CLOSED FORM DISPERSION CORRECTIONS INCLUDING A REAL SHIFTED WAVE NUMBER FOR FINITE DIFFERENCE DISCRETIZATIONS OF 2D CONSTANT COEFFICIENT HELMHOLTZ PROBLEMS

PIERRE-HENRI COCQUET, MARTIN J. GANDER AND XUESHUANG XIANG

Abstract. All grid-based discretizations of the Helmholtz equation suffer from the so-called pollution effect, which is caused by numerical dispersion: plane waves propagate at the discrete level at speeds which differ from the speed at the continuous level. This leads to poor numerical approximations, unless very fine meshes are used. We propose here a new finite difference scheme to do dispersion correction for the two dimensional Helmholtz equation with constant wavenumber on rectangular domains discretized with finite difference methods on uniform meshes. The main innovations are first the use of a real shift in the wave number used in the discrete scheme, which is different from the continuous wave number, and second, an asymptotic analysis which allows us to determine closed form finite difference schemes without any numerical optimization. We also prove that our asymptotically optimized scheme is sixth order accurate for plane wave solutions. Numerical experiments show that our new optimized scheme has a very small dispersion error when only few points per wavelength are used, and can be effective when solving Helmholtz problems using multigrid.

Key words. Helmholtz equation, Finite difference method, Numerical dispersion, Rectangular domains, Constant wavenumbers.

1. Introduction.

The Helmholtz equation is a model problem for time-harmonic wave propagation. Solving the Helmholtz equation numerically for medium to large wavenumbers is a hard task [15], since the continuous operator is not coercive, the continuous and discrete operators are complex symmetric but not hermitian, and finally the solutions are highly oscillatory. In addition to these difficulties, all grid based discretizations of the Helmholtz problem suffer from the so-called pollution effect [3, 4, 22, 12, 36], because of the numerical dispersion that causes plane waves at the discrete level to have a wavenumber different from the continuous one. Therefore the numerical solution propagates at a speed different from the correct one and, although having the correct magnitude, the numerical solution has a phase lead or lag (see e.g. [1, 19, 16]). In order to control these effects, one either needs to use high order methods, or consider fine enough meshes which lead to very large systems whose size increases in general more than linearly with the wavenumber.

Reducing numerical dispersion, also known as dispersion correction, for Finite Difference (FD) approximations of the Helmholtz equation has already been addressed by several authors. The authors of [23] consider a 9-point stencil with free parameters, which are computed numerically using a steepest descent method to obtain discrete phase curves closest to the continuous one. This leads to a FD method that still works quite well using only 4 grid points per wavelength. In [6] (see [7] for similar results in 3D), a second order 9-point stencil is modified to get pointwise consistent approximations of the Helmholtz problem with cartesian Perfectly Matched Layers (PML). Some free parameters involved in the definition of the stencil are then determined by minimizing the average L^2 error of the difference between the normalized numerical phase velocity and the continuous one. The optimization is done numerically for several values of the angles by considering a least-squares minimization problem. Similar results have been obtained in [8], where the optimal parameters are computed numerically by minimization of the difference between what the authors call numerical wavenumber (see Remark 2.2 for its precise definition) and continuous wavenumber.

Numerical simulations show that the scheme with dispersion correction has slightly less pointwise error than the non-optimized version. Also, for some specific angle and 6 or 12 grid points per wavelength, the optimized scheme has an error that does not grow with the wavenumber when approximating plane wave solution.

Dispersion correction has also been investigated for higher order schemes. In [34], a 9-point stencil with fourth-order accuracy is built and a sixth order one can be found in [35]. It is worth noting that both of these numerical methods are defined for variable wavenumber as well and that their construction is based on some results from [33]. For constant wavenumber, both stencils have free parameters that are later computed using a numerical minimization procedure to reduce the numerical dispersion. However, since the constants in the stencils depend in a non-trivial way on the wavenumber and the meshsize, the numerical wavenumber is not easily computable. As a result, the authors do not compute the free parameters by minimizing the difference between the numerical and continuous wavenumbers and define a refined strategy to determine them. The resulting optimized coefficients are piecewise constant functions of the number of grid points per wavelength. Numerical simulations show that the scheme with dispersion correction has slightly less pointwise error than its non-optimized version. In [28], the α-interpolation method is used on a fourth order finite difference stencil and some rectangular bilinear finite elements. The authors minimize the maximum norm (over all angles) of the relative phase error to get the parameter α and show, through some numerical simulations, that their scheme has a small pollution effet.

There are also some dispersion correction results which do not resort to numerical optimization. For instance, in [20] a high order FD method is designed using Padé approximations (see also [32]) and it is shown that the dispersion can be minimized along grid diagonals in 2 and 3 dimensions. For the one dimensional Helmholtz problem, the authors build their 3-point stencil with a more general definition of the derivative, as one can find also in non-standard FD methods [2], and they show that exact phase representation can be achieved. In [24], centered FD schemes are defined with an undetermined weight that is then computed in order to maximize the accuracy of the second order operator when approximating plane waves. Numerical simulations show that this choice greatly reduces the average error of the scheme on plane wave solutions for any angle.

Another advantage of FD schemes with reduced numerical dispersion is that they enhance the performance of multigrid solvers. In [30], FD schemes with dispersion correction and free parameters are designed. The coefficients depend piecewise polynomially on kh and are obtained by numerical minimization of two separate cost functions. The first one involves the phase error with a regularization term while the second one only considers the amplitude error. The resulting FD scheme is then used in a multigrid method with coarsest mesh having three points per wavelength. Numerical tests on 3-D examples show that this technique, compared to some existing methods from the literature, allow to significantly save computation time. A second order 9-point stencil with free parameters has been considered in [31]. The coefficients are again obtained through the minimization of phase slowness for several angles. Again, the multigrid algorithm with the FD scheme with dispersion correction performs much better than the classical algorithm, both in terms of iteration numbers and computational time. An even earlier result dealing with the performance of the multigrid algorithm for Helmholtz equations when doing dispersion correction can be found in [16]. The authors use, on the fine grid, a real shift on the wavenumber that allows to match the discrete and continuous dispersion relations exactly in 1D. They then use this scheme in a multigrid algorithm and show that the resulting algorithm converges for any number of levels inside a V-cycle and any wavenumber. An extension to the case of piecewise-constant wavenumber that shows similar performance can be found in [9]. These results are actually quite impressive, because with standard FD or finite element (FE) methods, the multigrid algorithm is known to be unusable, unless a complex shifted wavenumber with a large enough imaginary part is used [11, 13, 12]. For one-dimensional Helmholtz equations, using a real shift is actually equivalent to non-standard FD [2], and to design the stencil using a more general definition of the derivative [20, p. 4, Eq. (10)]. The idea of the real shift on the wavenumber has also been investigated in [10]. A second order 9-point stencil with free parameters has been used and the wavenumber has also been considered as a parameter. The shifted wavenumber depends explicitly on the other free parameters. All the remaining coefficients have then been computed by minimizing the average L^2 norm of the phase speed. Numerical simulations indicate that the FD scheme with dispersion correction is actually sixth order accurate for plane wave solutions, while formally being only second order accurate.

In this paper, we study dispersion correction based on the idea of using a real-shift on the wavenumber in the numerical scheme. In addition, we wish to compute the optimized coefficients explicitly without any numerical optimization. To achieve this, we rely on asymptotic dispersion correction. This means that we determine the free parameters of our FD stencil, including a shifted wavenumber $\widetilde{k} = \widetilde{k}(k, h)$, by requiring that its discrete wavenumber approximates the continuous one with best error when the number of grid points per wavelength $G = G(k, h) := 2\pi/(kh)$ goes to infinity[1]. We will see that this approach gives surprisingly good dispersion corrections for G already below ten points per wavelength.

Our paper is organized as follows: first we theoretically study the link between the order of the scheme on plane waves, the approximation order of the continuous wavenumber by its discrete counterpart, and the distance between the two dispersion relations. We prove that these three items are of the same order, which allows us to do asymptotic dispersion correction considering either the distance between the dispersion relations or the distance between discrete and continuous wavenumbers. We consider next a 9-point FD stencil and perform a first asymptotic dispersion correction without a real shift on the wavenumber. This allows us to determine some of the free parameters and to get a fourth order scheme whose discrete wavenumber is then also a fourth order approximation of the continuous one as $G \to +\infty$. The idea of using a shifted wavenumber $\widetilde{k} = \widetilde{k}(k, h)$ is then investigated. This allows us to determine explicitly the remaining coefficients that yield an FD stencil whose discrete wavenumber is shown to be a sixth order approximation of the continuous one as $G \to \infty$. We conclude with some numerical simulations which show that our new dispersion correction including the modified wave number $\widetilde{k}$ can give accurate solutions already for $G = 2.5$ points per wavelength for wave numbers up to about one hundred on the square $(-1, 1)^2$, and the asymptotically optimized coefficients also give a very good dispersion correction for typical engineering values of $G = 10 - 12$ points per wavelength. These numerical experiments involve the computation of the relative error on a model problem as well as the study of the performance of multigrid V and W-cycles.

[1] To simplify the notation, we will not explicitly write the dependence of $\widetilde{k}$ and G on k and h later in the analysis.

2. Approximation error of the discrete wavenumber. We now prove a general result giving a relationship between the accuracy of a FD scheme, the approximation order of the continuous wavenumber by its discrete counterpart, and the distance between the continuous and the discrete dispersion relations. For the FE method, the discrete wavenumber has been computed explicitly in some special cases and is known to approximate the continuous wavenumber with the same accuracy as the numerical method [4, 22, 3, 1]. A similar result holds also for FD schemes (see e.g. [6, Proposition 3.3], [8, Propositon 3.1], [34, Proposition 3.1]). All these results were obtained by direct computations, and no general result seems to be currently available. The computation of the distance between the two dispersion curves is actually directly linked to the goal of this paper, namely doing dispersion correction: having these two curves as close as possible reduces the numerical dispersion.

We consider a FD discretization of the Helmholtz operator $\mathcal{H}u = -(\Delta + k^2)u$, which can be written as

$$(\mathcal{H}(k,h)u)_{i,j} = (-\Delta_h u)_{i,j} - k^2(\mathcal{M}_h u)_{i,j},$$

where h is the meshsize, $\mathcal{M}_h$ is a mass term obtained for instance from a symmetric 9-pt stencil (see e.g. [10, p. 4]) and the subscripts i, j indicate the grid point (x_i, y_j) where the approximation is computed.

REMARK 2.1. *In this paper, we consider only rectangular domains with uniform rectangular meshes with meshsize h, and emphasize that the subscripts i, j are related to interior points only.*

For $\xi = (\xi_1, \xi_2)$ and $\mathbf{x} = (x, y)$, the (discrete) symbol is defined by

$$\sigma(k,h,\xi) {:=} \left(\mathrm{e}^{-\mathrm{i}\mathbf{x}\cdot\xi}\right)_{i,j} \left(\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}\mathbf{x}\cdot\xi}\right)_{i,j}, \tag{2.1}$$

and does not depend on the grid point (x_i, y_j) where it is computed, since the meshes we consider are uniform. The continuous and discrete dispersion relations are then defined by the sets

$$\mathcal{D}_{\mathrm{c}} := \left\{\xi \in \mathbb{R}^2 \mid |\xi|^2 - k^2 = 0\right\}, \qquad \mathcal{D}_h := \left\{\xi \in \mathbb{R}^2 \mid \sigma(k,h,\xi) = 0\right\}, \tag{2.2}$$

and setting $\mathbf{e}(\theta) = (\cos(\theta), \sin(\theta))$, the discrete wavenumber $k_d = k_d(k,h,\theta)$, which depends on the angle θ, is the solution to the equation

$$\sigma(k,h,k_d(k,h,\theta)\mathbf{e}(\theta)) = 0, \tag{2.3}$$

since we want discrete plane waves to satisfy $(\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}k_d\mathbf{x}\cdot\xi})_{i,j} = 0$.

REMARK 2.2. *Some authors call the discrete wavenumber also the numerical wavenumber. Nevertheless, it is worth pointing out that a different definition of the numerical wavenumber is used in [6, 8, 34, 36] where it is defined as k_N satisfying $\sigma(k_N, h, k\mathbf{e}(\theta)) = 0$.*

It is convenient to work with the number of grid-points per wavelength[2] defined as $G{:=}2\pi/(kh)$. Therefore, the discrete wavenumber actually depends on (k, G, θ) and satisfies

$$\sigma\left(k, \frac{2\pi}{kG}, k_d\mathbf{e}(\theta)\right) = 0. \tag{2.4}$$

[2] In this paper, we will both work with h and G. We note that $G \to +\infty$ will mean that $h \to 0$.

The next result shows that the accuracy of the FD scheme on plane wave solutions is the same as the difference between the discrete and continuous wavenumber when $G \to +\infty$. The same result also holds for the distance between the discrete and continuous dispersion curves defined in (2.2).

