}+1}-1 = \frac{1}{2\sqrt{2}}\int_0^y\frac{x}{\sqrt{1+x^2}\sqrt{\sqrt{1+x^2}+1}}dx \leq \frac{y^2}{4\sqrt{2}}.$$

With $y=\eta_0$, this gives $|k_r-k_0| \lesssim k_0\eta_0^2$. Regarding the imaginary part, we have when $y\geq 0$,

$$\sqrt{1+y^2}-1 = \int_0^y\frac{dx}{\sqrt{1+x^2}} \leq y,$$

so that

$$\frac{1}{\sqrt{2}}\sqrt{\sqrt{1+\eta_0^2}-1} \leq \frac{\eta_0}{\sqrt{2}}$$

and $k_r \lesssim k_0\eta_0^2$. As a conclusion, since $\eta_0\leq 1$, $\left|e^{ik_0(\eta)r}-e^{ik_0r}\right| \lesssim k_0r\eta_0$. Then,

$$\begin{aligned}|f(x)-f_\eta(x)| \quad &\leq \quad \int_{\mathbb{R}^2}|G(x-y)-G_\eta(x-y)||U(y)|dy\\ &\leq \quad \lesssim k_0\eta_0\int_{\mathbb{R}^2}|U(y)|dy.\end{aligned}$$

This concludes the proof of Theorem B.2.

E Removing evanescent modes

Consider

$$g(x)=\int_{\mathbb{R}^2}\frac{e^{i\sqrt{k_0^2-|\xi|^2}z+i\xi\cdot x}}{\sqrt{k_0^2-|\xi|^2}}\hat{U}(\xi)d\xi,$$

which is proportional to the Fourier transform of f (defined in (1)), neglecting irrelevant constant. The function $\hat{U}(\xi)$ is smooth, decays to 0 for $|\xi|\gg k_0$ but is not necessarily integrable. We change variables to arrive at

$$g(x)=k_0\int_{\mathbb{R}^2}\frac{e^{i\sqrt{1-|\xi|^2}k_0 z+ik_0\xi\cdot x}}{\sqrt{1-|\xi|^2}}\hat{U}(k_0\xi)d\xi,$$

We can use stationary phase techniques to analyze g. Suppose first that we are in a far field regime in all directions (i.e. transverse along x and longitudinal along z), so that both $k_0|x|=1/\varepsilon\gg 1$ and $zk_0\gg 1$. A short calculation shows that there is a stationary point in $|\xi|<1$ (equal to $\xi=\hat{x}/(z^2/x^2+1)^{1/2}$, $\hat{x}=x/|x|$), which gives the leading contribution. So as expected the contribution of evanescent modes, corresponding to $|\xi|>1$, is negligible compared to that of the propagating modes.

In MLB, z is chosen of the order of $\min(\lambda,\ell_c)$ for the numerical scheme to be accurate. This means that we are always in a near field regime going from one layer to the next, independently of how large $|x|$ is. In that case, there is no stationary point, and the leading contribution comes from the region around $|\xi|=1$ where the integrand is singular, and there is no particular reason for the evanescent modes to be negligible. We confirm this with the following analysis. Let

$$g_{\text{ev}}(x)=k_0\int_{|\xi|\geq 1}\frac{e^{-\sqrt{|\xi|^2-1}k_0 z+ik_0\xi\cdot x}}{i\sqrt{|\xi|^2-1}}\hat{U}(k_0\xi)d\xi,$$

be the evanescent modes contribution. To simplify notation and without lack of generality, we suppose that U is real-valued so that $\hat{U}(-\xi)=\hat{U}^*(\xi)$. We derive below an asymptotic expression of g_{ev} as $\varepsilon\to 0$. Going to polar coordinates and changing variables, we find

$$\begin{aligned}g_{\text{ev}}(x)&=k_0\int_0^{2\pi}\int_{r\geq 1}\frac{e^{-\sqrt{r^2-1}k_0 z+ir\cos\theta/\varepsilon}}{i\sqrt{r^2-1}}\hat{U}\big(k_0 r[\cos\theta\hat{x}+\sin\theta\hat{x}_\perp]\big)rdrd\theta\\&=k_0\int_0^{2\pi}\int_{r\geq 0}\frac{e^{-\sqrt{r(r+2)}k_0 z+i(r+1)\cos\theta/\varepsilon}}{i\sqrt{r(r+2)}}\hat{U}\big(k_0(r+1)[\cos\theta\hat{x}+\sin\theta\hat{x}_\perp]\big)(r+1)drd\theta.\end{aligned}$$