THEOREM 2.3. *Assume that the discrete symbol is smooth in all its variables for large enough G. Assume also that the first derivatives of the symbol satisfy*

$$\nabla_\xi \sigma \left(k, \frac{2\pi}{kG}, \xi \right) = c_2 \xi + N(k, G, \xi),$$

where c_2 is a constant and $\|N(k,G,\xi)\| \leq \widetilde{c}(\xi) G^{-1}$ where $\xi \in \mathbb{R}^2 \mapsto \widetilde{c}(\xi) > 0$ is continuous. Then the two following statements are equivalent:

i) The FD scheme is accurate of order $q \in \mathbb{N}^*$ for all plane wave solutions, that is

$$\forall\ \theta \in [0, 2\pi],\ (\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)})_{i,j} = O_{G\to\infty}(G^{-q})k^2.$$

ii) The discrete wavenumber admits the asymptotic expansion $k_d = k\,(1 + O_{G\to+\infty}(G^{-q}))$.
If in addition, the discrete dispersion relation admits a polar representation then ii) is also equivalent to

iii) The distance between the discrete and continuous dispersion relations satisfies $\mathrm{dist}(\mathcal{D}_h, \mathcal{D}_\mathrm{c}) = k\, O_{G\to\infty}(G^{-q})$.
Above, the $O_{G\to\infty}(.)$ only depends on θ except in statement iii).

Proof. *ii) ⇒ i)* We obtain by a direct computation that

$$\begin{aligned}(\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)})_{i,j} &= \left(\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)}\right)_{i,j} \left(\mathrm{e}^{-\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)}\right)_{i,j} (\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)})_{i,j} \\ &= \left(\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)}\right)_{i,j} (\sigma(k,h,k\mathbf{e}(\theta)) - \sigma(k,h,k_d\mathbf{e}(\theta)))) \\ &= \mathrm{e}^{\mathrm{i}k(x_i\cos(\theta)+y_j\sin(\theta))} \left(\int_0^1 \nabla_\xi \sigma(k,h,k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d)) \cdot (k-k_d)\mathbf{e}(\theta) ds \right),\end{aligned}$$

where we used that k_d satisfies (2.4). Using now the assumption on the derivative of the discrete symbol with respect to ξ, and that $G = 2\pi/(kh)$, we get

$$\begin{aligned}&\left| \int_0^1 \nabla_\xi \sigma(k,h,k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d)) \cdot (k-k_d)\mathbf{e}(\theta) ds \right| \\ &\leq |c_2||k-k_d| \int_0^1 \|k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d)\|\, ds + |k-k_d| \int_0^1 \|N(k,G,k_d\mathbf{e}(\theta)+s\mathbf{e}(\theta)(k-k_d))\|\, ds \\ &\leq |c_2|k_d|k-k_d| + |k-k_d| \left(|c_2||k-k_d| + G^{-1} \int_0^1 \widetilde{c}(k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d)))ds \right).\end{aligned}$$

Because of the continuity of $\xi \mapsto \widetilde{c}(\xi)$, we have

$$\lim_{G\to+\infty} \int_0^1 \widetilde{c}(k\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d)))ds = \widetilde{c}(k\mathbf{e}(\theta)),$$

and thus $(\mathcal{H}(k,h)\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)})_{i,j} = k^2\, O_{G\to+\infty}(G^{-q})$ which yields the desired result.

$\underline{i) \Rightarrow ii)}$ Recall that the discrete wavenumber satisfies (2.4). Since $\mathcal{H}(k,h)$ is a FD discretization of the Helmholtz operator, the discrete symbol converges, as $h \to 0$, to the symbol of the continuous operator. This gives

$$\lim_{h\to 0} \sigma(k,h,\xi) = |\xi|^2 - k^2. \tag{2.5}$$

As a result, we have $0 = \lim_{h\to 0} \sigma(k,h,k_d\mathbf{e}(\theta)) = -k^2 + \lim_{h\to 0} k_d(k,h,\theta)^2$, which shows that the zeroth-order term in the expansion of k_d is $\lim_{h\to 0} k_d(k,h,\theta) = k$. Note that $G \to +\infty$ as h goes to zero and thus we have also proved that $\lim_{G\to+\infty} k_d(k,G,\theta) = k$. From $i)$, we get for large enough G that

$$\mathrm{e}^{-\mathrm{i}k(x_i\cos(\theta)+y_j\sin(\theta))} \left(\mathcal{H}\left(k, \frac{2\pi}{kG}\right)\mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)}\right)_{i,j} = \sigma\left(k, \frac{2\pi}{kG}, k\mathbf{e}(\theta)\right) = O_{G\to+\infty}(G^{-q})k^2.$$

Since $\sigma(k,h,k_d\mathbf{e}(\theta)) = 0$, $\sigma\left(k, \frac{2\pi}{kG}, k\mathbf{e}(\theta)\right) - \sigma\left(k, \frac{2\pi}{kG}, k_d\mathbf{e}(\theta)\right) = O_{G\to+\infty}(G^{-q})k^2$, and a Taylor expansion gives

$$\begin{aligned} O_{G\to+\infty}(G^{-q})k^2 &= (k-k_d)\left(\int_0^1 \nabla_\xi \sigma(k, \frac{2\pi}{kG}, k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d))\cdot\mathbf{e}(\theta)ds\right) \\ &= (k-k_d)\int_0^1 F(k,G,k_d,\mathbf{e}(\theta),s)ds. \end{aligned}$$

We thus get that

$$(k-k_d) = \frac{O_{G\to+\infty}(G^{-q})k^2}{\int_0^1 F(k,G,k_d,\mathbf{e}(\theta),s)ds}, \tag{2.6}$$

and the assumption on $\nabla_\xi\sigma$ gives

$$\begin{aligned} \int_0^1 F(k,G,k_d,\mathbf{e}(\theta),s)ds &= \int_0^1 c_2 k_d + s(k-k_d)ds \\ &\quad + \int_0^1 N(k,G,k_d\mathbf{e}(\theta) + s\mathbf{e}(\theta)(k-k_d))\cdot\mathbf{e}(\theta)ds \\ &= c_2 k + o(1) + O(G^{-1}) = c_2 k + o(1), \end{aligned}$$

where we used the bound on N, the continuity of $\xi \mapsto \widetilde{c}(\xi)$ and that $k_d \to k$ as $G \to +\infty$. Using this in (2.6), we have

$$k - k_d = \frac{O_{G\to+\infty}(G^{-q})k^2}{c_2 k + o(1)} = O_{G\to+\infty}(G^{-q})k,$$

which gives the desired estimate.

$\underline{ii) \Leftrightarrow iii)}$ Because of the additional assumption, for any $\theta \in [0,2\pi]$, there is a unique $\xi(\theta) \in \mathcal{D}_h$ such that $\xi(\theta) = \|\xi(\theta)\|\,\mathbf{e}(\theta)$. Since the discrete wavenumber is a solution to (2.4) we obtain that $k_d(\theta) = \|\xi(\theta)\|$ and that

$$\begin{aligned} \operatorname{dist}(\mathcal{D}_h, \mathcal{D}_\mathrm{c}) &= \max\left\{\|\xi_h - \xi_\mathrm{c}\|,\, \xi_h \in \mathcal{D}_h \text{ and } \xi_\mathrm{c} \in \mathcal{D}_\mathrm{c}\right\} \\ &= \max_{\theta\in[0,2\pi]} \|\xi(\theta) - k\mathbf{e}(\theta)\| = \max_{\theta\in[0,2\pi]} \|k_d(\theta)\mathbf{e}(\theta) - k\mathbf{e}(\theta)\| \\ &= \max_{\theta\in[0,2\pi]} |k_d(\theta) - k|. \end{aligned}$$

The equivalence between statements *ii*) and *iii*) then easily follows. □

In the proof, we used (2.5) which means that the discrete symbol converges to the continuous one, namely $(|\xi|^2 - k^2)$, as $h \to 0$ (or $G \to +\infty$). If the discrete symbol is real-analytic with respect to $(k, \xi) \in \mathbb{R}_+^* \times \mathbb{R}^2$ one gets that the derivatives of the discrete symbol converge to those of the continuous symbol as $h \to 0$. This gives

$$\lim_{h \to 0} \nabla_\xi \sigma \left(k, h, \xi\right) = 2\xi,$$

and thus $c_2 = 2$. Nevertheless, in practice, it is easy to check that the standing assumption of Theorem 2.3 is satisfied for the 5 or 9-point stencil and by some FD methods with coefficients depending on $kh = 2\pi/G$ as those from [20, 32, 34, 35].

3. Asymptotic dispersion correction for the 9-point stencil.

We consider a FD discretization of the two-dimensional Helmholtz equation on a uniform grid with meshsize $h > 0$ given by a 9-point stencil. The Helmholtz operator at the discrete level is then defined by

$$\begin{aligned}
&\left(\mathcal{H}_h v\right)_{i,j} := \left(\frac{4a}{h^2} - k^2 b\right) v(x_i, y_j) \\
&+ \left(\frac{1-2a}{h^2} - \frac{k^2 c}{4}\right) \left(v(x_{i-1}, y_j) + v(x_{i+1}, y_j) + v(x_i, y_{j-1}) + v(x_i, y_{j+1})\right). \qquad (3.1)\\
&- \left(\frac{1-a}{h^2} + k^2 \frac{1-b-c}{4}\right) \left(v(x_{i-1}, y_{j-1}) + v(x_{i+1}, y_{j-1}) + v(x_{i-1}, y_{j+1}) + v(x_{i+1}, y_{j+1})\right),
\end{aligned}$$

where a, b, c are some positive constants. The above numerical scheme is second order accurate for any a, b, c. In this section, we are interested in having the most accurate discrete wavenumber as $G \to +\infty$ for our asymptotic dispersion correction. According to Theorem 2.3, the discrete wave number accuracy is directly linked to the accuracy of $\mathcal{H}_h$ on plane wave solutions which can be of order four for a suitable choice of the constants as we now show.

THEOREM 3.1. *Assume that u is a solution to the homogeneous Helmholtz equation $\mathcal{H}u = 0$ in a neighborhood $\mathcal{O}$ of (x_i, y_j) such that $(x_i, y_j) + h\mathbf{v} \in \mathcal{O}$ for any $\|\mathbf{v}\| = 1$. Then for*

$$a = \frac{5}{6}, \ \frac{c}{4} + \frac{b}{2} - \frac{5}{12} = 0,$$

we have $\left(\mathcal{H}_h u\right)_{i,j} = O_{h\to 0}(h^4)$. In addition, for plane-wave solutions, we have that $\left(\mathcal{H}_h \mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)}\right)_{i,j} = k^2 O_{G\to+\infty}(G^{-4})$, where this $O_{G\to+\infty}(.)$ only depends on θ.

Proof. Since the solution to $\mathcal{H}u = 0$ is smooth on $\mathcal{O}$, a Taylor expansion[3] gives

$$\begin{aligned}
\left(\mathcal{H}_h u\right)_{i,j} &= O_{h\to 0}(h^4) - k^2 u(x_i, y_j) - \Delta u(x_i, y_j) \qquad (3.2)\\
&+ h^2 \left(-\frac{1}{12}\partial_x^4 - \frac{1}{12}\partial_y^4 + \frac{ck^2}{4}\Delta + (a-1)\partial_x^2\partial_y^2 + (b-1)\frac{k^2}{2}\Delta\right) u(x_i, y_j)
\end{aligned}$$

We have $(\partial_{x^4}^4 + \partial_{y^4}^4)u = (-2\partial_x^2\partial_y^2 + \Delta^2)u$ since u is smooth. The second order term in (3.2) is then

$$h^2 \left(-\frac{1}{12}\Delta^2 + \left(\frac{1}{6} - 1 + a\right)\partial_{x^2y^2}^4 + k^2\left(\frac{c}{4} + \frac{b}{2} - \frac{1}{2}\right)\Delta\right) u(x_i, y_j). \qquad (3.3)$$

[3] Symbolic-type computations can be used for this, see Section 7.

Using now that $\mathcal{H}u = 0$, we obtain $\Delta \mathcal{H} u = 0$ from which we infer that $\Delta^2 u = -k^2 \Delta u$ and thus (3.3) becomes

$$h^2 \left(\left(-\frac{5}{6} + a \right) \partial^4_{x^2 y^2} + k^2 \left(\frac{c}{4} + \frac{b}{2} - \frac{1}{2} + \frac{1}{12} \right) \Delta \right) u(x_i, y_j).$$

Finally, chosing a, b, c as above gives that $(\mathcal{H}_h u)_{i,j} = O_{h \to 0}(h^4)$.

Direct computations with these parameters then yield

$$\begin{aligned}\left(\mathcal{H}_h \mathrm{e}^{\mathrm{i}k\mathbf{x}\cdot\mathbf{e}(\theta)} \right)_{i,j} &= -\frac{\mathrm{e}^{\mathrm{i}k(x_i,y_j)\cdot\mathbf{e}(\theta)}}{720} h^4 k^6 \left(3 + (90c - 16)(\cos(\theta)^4 - \cos(\theta)^2) \right) \\ &\quad + k^2 O((kh)^6), \end{aligned} \tag{3.4}$$

which proves the second statement of the theorem. □

Remark 3.2. *The previous theorem is only valid for homogeneous or harmonic right hand side which is enough for the goals of this paper, that is to build FD schemes with reduced numerical dispersion. The case of inhomogeneous right-hand-side can be handled by modifying the discrete right hand side as in [34, 35]. For our 9-point stencil, it is enough to replace the usual $f_{i,j}$ source term with the modified stencil $f_{i,j} + \frac{h^2}{12}(\Delta_h f)_{i,j}$, where $(\Delta_h f)_{i,j}$ denotes the approximation of the Laplace operator with the standard 5-point stencil.*

The discrete wavenumber k_d is defined by (2.4) where one can replace $h = 2\pi/(kG)$ if one wishes to have its asymptotics as the number of grid points goes to infinity. From Theorems 3.1 and 2.3 we get that the discrete wavenumber satisfies the asymptotic expansion $k_d = k + k\,O_{G \to +\infty}(G^{-4})$, where the $O(.)$ only depends on θ.

4. Asymptotic dispersion correction using a real-shift for the wavenumber.

We show in this section how the 9-point FD scheme (3.1) with a fourth order discrete wavenumber can be improved to get a sixth order discrete wavenumber. To reach this goal, we follow an idea from [16] which uses a different wavenumber at the discrete level to get a numerical scheme whose dispersion relation equals the continuous one in 1D. Since such exact dispersion correction can only be obtained in 1D, we are going to use the free parameters c and $\widetilde{k}$ to minimize the dispersion error as the number of grid points per wavelength goes to infinity.

Our approach is rather geometric since we wish to find $(\widetilde{k}, c)$ so that the discrete dispersion relation is as close as possible to the continuous one when $G = 2\pi/(kh)$ goes to $+\infty$. Because of Theorem 2.3, this will yield a discrete wavenumber closer to the continuous one. According to (2.2), the two sets of interest are defined by

$$\begin{aligned} \mathcal{D}_{\mathrm{c}} &:= \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \mid \xi_1^2 + \xi_2^2 - k^2 = 0 \right\}, \\ \widetilde{\mathcal{D}}_h &:= \{ (\xi_1, \xi_2) \in \mathbb{R}^2 \mid \begin{array}{l} (4ah^{-2} - \widetilde{k}^2 b) + 2(\frac{1-2a}{h^2} - \frac{\widetilde{k}^2 c}{4})(\cos(h\xi_1) + \cos(h\xi_2)) \\ -2(\frac{1-a}{h^2} + \widetilde{k}^2 \frac{1-b-c}{4})(\cos(h(\xi_1 + \xi_2)) + \cos(h(\xi_1 - \xi_2))) = 0 \}. \end{array} \end{aligned}$$

Due to the symmetries of the discrete dispersion relation, it is enough to work only in the upper right quarter plane $\xi_1 > 0$, $\xi_2 > 0$. Therefore we compute, for a given angle $\theta \in [0, \pi/2)$, the distance between the origin and the intersection of $\widetilde{\mathcal{D}}_h$ with the line $\xi_2 = \tan(\theta)\xi_1$. The case $\theta = \pi/2$ will be done later. We also set

$$\mathcal{L}(\theta) := \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 \mid \xi_1 > 0, \xi_2 > 0, \ \xi_2 = \tan(\theta)\xi_1 \right\}.$$

The next theorem gives the distance, as G goes to ∞, between the origin and the discrete dispersion relation for a given angle.

THEOREM 4.1. *Let* $\theta \in [0, \pi/2)$, $(x(\theta), y(\theta)) \in \widetilde{\mathcal{D}}_h \cap \mathcal{L}(\theta)$ *and* $\mathrm{d}(\theta) = \sqrt{x(\theta)^2 + y(\theta)^2}$. *Assume that* $\widetilde{k}_h = k_0 + k_1 G^{-1} + \cdots + k_6 G^{-6}$ *and* $c = c_0 + c_1 G^{-1} + c_2 G^{-2}$. *Then we have the asymptotic expansions* $\widetilde{k} = k - \frac{\pi^4 k}{30} G^{-4} + k_6 G^{-6}$, $c = \frac{8}{45} + c_2 G^{-2}$ *and*

$$\mathrm{d}(\theta) = k + k d_6(\theta, k_6, c_2) G^{-6} + O(G^{-7}),$$

with

$$(k_6, c_2) = \left(-\frac{\pi^2 k}{192}, -\frac{\pi^2}{54}\right) = \arg\min_{k_6, c_2} \left(\max_{\theta}(|d_6(\theta, k_6, c_2)|)\right).$$

Proof. Note first that $(x(\theta), y(\theta))$ satisfies the set of equations

$$\begin{aligned}
&x(\theta) > 0, y(\theta) > 0 \text{ and } y(\theta) = \tan(\theta) x(\theta),\\
&F(\widetilde{k}, c, x(\theta), y(\theta), G) :=\\
&(4aG^2/\pi^2 - \widetilde{k}^2 b) + 2(\tfrac{1-2a}{\pi^2} G^2 - \tfrac{\widetilde{k}^2 c}{4})(\cos(x(\theta)\tfrac{2\pi}{kG}) + \cos(y(\theta)\tfrac{2\pi}{kG}))\\
&-2(\tfrac{1-a}{\pi^2} G^2 + \widetilde{k}^2 \tfrac{1-b-c}{4})(\cos((x(\theta) + y(\theta))\tfrac{2\pi}{kG}) + \cos((x(\theta) - y(\theta)))\tfrac{2\pi}{kG}) = 0,
\end{aligned}$$

and is uniquely defined. Since the equation $F = 0$ does not have explicit solutions, we use asymptotic analysis, assuming that $x(\theta)$ can be written as

$$x(\theta) = \sum_{j=0}^{6} x_j G^{-j} + O_{G\to\infty}(G^{-7}). \tag{4.1}$$

Inserting (4.1) into F and doing some lengthy but not difficult computations, we get

$$\begin{aligned}
F(\widetilde{k}, c, x(\theta), y(\theta), G) &= \frac{1}{3} \frac{(-3\cos(\theta)^6 k^4 \widetilde{k}^2 + 3\cos(\theta)^4 k^4 x_0^2)}{k^4 \cos(\theta)^6} + \frac{1}{G}\left(\frac{2x_0 x_1}{\cos(\theta)^2}\right)\\
&+ \frac{1}{3G^2}\left(\frac{\widetilde{k}^2 \pi^2 x_0^2 \cos(\theta)^2 + 6x_2 x_0 k^2 \cos(\theta)^2 + 3x_1^2 k^2 \cos(\theta)^2 - \pi^2 x_0^4)}{k^2 \cos(\theta)^4}\right)\\
&+ \frac{1}{3G^3} \frac{2\widetilde{k}^2 \pi^2 x_1 x_0 \cos(\theta)^2 + 6x_3 x_0 k^2 \cos(\theta)^2 + 6x_2 x_1 k^2 \cos(\theta)^2 - 4\pi^2 x_1 x_0^3}{k^2 \cos(\theta)^4}\\
&+ O_{G\to\infty}(G^{-4}),
\end{aligned}$$

from which we infer the first coefficients of $x(\theta)$ and $\widetilde{k}$

$$\lim_{G\to+\infty} \widetilde{k} = k, \quad x_0 = k\cos(\theta), \quad x_1 = 0, \quad x_2 = 0 \quad \text{and} \quad x_3 = 0.$$

Continuing the asymptotic expansion of F as $G \to +\infty$ and using the above constants, we obtain

$$\begin{aligned}
&F(\widetilde{k}, c, x(\theta), y(\theta), G) = (k^2 - \widetilde{k}^2) + \frac{\pi^2}{3G^2}(-k^3 + k\widetilde{k}^2)\\
&+\frac{\widetilde{k}^2\pi^4}{945G^4}(c(-1890\cos(\theta)^6 + 420\cos(\theta)^2 + 1890)\cos(\theta)^2 + 420\cos(\theta)^2(\cos(\theta)^2 - 1) - 105)\\
&+\frac{1}{945G^4\cos(\theta)}(-84\pi^4\cos(\theta)^5k^3 + 84\pi^4\cos(\theta)^3k^3 + 42\pi^4\cos(\theta)k^3 + 1890k^2x_4) + \frac{2kx_5}{\cos(\theta)G^5}\\
&+\frac{2c\widetilde{k}\pi^6}{3G^6}(\cos(\theta)^4 - \cos(\theta)^2)\\
&+\frac{\widetilde{k}^2}{945k\cos(\theta)G^6}(-168\pi^6\cos(\theta)^5k + 168\pi^6\cos(\theta)^3k + 14\pi^6\cos(\theta)k + 630\pi^2x_4)\\
&+\frac{1}{945k\cos(\theta)G^6}(-20\pi^6\frac{\widetilde{k}^2}{945k\cos(\theta)G^6}k^3 + 40\pi^6\cos(\theta)^7k^3 - 4\pi^6\cos(\theta)^5k^3\\
&-16\pi^6\cos(\theta)^3k^3 - 3\pi^6\cos(\theta)k^3 - 1260\pi^2k^2x_4 + 1890k^2x_6)\\
&+\frac{x_5\pi^2}{945k\cos(\theta)G^7}(-1260k^2 + 630\widetilde{k}^2) + kO(G^{-8}),
\end{aligned}$$

which gives $x_5 = 0$. We now assume that the shifted wavenumber $\widetilde{k}$ and c can be written as

$$\widetilde{k} = \sum_{j=0}^{6} k_j G^{-j} + O_{G\to\infty}(G^{-7}), \qquad c = \sum_{j=0}^{2} c_j G^{-j} + O_{G\to\infty}(G^{-3}), \tag{4.2}$$

where the k_j and c_j do not depend on θ and $k_0 = \lim_{G\to+\infty} k_d = k$. Note that we expand c only up to order 2 since it appears at order 4. Inserting (4.2) into the asymptotic expansion of F, we find

$$\begin{aligned}
F(\widetilde{k}, c, x(\theta), y(\theta), G) = (k - k_0) - 2\frac{k_1k}{G}&\\
&+\frac{1}{45G^2}(-15\pi^2k^2 + 15\pi^2k_0^2 - 90k_0k_2 - 45k_1^2)\\
&+\frac{1}{45G^3}(30\pi^2k_0k_1 - 90k_0k_3 - 90k_1k_2) + k\,O_{G\to\infty}(G^{-4}),
\end{aligned}$$

which gives $k_0 = k$, $k_1 = 0$, $k_2 = 0$ and $k_3 = 0$. Continuing the asymptotic expansion of F yields

$$\begin{aligned}
&F(\widetilde{k}, c, x(\theta), y(\theta), G) = k\,O_{G\to\infty}(G^{-8})\\
&+\frac{k}{\cos(45\theta)G^4}(90\pi^4\cos(\theta)^5c_0k - 16\pi^4\cos(\theta)^5k - 90\cos(\theta)^3c_0k + 16\pi^4\cos(\theta)^3k\\
&+3\pi^4\cos(\theta) + 90k_4\cos(\theta) - 90x_4)\\
&-\frac{2k}{G^5}(\pi^4\cos(\theta)^4c_1k - 1890\pi^4\cos(\theta)^2c_1k + k_5\cos(\theta))\\
&-\frac{k}{945\cos(\theta)G^6}(20\pi^6\cos(\theta)^9k - 40\pi^6\cos(\theta)^7k - 630\pi^6\cos(\theta)^5c_0k - 1890\pi^4\cos(\theta)^3c_2k\\
&+172\pi^6\cos(\theta)^5k + 630\pi^6\cos(\theta)^3c_0k + 1890\pi^4\cos(\theta)^5c_2k - 152\pi^6\cos(\theta)^3k\\
&-11\pi^6k\cos(\theta) - 630k_4\pi^2\cos(\theta) + 630x_4\pi^2x_4 + 1890k_6\cos(\theta) - 1890x_6).
\end{aligned}$$

This finally gives that if

$$k_4 = -\frac{\pi^4 k}{30}, \quad k_1 = 0, \quad c_0 = \frac{8}{45} \quad \text{and} \quad c_1 = 0,$$

then $x_4 = 0$ and x_6 can be determined to have $F = O(G^{-8})$. The asymptotic expansion of $x(\theta)$ is then

$$\begin{aligned} x(\theta) &= k\cos(\theta) + O(G^{-7}) \\ &+ G^{-6}\frac{k\cos(\theta)}{189}(2\cos(\theta)^8\pi^6 - 4\cos(\theta)^6\pi^6 + 6\cos(\theta)^4\pi^6 \qquad (4.3) \\ &+ 189\cos(\theta)^4\pi^4 c_2 - 4\cos(\theta)^2\pi^6 - 189\cos(\theta)^2\pi^4 c_2 + \pi^6 + 189k_6/k) \\ &= k\cos(\theta) + k\cos(\theta)X_6(\theta, k_6, c_2)G^{-6} + O(G^{-7}), \qquad (4.4) \end{aligned}$$

where one can determine the expression of X_6 easily by identification.