Above, $\hat{x}=x/|x|$ and $\hat{x}_\perp$ is the vector obtained after a $\pi/2$ rotation of $\hat{x}$. We then set $r\to\varepsilon r$ and neglect lower order terms in ε, and find, after the change of variables $\cos\theta\to u$ in the second line,

$$\begin{aligned}g_{\text{ev}}(x)&\simeq\varepsilon^{1/2}k_0\int_0^{2\pi}\int_{r\geq 0}\frac{e^{i\cos\theta/\varepsilon+ir\cos\theta}}{i\sqrt{2r}}\hat{U}(k_0[\cos\theta\hat{x}+\sin\theta\hat{x}_\perp])drd\theta\\&\simeq\varepsilon^{1/2}k_0\int_{-1}^{1}\int_{r\geq 0}\frac{1}{i\sqrt{2r}\sqrt{1-u^2}}\left[e^{iu/\varepsilon+iru}\hat{U}(k_0[u\hat{x}+\sqrt{1-u^2}\hat{x}_\perp])-cc\right]drdu.\end{aligned}$$

Above, cc means complex conjugate. The integral over r can be performed first, giving

$$\int_0^\infty \frac{e^{iru}}{\sqrt{2r}}dr = \frac{1}{\sqrt{|u|}}F_0(\text{sign}(u)) := \frac{1}{\sqrt{|u|}}\int_0^\infty \frac{e^{ir\text{sign}(u)}}{\sqrt{2r}}dr.$$

After the change of variables $u \to \varepsilon u$ in the second line below and neglecting lower order terms in ε, we finally find

$$\begin{aligned} g_{\text{ev}}(x) &\simeq \varepsilon^{1/2}k_0 \int_{-1}^1 \frac{1}{i\sqrt{|u|}\sqrt{1-u^2}}\Big[F_0(\text{sign}(u))e^{iu/\varepsilon}\hat{U}(k_0[u\hat{x}+\sqrt{1-u^2}\hat{x}_\perp]) - cc\Big]du \\ &\simeq -i\varepsilon k_0[F_1\hat{U}(k_0\hat{x}_\perp) - cc], \end{aligned}$$

where

$$F_1 = \int_{\mathbb{R}} \frac{F_0(\text{sign}(u))e^{iu}}{\sqrt{|u|}}du.$$

Note that an integration by parts shows that F_0 and F_1 are well defined. A similar calculation for the propagative modes yields

$$\begin{aligned} g_{\text{prop}}(x) &= k_0 \int_{|\xi|\le 1} \frac{e^{i\sqrt{1-|\xi|^2}k_0 z + ik_0\xi\cdot x}}{\sqrt{1-|\xi|^2}}\hat{U}(k_0\xi)d\xi \\ &\simeq \varepsilon k_0[F_2\hat{U}(k_0\hat{x}_\perp) - cc], \end{aligned}$$

where F_2 is equal to F_1 except that F_0 is replaced by F_0^*. This shows that evanescent and propagating modes have nearly identical expressions, which is due to the fact that the leading contribution to g comes from the singularity at $|\xi| = k_0$. Hence, when the field $\hat{U}$ has a non-zero contribution on $|\xi| = k_0$, removing evanescent modes creates an error as large as the contribution of the propagating modes in the transverse far field regime where $k_0|x| \gg 1$.

References

[1] M. Chen, D. Ren, H.-Y. Liu, S. Chowdhury, and L. Waller, *Multi-layer born multiple-scattering model for 3d phase microscopy*, Optica, 7 (2020), pp. 394–403.

[2] C. Gasquet and P. Witomski, *Fourier analysis and applications: filtering, numerical computation, wavelets*, vol. 30, Springer Science & Business Media, 2013.

[3] L. N. Trefethen, *Spectral Methods in Matlab*, SIAM, Philadelphia, PA, 2000.