The distance between the discrete and continuous dispersion relations then has for G large the expansion

$$\begin{aligned} \mathrm{d}(\theta) &= \sqrt{x(\theta)^2 + y(\theta)^2} = \sqrt{1 + \tan(\theta)^2}\,x(\theta) \\ &= k + kG^{-6}\sqrt{1 + \tan(\theta)^2}\cos(\theta)X_6(\theta, k_6, c_2) + O(G^{-7}), \end{aligned}$$

which gives $d_6(\theta, k_6, c_2) = X_6(\theta, k_6, c_2)$. We now wish to determine (c_2, k_6) such that $\max_{\theta\in[0,\pi/2]}|d_6(\theta, k_6, c_2)|$ is minimal since we want these parameters independent of θ. Therefore, we have to solve the min-max problem

$$\min_{c_2, k_6}\max_{\theta\in[0,\pi/2]}|d_6(\theta, k_6, c_2)|. \qquad (4.5)$$

First, to simplify the overall computations, we define $H(Y, k_6, c_2) = d_6(\arccos(Y), k_6, c_2)$ that is

$$H(Y, k_6, c_2) = \frac{2}{189}Y^8\pi^6 - \frac{4}{189}Y^6\pi^6 + \frac{2}{63}\pi^6 Y^4 + c_2 Y^4\pi^4 - \frac{4}{189}\pi^6 Y^2 - \pi^4 Y^2 c_2 + \frac{1}{189}\pi^6 + \frac{k_6}{k}.$$

Now note that $\max_{\theta\in[0,\pi/2]}|d_6(\theta, k_6, c_2)| = \max_{Y\in[-1,1]}|H(Y, k_6, c_2)|$, and thus solving the min-max problem (4.5) is equivalent to solving the min-max problem associated with H. We start by computing the critical points of H, which satisfy the polynomial equation

$$0 = \partial_Y H = \frac{16}{189}\pi^6 Y^7 - \frac{8}{63}Y^5\pi^6 + \left(\frac{8}{63}\pi^6 + 4\pi^4 c_2\right)Y^3 - \left(\frac{8}{189}\pi^6 + 2\pi^4 c_2\right)Y,$$

whose solutions are

$$Y_1 = 0,\ Y_2 = \frac{\pm 1}{\sqrt{2}},\ Y_3^{\pm,\pm} = \pm\frac{1}{\sqrt{2}\pi}\sqrt{\pi^2 \pm \sqrt{-3\pi^4 - 189\pi^2 c_2}},$$

where there are four different Y_3. Since $H(-Y) = H(Y)$, we get that

$$\begin{aligned} \max_{Y\in[-1,+1)}|H(Y, k_6, c_2)| &= \max\left\{|H(Y_1)|, |H(Y_2^{\pm})|, |H(Y_3^{\pm,\pm})|\right\} \\ &= \max\{\left|\frac{\pi^6}{189} + \frac{k_6}{k}\right|, \left|-\frac{\pi^6}{189} - \pi^4 c_2 - \frac{189}{8}\pi^2 c_2^2 + \frac{k_6}{k}\right| \\ &\quad, \left|\frac{\pi^6}{1512} - \frac{\pi^4}{4}c_2 + \frac{k_6}{k}\right|\} \\ &= \max\left\{|R_1(k_6/k, c_2)|, |R_2(k_6/k, c_2)|, |R_3(k_6/k, c_2)|\right\}. \end{aligned}$$

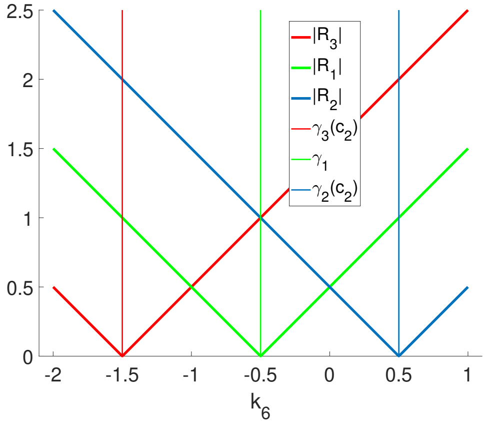

FIG. 4.1. *Representation of* $|R_j(\widetilde{k_6}, c_2)|$ *for a fixed* c_2.

Note that the functions $k_6 \mapsto R_j$ only depend on k_6/k. As a result, denoting by $\widetilde{k_6} := k_6/k$, problem (4.5) reduces to find $(c_2, \widetilde{k_6})$ that satisfy

$$\min_{c_2, \widetilde{k_6}} \max \left\{ |R_1(\widetilde{k_6}, c_2)|, |R_2(\widetilde{k_6}, c_2)|, |R_3(\widetilde{k_6}, c_2)| \right\}. \tag{4.6}$$

We are going to work first with fixed c_2, hence we wish to find $\widetilde{k_6}(c_2)$ that minimizes the $\max_j\{|R_j(\widetilde{k_6}, c_2)|\}$ for each c_2. The solution to problem (4.6) will then be obtained by computing the minimum of the real function $c_2 \mapsto \max_j\{|R_j(\widetilde{k_6}(c_2), c_2)|\}$. We introduce the useful notation

$$R_1 := -\gamma_1 + \widetilde{k_6},\ R_2 := -\gamma_2(c_2) + \widetilde{k_6},\ R_3 := -\gamma_3(c_2) + \widetilde{k_6}.$$

Note that for any c_2, the functions $\widetilde{k_6} \mapsto R(\widetilde{k_6}, c_2)$ are affine with the same slope. As seen from Figure 4.1, the optimal value $\widetilde{k_6}(c_2)$ that minimizes $\max_j\{|R_j(\widetilde{k_6}, c_2)|\}$ for each c_2 is reached when

$$R_{\mathrm{l}}(\widetilde{k_6}, c_2) = -R_{\mathrm{r}}(\widetilde{k_6}, c_2), \tag{4.7}$$

where the subscripts l, r stand for *left* and *right* and $R_{\mathrm{l}} = R_j$ if $\gamma_j(c_2) < \gamma_n(c_2)$ for $n \neq j$, and $R_{\mathrm{r}} = R_j$ if $\gamma_j(c_2) > \gamma_n(c_2)$ for $n \neq j$. It can be observed by direct calculation that $\gamma_2(c_2) \geq \gamma_3(c_2)$ and $\gamma_2(c_2) \geq \gamma_1$ for any c_2 and thus (4.7) reduces to the two following cases:

a) $\gamma_3 \leq \gamma_1 \leq \gamma_2$ that is $R_3(\widetilde{k_6}, c_2) = -R_2(\widetilde{k_6}, c_2)$. This gives $\widetilde{k_6}^{(1)}(c_2) = \frac{\pi^6}{432} + \frac{5}{8}\pi^4 c_2 + \frac{189}{16}\pi^2 c_2^2$.

b) $\gamma_1 \leq \gamma_2 \leq \gamma_3$ that is $R_1(\widetilde{k_6}, c_2) = -R_2(\widetilde{k_6}, c_2)$. This gives $\widetilde{k_6}^{(2)}(c_2) = \frac{\pi^4}{2} c_2 + \frac{189}{16}\pi^2 c_2^2$.

The optimization problem (4.6) is then equivalent to solve

$$\begin{aligned}
&\min_{c_2} \max \left\{ -R_2(\widetilde{k_6}^{(1)}(c_2), c_2), -R_2(\widetilde{k_6}^{(2)}(c_2), c_2) \right\} \\
&= \min_{c_2} \max \left\{ \frac{\pi^6}{336} + \frac{3\pi^4}{8} c_2 + \frac{189}{16}\pi^2 c_2^2, \frac{\pi^6}{189} + \frac{\pi^4}{2} c_2 + \frac{189}{16}\pi^2 c_2^2 \right\}.
\end{aligned}$$

Since both $c_2 \mapsto -R_2(\widetilde{k}_6^{(j)}(c_2), c_2)$ are positive second order polynomials, the solution to the previous min-max problem is reached when $R_2(\widetilde{k}_6^{(1)}(c_2), c_2) = R_2(\widetilde{k}_6^{(2)}(c_2), c_2)$, which gives $c_2 = -\frac{\pi^2}{54}$. Computing the value of $\widetilde{k}_6^{(j)}(-\pi^2/54) = -\pi^6/192$ finally gives that the solution to problem (4.5) is reached at $c_2 = -\frac{\pi^2}{54}$, $k_6 = k\widetilde{k}_6 = -k\frac{\pi^6}{192}$, and the proof is therefore complete. □

Note that we excluded the case $\theta = \pi/2$ in Theorem 4.1 since this would imply division by zero in the asymptotic expansion. Actually, this case can be explicitly computed: we find

$$x\left(\frac{\pi}{2}\right) = 0,\ y\left(\frac{\pi}{2}\right) = \frac{kG}{2\pi}\arccos\left(\frac{3G^2k^2 - 5\pi^2\widetilde{k}^2}{3G^2k^2 + \pi^2\widetilde{k}^2}\right),$$

and it is then easy to see that under the assumptions of Theorem 4.1, one has

$$\mathrm{d}\left(\frac{\pi}{2}\right) = k + k\,O_{G\to\infty}(G^{-6}), \tag{4.8}$$

and that the sixth order term of $\mathrm{d}(\pi/2)$ is also minimal for (c_2, k_6) as given above. The distance between the discrete and continuous dispersion relation can now be computed as $G \to +\infty$. First note that $\mathcal{D}_{\mathrm{c}} \cap \mathcal{L}(\theta) = \{k\mathbf{e}(\theta)\}$. Using now the symmetries of the discrete and continuous dispersion relations and Theorem 4.1 together with (4.8), we have that

$$\mathrm{dist}(\mathcal{D}_{\mathrm{c}}, \widetilde{\mathcal{D}}_h) = \max_{\theta\in[0,\pi/2]} |\mathrm{d}(\theta) - k| = k\,O_{G\to\infty}(G^{-6}).$$

We define now a new 9-point FD method $\mathcal{H}_h^{\mathrm{asympt}}$ given by (3.1) with k replaced by $\widetilde{k}$. The discrete wavenumber associated to the FD operator $\mathcal{H}_h^{\mathrm{asympt}}$ is defined as in (2.4). Its asymptotic behavior as $G \to +\infty$ can be computed using Theorems 2.3 and 4.1 from which we infer the next result.

THEOREM 4.2. *Assume that* $(a, b, c^{\mathrm{asympt}})$ *are as in Theorem 3.1 and that* $\widetilde{k}^{\mathrm{asympt}}$ *and* c^{asympt} *are as given by Theorem 4.1,* $\widetilde{k}^{\mathrm{asympt}} = k - \frac{\pi^4 k}{30}G^{-4} - k\frac{\pi^6}{192}G^{-6}$, $c^{\mathrm{asympt}} = \frac{8}{45} - \frac{\pi^2}{54}G^{-2}$. *Then the discrete wavenumber* $\widetilde{k}_{\mathrm{d}}^{\mathrm{asympt}}$ *associated to the FD stencil* $\mathcal{H}^{\mathrm{asympt}}$ *satisfies* $\widetilde{k}_{\mathrm{d}}^{\mathrm{asympt}} = k + k\,O_{G\to\infty}(G^{-6})$, *where the* $O_{G\to\infty}(.)$ *only depends on* θ.

We emphasize that the combination of the results from Theorem 2.3 and Theorem 4.2 ensures that the FD scheme $\mathcal{H}_h^{\mathrm{asympt}}$ defined in 4.2 is sixth-order accurate on plane-wave solution. Also, we now have a numerical method whose discrete wavenumber is, due to the real shift $\widetilde{k}$, a sixth order approximation of k while being only fourth order without it.

5. Numerical experiments. So far, we have been interested in asymptotic dispersion correction, that is as G goes to infinity. In this section we would like to see if our FD scheme $\mathcal{H}^{\mathrm{asympt}}$ can also reduce numerical dispersion for smaller values of G. In order to reach this goal, we are going to compute first the values of G for which the dispersion relation is non-empty, and also for which it becomes disconnected, since then the dispersion correction can no longer be efficient. We will next optimize numerically the parameters $\widetilde{k}$ and c by minimizing the distance and then compare our formulas with the numerically optimized parameters. Next, we compute and compare the distance between the discrete and continuous dispersion relation for several FD methods from the literature which use dispersion correction or not. We then solve a Helmholtz equation on the square $(-1,1)\times(-1,1)$ with inhomogeneous Dirichlet boundary conditions whose exact solution is a plane wave to show the interest of using FD schemes with dispersion correction. We end this section by showing numerically that using dispersion correction in a multigrid algorithm yields a convergent algorithm.

5.1. Geometric properties of the discrete dispersion relation. We emphasize that dispersion correction is done by minimizing the distance between the discrete and continuous dispersion relations. Also, as pointed out in the introduction, FD schemes with dispersion correction can enhance the performance of the multigrid algorithm. It is therefore of interest to know for which values of k and h the discrete dispersion relation $\mathcal{D}_h$ is non-empty. We are also going to look for which values of k and h the discrete dispersion relation becomes disconnected, which is the threshold below which the dispersion correction can no longer be efficient.

THEOREM 5.1. *Assume that* $2b+c-1\geq 0$*,* $(1-2c)/(16a-8)>0$ *and that* $4a(2b+c-1)-4b+1\geq 0$*. Then the discrete dispersion relation* $\widetilde{\mathcal{D}}_h$ *of* $\widetilde{\mathcal{H}}_h$ *is disconnected if and only if*

$$\widetilde{G}:=\frac{2\pi}{\widetilde{k}h}<\widetilde{G}^*=\pi\sqrt{2b+c-1}.$$

In addition, the discrete dispersion relation is non-empty if and only if

$$\widetilde{G}:=\frac{2\pi}{\widetilde{k}h}\geq\widetilde{G}^{\min}:=2\pi\sqrt{\frac{1-2c}{16a-8}}.$$

Proof. To prove the theorem, we are going to study the parametric representation of the representative curve of $\widetilde{\mathcal{D}}_h$. From its definition, we have that any (ξ_1,ξ_2) belong to $\widetilde{\mathcal{D}}_h$ if and only if it satisfies the equation

$$\cos(h\xi_2)=f(\cos(h\xi_1),\widetilde{G}),\tag{5.1}$$

where

$$f(X,\widetilde{G})=\frac{((2a-1)\widetilde{G}^2+c\pi^2)X-2a\widetilde{G}^2+2b\pi^2}{((2a-2)\widetilde{G}^2+2b\pi^2+2c\pi^2-2\pi^2)X+(-2a+1)\widetilde{G}^2-c\pi^2}.$$

Equation (5.1) defines a smooth curve $\xi_1\mapsto\xi_2(\xi_1)$ with $(\xi_1,\xi_2(\xi_1))\in\widetilde{\mathcal{D}}_h$ as soon as $f(\cos(h\xi_1),\widetilde{G})\in[-1,1]$. We show below that there exists $\widetilde{G}^*$ such that $\forall\widetilde{G}<\widetilde{G}^*$, $f\left([-1,1],\widetilde{G}\right)\supsetneqq[-1,1]$. This will prove that, for all $\widetilde{G}<\widetilde{G}^*$, there exists at least one $\xi_1=\arccos(X)$ for which we can not find a corresponding ξ_2 satisfying $(\xi_1,\xi_2)\in\widetilde{\mathcal{D}}_h$. In other words, the intersection of $\widetilde{\mathcal{D}}_h$ with the vertical line at such ξ_1 is not empty for all $\widetilde{G}\geq\widetilde{G}^*$ but becomes empty for all $\widetilde{G}<\widetilde{G}^*$ which means that the discrete dispersion relation becomes disconnected.

Note that there are some $a_1,b_1,a_2,b_2\in\mathbb{R}$ such that $f(X,\widetilde{G})=\frac{a_1X+b_1}{a_2X+b_2}$, and thus $f'(X,\widetilde{G})=(a_1b_2-a_2b_1)/(a_2X+b_2)^2$ has constant sign hence $X\in[-1,1]\mapsto f(X,\widetilde{G})\in\mathbb{R}$ is monotonic. As a result, we have the estimate

$$\forall X\in[-1,1],\ \min\left\{f(1,\widetilde{G}),f(-1,\widetilde{G})\right\}\leq f(X,\widetilde{G})\leq\max\left\{f(1,\widetilde{G}),f(-1,\widetilde{G})\right\},$$

with

$$f(1,\widetilde{G}) = \frac{2bh^2\widetilde{k}^2 + ch^2\widetilde{k}^2 - 4}{2bh^2\widetilde{k}^2 + ch^2\widetilde{k}^2 - 2h^2\widetilde{k}^2 - 4} = \frac{\widetilde{G}^2 - \pi^2(2b+c)}{\widetilde{G}^2 + (-2b-c+2)\pi^2},$$

$$f(-1,\widetilde{G}) = \frac{-2bh^2\widetilde{k}^2 + ch^2\widetilde{k}^2 + 16a - 4}{2bh^2\widetilde{k}^2 + 3ch^2\widetilde{k}^2 - 2h^2\widetilde{k}^2 + 16a - 12} = \frac{(4a-1)\widetilde{G}^2 + (-2b+c)\pi^2}{(4a-3)\widetilde{G}^2 + (2b+3c-2)\pi^2}.$$

Some computations give that $\widetilde{G} \mapsto f(-1,\widetilde{G})$ is monotonic and that $\widetilde{G} \mapsto f(1,\widetilde{G})$ is increasing with $\lim_{\widetilde{G}\to+\infty} f(1,\widetilde{G}) = 1$. In addition, one can see that the derivative of $\widetilde{G} \mapsto f(-1,\widetilde{G})$ is positive if and only if $4a(2b+c-1) - 4b + 1 \geq 0$. We now only need to compute the $\widetilde{G} > 0$ for which $f(1,\widetilde{G}_{1,-}) = -1$ and $f(-1,\widetilde{G}_{-1,\pm}) = \pm 1$. This gives

$$\widetilde{G}_{1,-} = \widetilde{G}_{-1,+} = \pi\sqrt{2b+c-1}, \qquad \widetilde{G}_{-1,-} = 2\pi\sqrt{\frac{1-2c}{16a-8}}.$$

Since $\widetilde{G} \mapsto f(-1,\widetilde{G})$ is increasing, we have that $\widetilde{G}_{-1,+} = \widetilde{G}_{1,-} \geq \widetilde{G}_{-1,-}$ and then, due to the monotonic behavior of the upper and lower bound of $X \mapsto f(X)$, we have that

$$\forall \widetilde{G} \leq \widetilde{G}_{1,-},\ f(1,\widetilde{G}) \leq f(1,\widetilde{G}_{1,-}) = -1,$$
$$f(-1,\widetilde{G}) \leq f(-1,\widetilde{G}_{-1,+}) = 1 \text{ or } f(-1,\widetilde{G}) \geq f(-1,\widetilde{G}_{-1,+}) = 1.$$

This observation gives that for all $\widetilde{G} \geq \widetilde{G}_{1,-}$, the vertical line at $\xi_1 = 0$ (that is for $X = 1$) intersects the discrete dispersion relation characterized by (5.1) while for all $\widetilde{G} < \widetilde{G}_{1,-} := \widetilde{G}^*$, this intersection is empty which means that $\widetilde{\mathcal{D}}_h$ becomes disconnected. Actually, we also show that for all $\widetilde{G} < \widetilde{G}_{-1,-}$, we have either that $f(X,\widetilde{G}) < -1$ or $f(X,\widetilde{G}) > 1$ and thus, because of (5.1), the discrete dispersion relation is empty. □

Remark 5.2. *We emphasize that the assumption on the coefficients (a,b,c) of Theorem 5.1 are satisfied for the 5-point stencil $(a,b,c) = (1,1,0)$ and the 4-th order FD scheme from Theorem 3.1 with $(a,b,c) = (5/6, 5/6 - c/2, c)$ for $c \geq 1/18$. We also compute below the values of G^* and $G^{\min}$ (without tildes since we do not use a shifted wavenumber) for these two schemes. Since the 5-point stencil can be obtained from (3.1) with $a = 1$, $b = 1$ and $c = 0$, $\mathcal{D}_{\mathrm{h}}^{5-\mathrm{pts}}$ is non-empty for any $G \geq G_{5-pts}^{\min} = \frac{\pi}{\sqrt{2}} = 2.221441469$. Theorem 5.1 applied to the 5-point stencil gives that the discrete dispersion relation is disconnected for any*

$$G < G_{5-pts}^* = \pi. \tag{5.2}$$

We also consider the 9-point stencil with the parameters

$$c = 8/45,\ a = 5/6,\ b = 2(5/12 - c/4), \tag{5.3}$$

which is fourth order accurate according to Theorem 3.1. We get that the discrete dispersion relation is non-empty for any $G \geq G_{9-pts}^{\min} = 2\pi\sqrt{\frac{29}{240}} = 2.184103659138$, and becomes disconnected for any $G < G_{9-pts}^* = \frac{\sqrt{6}\pi}{3} = 2.565099660324$. We emphasize that both values are smaller than the thresholds obtained for the 5-point stencil.

Theorem 5.1 is valid with or without a real shift $\widetilde{k}$ on the wavenumber. It is worth noting that $G = 2\pi/(kh)$ is the genuine number of grid points per wavelength

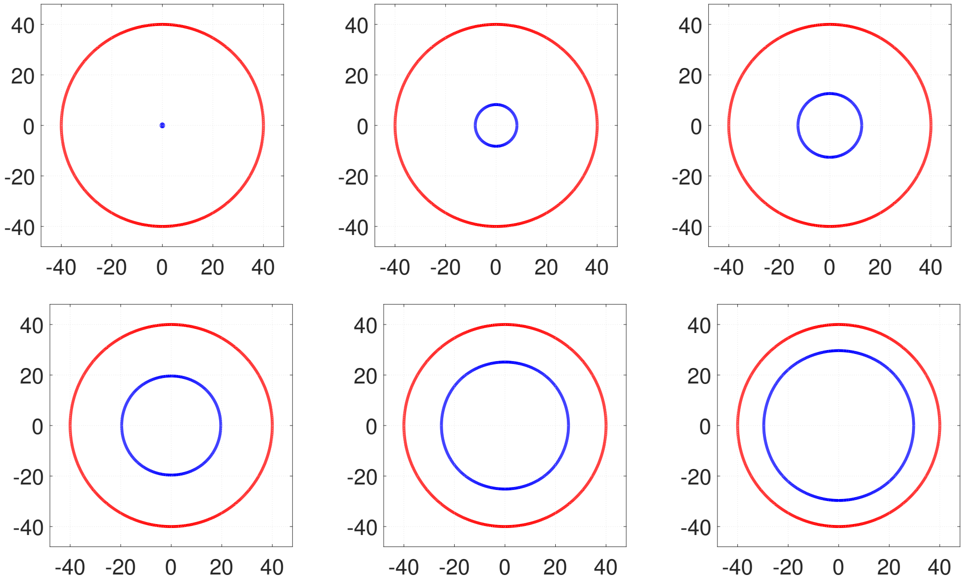

FIG. 5.1. *Discrete dispersion relation for* $G = 1.52,\ 1.6,\ 1.65,\ 1.75,\ 1.85,\ 1.95$ *from top left to bottom right. Blue: Discrete dispersion relation of the optimized FD scheme* $\mathcal{H}^{\text{asympt}}$*, Red: Continuous dispersion relation.*

while $\widetilde{G} = 2\pi/(\widetilde{k}h)$ is the relevant parameter involved in the dispersion relation of the asymptotically optimized finite difference scheme $\mathcal{H}_h^{\text{asympt}}$. They are related by the formula $G = \frac{\widetilde{k}}{k}\widetilde{G}$. For the FD scheme $\mathcal{H}_h^{\text{asympt}}$, we have $\widetilde{k} < k$ (see Theorem 4.1) and thus $G < \widetilde{G}$. Since we have lower bounds on $\widetilde{G}$ for the discrete dispersion relation to be non-empty and connected, we get that using a real shift allows us to consider values of G that are actually smaller than G^* and $G^{\min}$. As a result, the range of G for which the dispersion relation is either non-empty or connected is larger when using the real shift than without.

We now illustrate this fact using the parameters (5.3) that give $\widetilde{G}^{\min}_{9-pts} \geq 2.18$. Using the definition of $\widetilde{k}$, this requirement translates to

$$2\pi\sqrt{\frac{1-2c^{\text{asympt}}}{16a-8}} \leq \widetilde{G} = \frac{2\pi}{\widetilde{k}^{\text{asympt}}} = \frac{G}{(1-1\pi^4G^{-4}-\pi^6G^{-6}/192)},$$

which is actually satisfied for any G such that $(1-1\pi^4G^{-4}-\pi^6G^{-6}/192) > 0$ that is $G \geq 1.52450064$. We represent the dispersion relation of $\mathcal{H}_h^{\text{asympt}}$ in Figure 5.1for $1.52 \leq G \leq 1.95$. This shows that, for the asymptotically optimized scheme, discrete waves can still propagate for less than two points per wavelength even though the dispersion correction is no longer working properly.

5.2. Numerical optimization of the parameters. In this section, we numerically compute for a given G the values of $(\widetilde{k}, c)$ that minimize the distance between the discrete and continuous dispersion relation. We use the (a, b, c) from (5.3) and $k \in \{20, 40, 80, 100, 160, 180, 200, 250, 300\}$, $G \in \{40, 30, 20, 15, 10, 5, 4, 3, 2.5\}$. We chose to work with this range of G because, in the next section, we compare the dispersion error of the asymptotically optimized FD scheme with the sixth order FD

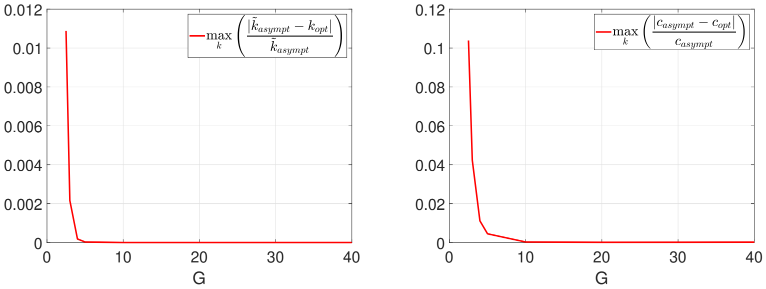

FIG. 5.2. *Left: Relative error between optimized k and $\widetilde{k}^{\mathrm{asympt}}$ from Theorem 4.1. Right: Relative error between numerically optimized value of c and the asymptotically optimized c^{asympt} from Theorem 4.2*

scheme from [35, Eq. (2.15)] which involves constants that are no longer defined when $G < 2.5$. We minimize the distance between the discrete and continuous dispersion relation for the above (k, G) using the Nelder-Mead method with the MATLAB function *fminsearch*. Figure 5.2 gives the relative error between the asymptotically optimized shift, namely $\widetilde{k}^{\mathrm{asympt}} = k - k\frac{\pi^4}{30G^4} - k\frac{\pi^6}{192G^6}$, and the numerically optimized one k^{opt}.

We see from these results that our shifted wavenumbers computed to get the best asymptotic (as $G \to +\infty$) dispersion error are actually good approximations to the numerically optimized one for $G \geq 5$. Figure 5.2 shows the relative error between the asymptotically optimized c, that is $c^{\mathrm{asympt}} = \frac{8}{45} - \frac{8}{45}G^{-2}$, and the numerically optimized one. We also get that the numerically optimized value of c does not depend on the wavenumber. For large G, c^{asympt} is a good approximation of the optimal c for $G \geq 10$. To conclude, the asymptotically optimized FD scheme has similar accuracy as the optimized one for any $G \geq 3$.

5.3. Distance between the discrete and continuous dispersion relations. We compute here the relative distance between the discrete and continuous dispersion curves for various FD method. That is

$$\mathrm{dist}_{\mathrm{r}} = \frac{\mathrm{dist}(\mathcal{D}_{\mathrm{c}}, \mathcal{D}_h)}{k},$$

where $\mathcal{D}_h$ is going to be the dispersion relation of some FD schemes. It is worth noting that $100 \times \mathrm{dist}_{\mathrm{r}}$ represents the percentage of error of the distance between the discrete and continuous dispersion curves. We introduce the following notations in order to consider various FD schemes: $\mathcal{H}^{\mathrm{asympt}}$ is given by (3.1) with asymptotic parameters from Theorem 4.2, $\mathcal{H}^{\mathrm{opt}}$ is given by (3.1) with numerically optimized parameters from Section 5.2, $\mathcal{H}^{\mathrm{Wu}}$ is the sixth order scheme from [35, Eq. (2.15)], $\mathcal{H}^{\mathrm{JSS}}$ is defined in [23], $\mathcal{H}^{\mathrm{fd5}}$ is the standard 5-point stencil, $\mathcal{H}^{\mathrm{Lambe}}$ is the FD stencil from [24], $\mathcal{H}^{\mathrm{Sutmann}}$ is the sixth-order stencil from [32]. It is the same stencil as in [35] but without dispersion correction. The numerical results are represented in Figure 5.3. The relative distance of the FD schemes with reduced numerical dispersion behave like $\mathrm{Cste} \times G^{-r}$ where r is their own order of accuracy. Note that this is expected for large G because of Theorem 2.3 and that this formula also holds for smaller values of G.

The results indicate that only $\mathcal{H}^{\mathrm{fd5}}$, that is the standard 5-point stencil, has a very bad numerical dispersion. This actually comes from the fact that its dispersion curve

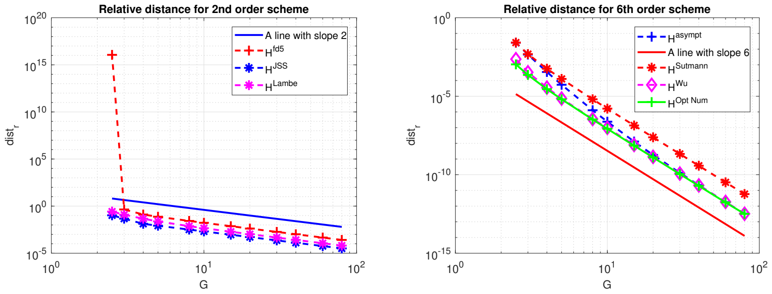

Fig. 5.3. *Relative distance between the discrete and continuous dispersion relation for several FD schemes*

becomes disconnected below $G = \pi$, see (5.2). The dispersion error of $\mathcal{H}^{\text{Lambe}}$ and $\mathcal{H}^{\text{JSS}}$ is smaller than the 5-point stencil because these FD schemes feature a dispersion correction.

The scheme with the least numerical dispersion is $\mathcal{H}^{\text{opt}}$ which also has similar dispersion error as the sixth order scheme with dispersion correction $\mathcal{H}^{\text{Wu}}$ from [35]. The asymptotically optimized scheme $\mathcal{H}^{\text{asympt}}$ also has similar dispersion error as $\mathcal{H}^{\text{Wu}}$ and $\mathcal{H}^{\text{opt}}$ for $G \geq 10$. The stencil $\mathcal{H}^{\text{Sutmann}}$ has the worst dispersion error for any considered G which is because this is the only scheme without dispersion correction. As a result, our asymptotically optimized FD scheme, which is formerly only fourth order accurate and performs dispersion correction with explicit formulas for the parameters, have numerical dispersion comparable to those of some genuinely sixth order schemes. These two observations show that using a real shift on the wavenumber yields some substantial improvement.

5.4. Limits of equi-oscillation for dispersion correction. We now show that finding the best dispersion correction by equioscillation is limited for values of G bigger than some limiting value $\widetilde{G}_{\text{equi}}$. For smaller G, the best dispersion correction is not characterized by equioscillation any more. We do this for simplicity with the 5-point finite difference stencil using only the shifted wave number $\widetilde{k}$ as parameter. We show in Figure 5.4 for different values of G the original dispersion curve of the 5-point scheme in black, the exact circular dispersion curve in red, and in blue the best dispersion correction possible using only the modified wave number $\widetilde{k}$. We see that for $G = 3.5$ and $G = \pi$ (see (5.2) for the latter), the original black dispersion curve is still a connected set, but much worse an approximation of the circular red exact dispersion curve than the dispersion corrected blue one. We also see that the blue dispersion corrected curve equioscillates around the red continuous circular dispersion curve, with the maxima of the difference attained at zero, 45 and 90 degree angles. For $G = 3, 2.5, 2.343$, the original dispersion curve is disconnected and lost its physical meaning (see also Remark 5.2), while the dispersion corrected curve still equioscillates and best approximates the continuous circular one, albeit for $G = 2.343$ now by a square. For smaller $G = 2, 1.5, 1, 0.5$, this is not possible any more and the blue dispersion corrected dispersion curve is just staying as close as possible to the red continous circular dispersion curve, remaining a square and shrinking. To compute this best possible dispersion curve, we recall that the dispersion curve for the 5-point FD scheme is given by

$$\cos hx + \cos hy - \left(2 - \frac{\widetilde{k}^2 h^2}{2}\right) = 0. \tag{5.4}$$

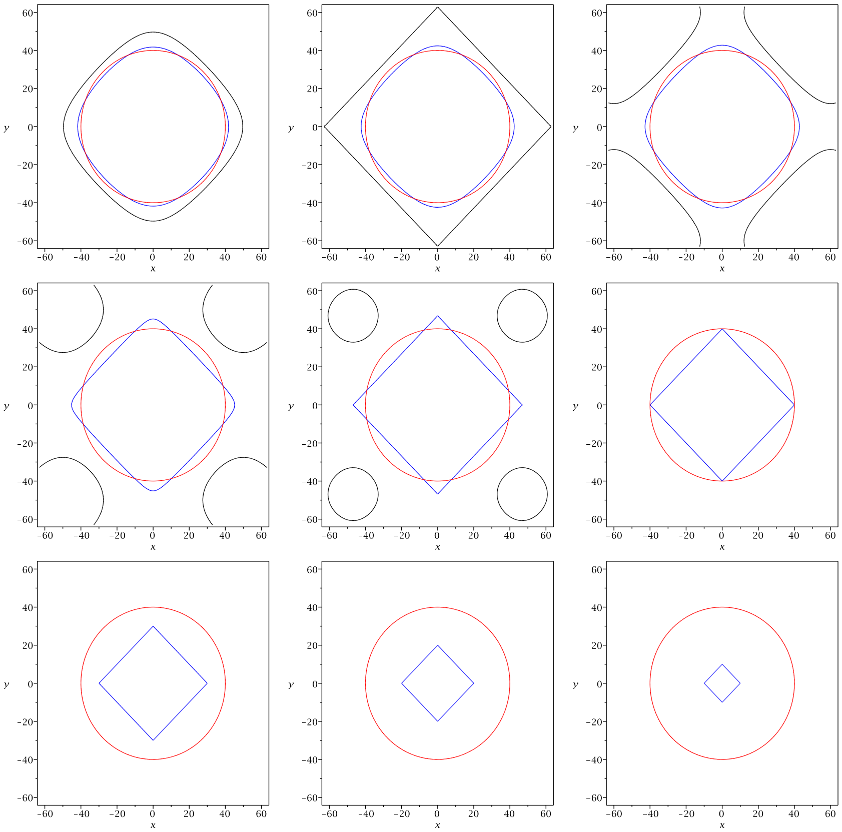

FIG. 5.4. *Best dispersion correction (blue) for the 5-point finite difference scheme for* $G = 3.5, \pi, 3, 2.5, 2.343, 2, 1.5, 1, 0.5$*, compared to the uncorrected dispersion relation (black) and the exact continuous dispersion relation (red).*

The signed distance from the circle at zero and 45 degree angle is therefore given by

$$d_0 := \frac{\pi - \arccos(\frac{1}{2}\widetilde{k}^2h^2 - 1)}{h} - k, \quad d_{45} := \sqrt{2}\frac{\pi - \arccos(\frac{1}{4}\widetilde{k}^2h^2 - 1)}{h} - k.$$

Setting $h := \frac{2\pi}{Gk}$ and solving $d_0 + d_{45} = 0$, we find the dispersion correction parameter $\widetilde{k}$ which mininizes the distance by equioscillation. Now we see in Figure 5.4 that as G becomes smaller, the dispersion corrected dispersion relation becomes more and more square, and the value when the dispersion corrected dispersion relation is exactly a square on its tip is when the dispersion corrected dispersion relation becomes a linear function. Solving the dispersion relation (5.4) for y, we find

$$y = \frac{\pi - \arccos(\cos(hx) - 2 + \frac{1}{2}\widetilde{k}^2h^2)}{h},$$

which becomes a linear function of x if

$$\frac{1}{2}\widetilde{k}^2h^2 = 2 \Longleftrightarrow \frac{k}{\widetilde{k}}G = \pi \Longleftrightarrow \widetilde{k} = \frac{kG}{\pi}.$$

Replacing this value into the expressions for the distances, we obtain

$$d_0 + d_{45} = \frac{1}{4}(\sqrt{2}G - 8 + 2G)k = 0,$$

and it is therefore only possible to do dispersion correction by equioscillation up to

$$\widetilde{G}_{\text{equi}} := \frac{8}{2+\sqrt{2}} = 2.343145750.$$

For smaller G, the closest we can get to the exact dispersion relation is when making it a square, $\widetilde{k} = \frac{kG}{\pi}$, which is plotted in Figure 5.4 for values $G < \widetilde{G}_{\text{equi}}$.

5.5. Accuracy on plane wave solutions. In this part, we investigate the accuracy of the FD schemes $\mathcal{H}^{\text{asympt}}$, $\mathcal{H}^{\text{opt}}$, $\mathcal{H}^{\text{Wu}}$ and $\mathcal{H}^{\text{Sutmann}}$ when solving a boundary value problem whose exact solution is a plane wave,

$$\begin{cases} \Delta u(\theta,\mathbf{x}) + k^2 u(\theta,\mathbf{x}) &= 0, \qquad\qquad\qquad \text{in } \Omega := (-1,+1)\times(-1,+1), \\ \qquad\qquad\quad u(\theta,\mathbf{x}) &= \exp(\mathrm{i}k\mathbf{x}\cdot\mathbf{d}(\theta)), \quad \text{on } \partial\Omega, \end{cases} \tag{5.5}$$

where $\mathbf{d}(\theta) = (\cos(\theta), \sin(\theta))$ is a unit vector. We emphasize that $u(\theta,\mathbf{x}) = \exp(\mathrm{i}k\cdot \mathbf{x}\mathbf{d}(\theta))$ is the unique solution to problem (5.5) only if k^2 is not an eigenvalue of the unbounded operator defined as the Laplace operator with domain $H_0^1(\Omega)$. In addition, since the distance between the continuous and discrete dispersion relations of $\mathcal{H}^{\text{asympt}}$ and $\mathcal{H}^{\text{opt}}$ both behave like $O(G^{-6})$ (see Theorem 4.1 and Figure 5.3), Theorem 2.3 ensures that these two schemes are going to give 6-th order approximations of $u(\theta,\mathbf{x})$ for any angle θ.

For a given meshsize h, we consider a uniform grid of points $\mathbf{x}_{i,j} = (x_i, y_j) \in \Omega$ which thus verify $|x_i - x_{i\pm1}| = h$ and $|y_j - y_{j\pm1}| = h$ for any i, j. We are going to solve the discrete problems associated to (5.5) for the four FD schemes and the angles $\theta \in \left\{\frac{2\pi}{N}l,\ l = 0, \cdots, N\right\}$. Denoting by $\mathbf{u}_h(\theta) := (\mathbf{u}_{h,i,j}(\theta))_{i,j}$ the discrete solution, we compute the averaged relative error

$$\text{err}(h) = \frac{1}{N+1}\sum_{l=0}^{N} \frac{\|u(\theta_l, \mathbf{x}_{i,j}) - \mathbf{u}_h(\theta_l)\|_2}{\|\mathbf{u}_h(\theta_l)\|_2},\ \mathbf{x}_{i,j} = (x_i, y_j), \tag{5.6}$$

where $\|\mathbf{v}\|_2 = \sqrt{\sum_{i,j}|\mathbf{v}_{i,j}|^2}$, is the usual Euclidean norm.

We first give convergence results for fixed wavenumbers and decreasing meshsize to get the accuracy as $h \to 0$ of each FD stencil. Next, we investigate the pollution effect [4, 22, 3, 36] which can be defined as the fact that, as the wavenumber increases, it is not enough to have a constant number of grid points per wavelength to keep the relative error bounded. We are then going to compute the averaged relative error (5.6) for G fixed and k increasing.

We choose $N = 20$ and for the meshsize $h \in \{0.005, 0.008, 0.01, 0.02, 0.03\}$ and the wavenumbers $k \in \{10, 20, 40, 80\}$. The numerical results are shown in Figure 5.5. As indicated by Figure 5.3 which shows the relative distance between the discrete and continuous dispersion relations, the stencils $\mathcal{H}^{\text{Wu}}$ and $\mathcal{H}^{\text{opt}}$ have very similar accuracy. Also, as expected, $\mathcal{H}^{\text{asympt}}$ has similar accuracy as these two if the number of grid points per wavelength is large enough (see Figure 5.5 for $k = 10, 20$). Note that, as k increases, G becomes smaller and we are thus no longer in the regime $G \to +\infty$ in which the asymptotically optimized scheme has been designed to be

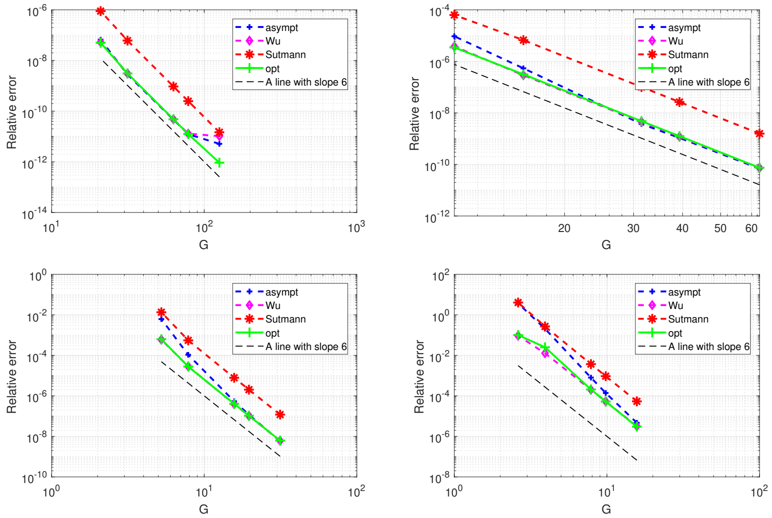

FIG. 5.5. *Relative errors for $k = 10$ (top left), $k = 20$ (top right), $k = 40$ (bottom left) and $k = 80$ (bottom right).*

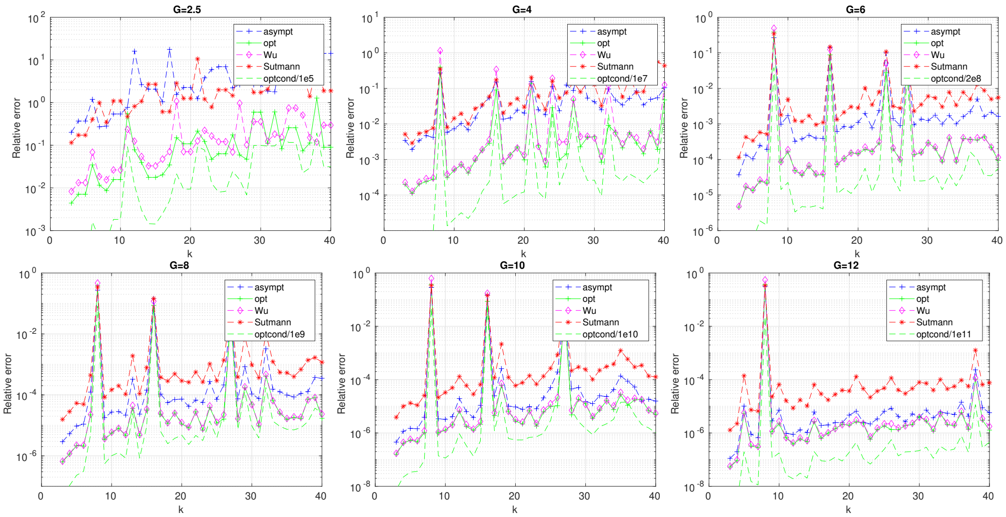

FIG. 5.6. *Relative error for G fixed and k increasing.*

efficient. Nevertheless, $\mathcal{H}^{\text{asympt}}$ is still a 6-th order scheme for any k and G considered. It is worth noting that $\mathcal{H}^{\text{Sutmann}}$ is less accurate than the three other FD schemes because this is the only stencil that does not feature a dispersion correction. Note that the pollution effect has not been overcome since the relative error increases with the wavenumber. Nevertheless, comparing with the accuracy of $\mathcal{H}^{\text{Sutmann}}$, this effect is indeed reduced by our dispersion correction.

To study the pollution effect further, we choose $N = 8$ and work now with G fixed given successively by $G \in \{2.5, 4, 6, 8, 10, 12\}$, and for every G, we use the wavenumbers $k \in \{3, 4, \ldots, 40\}$. The relative errors we measured are shown in Figure 5.6. First note that for some values of G, k, there are some bumps shared by the four

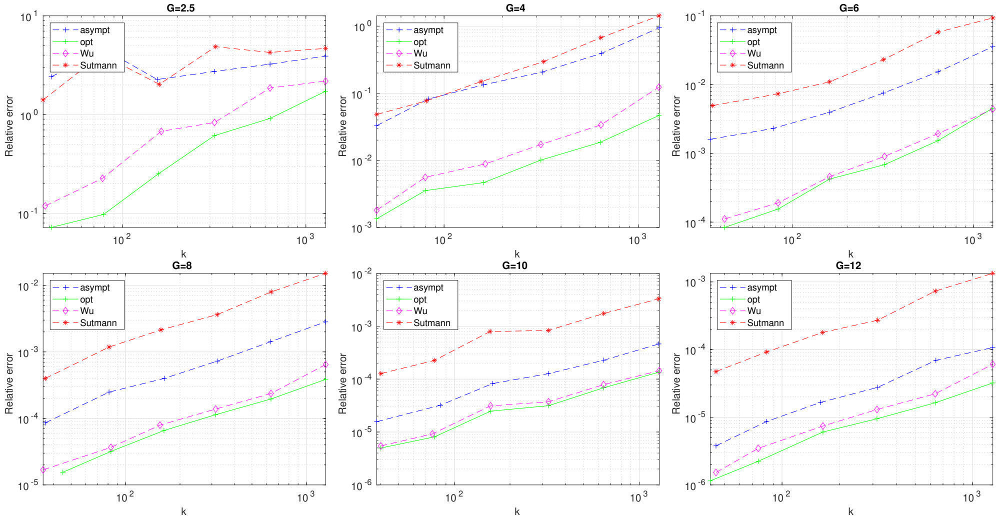

Fig. 5.7. *Relative error for G fixed and k increasing.*

schemes considered. To explain these, we also computed numerically the condition number of each matrix, and we show as example the scaled condition number of our optimized scheme also in Figure 5.6. We see that the condition number has bumps in the same location as the relative error shown in Figure 5.6. This indicates that the reason for the bumps is the fact that we approximate problem (5.5) which is singular for certain k, h and thus ill conditioned when close to a discrete eigenvalue, and we thus focus on the other values, away from the bumps. We first note that for $G = 2.5$ and this wavenumber range, both the Wu scheme and our new optimized scheme lead to usable solutions with error of about 10 percent, while the Sutmann scheme and the asymptotic dispersion correction can not retain any accuracy. For $G = 4$ and this wave number range, the Wu scheme and our optimized dispersion correction give errors below 1 percent, and the Sutmann scheme and our asymptotic formula are below 10 percent. This trend continues, but most importantly we see that the asymptotic closed formula scheme becomes better as G is increasing, and for the typical engineering practice of $G = 10$ or $G = 12$ points per wavelength, the asymptotic closed form scheme now performs almost as well as the Wu scheme and our optimized dispersion correction, which both rely on numerical optimization. To study the behavior for much larger wave numbers, we consider now $k \in \{40, 80, 160, 320, 640, 1280\}$, and to avoid the bumps due to ill conditioning, we simulate in an interval around each wave number, e.g. for $k = 40$ we run all wavenumbers $\{35, 36, 37, \ldots, 45\}$, and pick for each scheme the case with the lowest error. The results are displayed in Figure 5.7. We can now clearly see how the dispersion correction lowers the pollution effect: in all results, our new optimized dispersion correction gives the lowest error, followed by the Wu scheme, and the Sutmann scheme without dispersion correction has an error which is between one and two orders of magnitudes larger. For only $G = 2.5$ points per wavelength (!), our new optimized dispersion correction scheme can still give a 10 percent accuracy for k up to about 100, and with $G = 4$ we get an error of $5e - 2$ for $k = 1280$. We also see that the asymptotic closed formula becomes more and more effective as G increases, and for $G = 12$ we approach the numerically optimized formula. In terms of computing time, we measured the average of 18 solutions on a modern workstation: for low accuracy and $k = 80$, the Sutmann scheme needs $G = 4$ for an error $0.8e-1$ which takes 0.032 sec., and our new optimized scheme with $G = 2.5$ and similar error $1e-1$ takes 0.012 sec., a factor 3 faster. For intermediate accuracy, and $k = 160$, the Sutmann scheme needs $G = 10$ for an error $8e-4$ which takes 2.4 sec., and our new optimized scheme with $G = 6$ and similar error $4e-4$ takes 0.58 sec., a factor 4 faster. For high accuracy, and $k = 1280$, the Sutmann scheme needs $G = 12$ for an error $1.5e-3$ which takes 371 sec., and our new optimized scheme with $G = 6$ and similar error $3e-3$ takes 78.4 sec., a factor 5 faster. We finally observe that for a given number of points per wavelength G, the Sutmann scheme without dispersion correction gives for $k = 40$ a similar or even lower accuracy than our new optimized dispersion correction for $k = 1280$.

REMARK 5.3. *Due to [26, p. 1223, Theorem 4.10], any solution to a Helmholtz problem can be written as $u(\mathbf{x}) = u_{\mathcal{A}}(\mathbf{x}) + u_{H^2}(\mathbf{x})$, where $u_{H^2} \in H^2(\Omega)$ has H^2-bounds that do not depend on k, and $u_{\mathcal{A}}$ is oscillatory. When doing dispersion correction, we thus have a scheme that is more accurate only for approximating $u_{\mathcal{A}}$. As a result, even if the asymptotically optimized and the optimized FD stencils are 6-th order accurate for plane wave solutions (see Theorems 4.2, 2.3 and the numerical results in Figure 5.5), these two schemes are going to be 4-th order accurate (see Theorem 3.1) for general right hand side.*

5.6. Multigrid with dispersion correction. We investigate now the performance of the multigrid algorithm [21, 29] for the 5-point stencil, the fourth-order 9-point stencil (5.3), and the asymptotically optimized and the optimized stencils. A general multigrid algorithm with l levels can be written as

```
function $\mathbf{z}_l = MGM_l(\mathbf{z}_l, \mathbf{b}_l)$
 if l = 1 then $\mathbf{z}_1 = A_1^{-1}\mathbf{b}_0$ else
    $\mathbf{z}_l = S^{\nu_1}(\mathbf{z}_l, \mathbf{b}_l)$;                    % pre – smoothing
    $\mathbf{d}_{l-1} = R_l(\mathbf{b}_l - A_l\mathbf{z}_l)$;
    $\mathbf{e}_{l-1}^0 = \mathbf{0}$;
    for $j = 1$ to $\tau$ do                                        (5.7)
      $\mathbf{e}_{l-1}^j = MGM_{l-1}(\mathbf{e}_{l-1}^{j-1}, \mathbf{d}_{l-1})$;
    end
    $\mathbf{z}_l = \mathbf{z}_l + P_l\mathbf{e}_{l-1}^\tau$;
    $\mathbf{z}_l = S^{\nu_2}(\mathbf{z}_l, \mathbf{b}_l)$;                    % post – smoothing
 end,
```

where A_l are the discrete operators defined on the grid at level l, R_l are restriction operators acting from fine (level $l+1$) to coarse grids (level l), P_l are prolongation operators acting from coarse to fine grids, S_l are linear iterative methods known as smoothers, and $l = 1$ is actually the coarsest level on which the matrix is finally inverted. The parameters ν_1 and ν_2 correspond to the number of pre and post-smoothing steps, and τ permits to consider either a V-cycle ($\tau = 1$) or a W-cycle ($\tau = 2$). From [29, p.22, Theorem 7.1] (see also [21, Lemma 7.14]), the algorithm 5.7 is a linear iterative method whose iteration matrix is given by

$$\begin{aligned} C_{MG,1} &= 0, \\ C_{MG,l} &= S_l^{\nu_2}\left(I - P_l\left(I - C_{MG,l-1}^{\tau}\right)A_{l-1}^{-1}R_lA_l\right)S_l^{\nu_1}. \end{aligned} \tag{5.8}$$

We emphasize that $C_{MG,l} = C_{MG,l}(\nu_2, \nu_1)$, and that

$$\rho\left(C_{MG,l}(\nu_2, \nu_1)\right) = \rho\left(C_{MG,l}(0, \nu_1 + \nu_2)\right),$$

$L=2,\ k=64 \setminus \nu$	1	5	10	20
5-pts	6.6410×10^6	6.1082×10^6	5.5040×10^6	4.4746×10^6
9-pts	35.0880	18.5364	13.3576	12.1106
asympt	22.4954	6.9003	1.6712	0.6875
opt	60.0441	18.4501	4.6126	0.4802
$L=3,\ k=32$				
5-pts	2.4703×10^5	2.0306×10^5	1.7001×10^5	1.3084×10^5
9-pts	13.6200	6.3527	5.6742	5.6195
asympt	25.8102	5.2241	0.8920	0.6922
opt	68.5569	14.1023	2.8002	0.3742
$L=4,\ k=16$				
5-pts	7.6109×10^3	5.9310×10^3	4.8903×10^3	3.8032×10^3
9-pts	7.3096	3.0443	2.2077	2.0340
asympt	25.2834	4.5225	0.8597	0.6956
opt	67.1272	12.3786	2.4282	0.3870

Table 5.1

Spectral radius of the multigrid iteration matrix for V-cycles ($\tau = 1$) and $p = 6$. We have approximately π grid points per wavelength at the coarsest level and performed ν smoothing steps on each level.

so it suffices to study only the one parameter case $\nu = \nu_1 + \nu_2$.

We use standard restriction and prolongation operators (see e.g. [31, p. 12, Eq. (23)]), defined for any grid function $\mathbf{u}$ at level $l+1$ by

$$\begin{aligned} (R_l\mathbf{u})_{i,j} &:= \tfrac{1}{16}(4u_{2i,2j} + 2(u_{2i-1,2j} + u_{2i+1,2j} + u_{2i,2j-1} + u_{2i,2j+1}) \\ &\quad +u_{2i-1,2j-1} + u_{2i+1,2j-1} + u_{2i-1,2j+1} + u_{2i+1,2j+1}), \\ P_l &:= 4R_l^T. \end{aligned} \tag{5.9}$$

For the smoother, we use a damped Kacmarz-like smoother (see e.g. [5, p. 9, Section 4] and [9]) whose iteration matrix is given by $S_l = \mathbb{I} - \omega_l A_l^* A_l$, with $\omega_l = \rho(A_l)^{-2}$.

Remark 5.4. *Since $S_l^* = S_l$, the smoother satisfies the two properties $\sigma(S_l) \subset [0,1]$ and $\|S_l\|_2 = \rho(S_l) \leq 1$. Since the smoother is bounded, it will not lead to instabilities like the ω-Jacobi smoother used in [31, p. 15, Section 4.2]. These instabilities can yield divergence of the multigrid algorithm if too many smoothing steps are performed as pointed out in [27].*

We now study the performance of Algorithm 5.7 applied to the Helmholtz equation with homogeneous Dirichlet boundary conditions on $\Omega = (0,1)\times(0,1)$. At any level l, the grid is an $n_l \times n_l$ uniform grid of Ω with

$$n_l = 2^l - 1,\ h_l = \frac{1}{n_l + 1},$$

so the grids are nested. We use the same number ν of smoothing steps on each level. For a number of levels L, we will then have done $\nu(L-1)$ smoothing steps at the end of the cycle. In all our numerical computations, we fix the number of levels L, and next chose k such that we have approximately π grid points per wavelength at the coarsest level. This ensures that the dispersion relation of the 5-point stencil is connected as shown in Remark 5.2. On the finest level, we use an $n_p \times n_p$ uniform grid with $p = 6$ hence we have $n_p \times n_p = 63^2 = 3969$ grid points on the finest grid. We show in Table 5.1 the spectral radius of the iteration matrices for V-cyles ($\tau = 1$) with $L = 2, 3, 4$, and in Table 5.2 the corresponding results for W-cycles ($\tau = 2$). For each stencil, one can see that the spectral radius is decreasing as ν increases. As a result the multigrid method is convergent if one uses a large enough number of smoothing

$L = 3,\ k = 32 \backslash \nu$	1	5	10	20
5-pts	5.0795×10^{10}	3.8270×10^{10}	2.8781×10^{10}	1.7963×10^{10}
9-pts	176.9828	43.2154	33.1786	32.2884
asympt	573.3670	36.0669	1.8799	0.4950
opt	4.0956×10^{3}	259.5673	13.6469	0.2512
$L = 4,\ k = 16$				
5-pts	2.1604×10^{15}	1.1155×10^{15}	0.6010×10^{15}	0.2335×10^{15}
9-pts	2.4772×10^{3}	0.1379×10^{3}	0.0335×10^{3}	0.0210×10^{3}
asympt	3.1880×10^{5}	1.105×10^{3}	3.0124	0.2558
opt	1.6271×10^{7}	5.7142×10^{4}	155.0806	0.1377

Table 5.2

Spectral radius of the multigrid iteration matrix for W-cycles ($\tau = 2$) and $p = 6$. We have approximately π grid points per wavelength at the coarsest level and performed ν smoothing steps on each level.

steps. This property of the multigrid algorithm is known to hold for elliptic coercive problems (see e.g. [21]) and seems to hold here for the Helmholtz equation due to the boundedness of the smoother (see Remark 5.4). However, for the 5-point and 9-point stencils, the minimal number of smoothing steps required for the iteration matrix to be a strict contraction is too large for this algorithm to be used in practice. As shown in [13], this behavior is also observed when dealing with the Helmholtz equation with a complex wavenumber whose imaginary part is not large enough. More importantly, note how the dispersion correction tremendously reduces the value of the minimal ν.

Note however also that, whatever the number of levels we use, the minimal value of ν for which multigrid converges is also increasing with the wavenumber k. To illustrate this, we show in Figure 5.8 the value of $\nu_{\min}$ defined as

$$\nu_{\min} := \min \left\{ \nu \in \mathbb{N}^* \middle| \ \rho \left(C_{MG,l}(0, \nu) \right) < 1 \right\},$$

as well as the spectral radius of the iteration matrix if $\nu_{\min}$ smoothing steps are done at each level. The performance of the asymptotically optimized scheme and the optimized scheme are very similar for small wavenumbers. This behavior is expected from the results of Section 5.2 since such cases correspond to large G. Note however that the reduction factor of the (numerically) optimized scheme is roughly two-times smaller than the one of the asymptotically optimized scheme for $\nu = 20$ (see Tables 5.1 and 5.2).

From the results of Tables 5.1 and 5.2, one can see that the performance of the W-cycle is bad for a small number of smoothing steps. Nevertheless, when we use 20 smoothing steps at each level, the spectral radius is smaller for the W-cycle than for the V-cycle and, for both V and W-cycles, the value of $\nu_{\min}$ is then quite similar (see Figure 5.8). Note finally that the asymptotically optimized scheme whose definition does not rely on numerical optimization can then yield a convergent multigrid algorithm even for a number of grid points per wavelength at the coarsest level that is out of the asymptotic range for which our formula has been derived.

We end this section by solving a Helmoltz problem on $(0, 1) \times (0, 1)$ with homogeneous Dirichlet boundary conditions and a non-zero source term given by

$$f(x, y) = \sin\left(\frac{x}{2}\right) \sin\left(ky\right) k.$$

We solve this problem for several values of k using the MGM algorithm with the same number of smoothing steps at each level. We emphasize that the multigrid algorithm

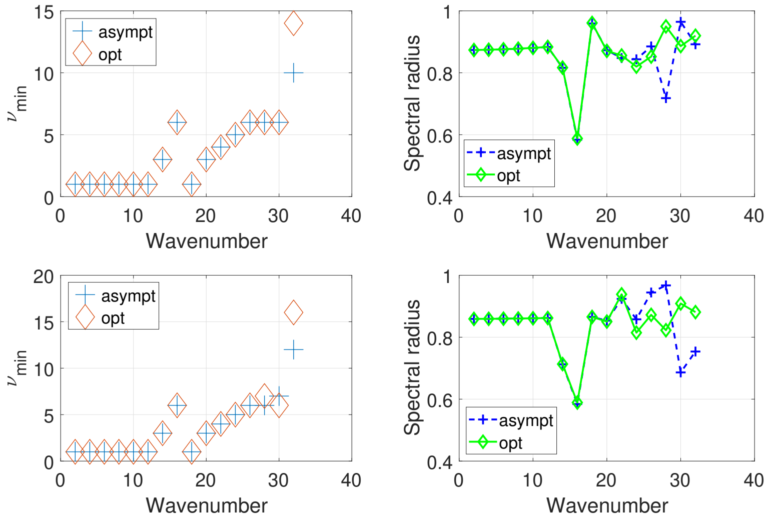

FIG. 5.8. *Value of* $\nu_{\min}$ *(left) and spectral radius (right) of the iteration matrix for* $L = 3$*,* $p = 6$*, V-cycle (top) and W-cycle (bottom).*

(5.7) is a fixed point iteration $z_l^{j+1} = MGM(z_l^j, b_l)$. In all our numerical experiments, we start with a random initial guess and used the stopping criterion

$$\sqrt{\sum_{m=1}^{n_p \times n_p} \frac{1}{n_p \times n_p} \left(z_{j+1}^m - z_j^m\right)^2} = \frac{1}{n_p} \left\| z_{j+1} - z_j \right\|_2 \leq 10^{-6},$$

where $n_p \times n_p$ is the number of grid points at the finest level. We define the reduction factor after some iterations as

$$\text{Reduction factor} := \max_j \frac{\left\| z_{j+1} - z_j \right\|_2 / n_p}{\left\| z_j - z_{j-1} \right\|_2 / n_p} = \max_j \frac{\left\| z_{j+1} - z_j \right\|_2}{\left\| z_j - z_{j-1} \right\|_2}.$$

We only consider the asymptotically optimized and the numerically optimized FD schemes, since the standard 5 and 9-pt stencils have convergence issues as we have already seen in Tables 5.1 and 5.2. We also chose to work with a small number of grid points at the coarsest level ($G_{coarse} < 6$) since, in this range, the asymptotically optimized scheme is far from the optimized one as indicated by Figure 5.2. The numerical results are given in Tables 5.3 for the V-cycle and 5.4 for the W-cycle.

For the V-cycle (see Table 5.3), one can see that the numerically optimized scheme performs much better than the asymptotically optimized scheme. In addition, while the performance of the asympt scheme collapses as k increases, those of the numerically optimized scheme remain similar. Regarding the W-cycle (see Table 5.4), both schemes almost have the same performances with reduction factors being the same up to 10^{-5}.

To conclude, the results presented in this section confirm that correcting the dispersion is one of the main features needed to design a convergent multigrid algorithm for Helmholtz problems.

6. Conclusions and outlook. We explored the idea of using a real shift for the wavenumber to do dispersion correction. Contrary to most of the previous results

Wavenumber	MGM settings	G finest grid	FD scheme	Reduction factor	Number of iterations
$k = 90$	$p = 8,\ L = 3$	17.8722	asympt	0.4944	15
			opt	0.1475	7
	$p = 9,\ L = 4$	35.7443	asympt	0.4982	15
			opt	0.1538	7
$k = 180$	$p = 8,\ L = 2$	8.9361	asympt	0.7764	28
			opt	0.1142	7
	$p = 9,\ L = 3$	17.8722	asympt	0.7273	23
			opt	0.1437	7
$k = 360$	$p = 9,\ L = 2$	8.9361	asympt	-	divergence
			opt	0.2620	8

Table 5.3

Results obtained when solving a Helmholtz problem on $(0, 1) \times (0, 1)$ *with homogeneous Dirichlet boundary conditions and source term* $f(x, y) = k \sin(ky) sin(x/2)$ *using a multigrid V-cycle with* $\nu = 30$ *smoothing steps at each level. In all these experiments, we have roughly* 4.5 *points per wavelength at the coarsest level.*

Wavenumber	MGM settings	G finest grid	FD scheme	Reduction factor	Number of iterations
$k = 90$	$p = 8,\ L = 3$	17.8722	asympt	9.006×10^{-3}	4
			opt	9.012×10^{-3}	4
	$p = 9,\ L = 4$	35.7443	asympt	7.207×10^{-3}	4
			opt	7.211×10^{-3}	4
$k = 180$	$p = 9,\ L = 3$	17.8722	asympt	9.020×10^{-3}	4
			opt	9.043×10^{-3}	4

Table 5.4

Results obtained when solving a Helmholtz problem on $(0, 1) \times (0, 1)$ *with homogeneous Dirichlet boundary conditions and source term* $f(x, y) = k \sin(ky) sin(x/2)$ *using a multigrid W-cycle with* $\nu = 30$ *smoothing steps at each level. In all these experiments, we have roughly* 4.5 *points per wavelength at the coarsest level.*

from the literature, our optimized coefficients are explicitly determined and do not rely on a numerical optimization procedure. Our optimized FD scheme has a dispersion relation that is not empty for a number of grid points per wavelength below 2, but even with optimized dispersion correction the dispersion error becomes large then. The real shift on the wavenumber allows us also to have a fourth-order FD scheme which is sixth-order accurate on plane wave solutions, which means having numerical dispersion comparable to those of some formerly sixth order schemes. We also showed that our asymptotically optimized scheme can be used in a multigrid algorithm and that the resulting linear iterative method is convergent if the number of smoothing steps is large enough. The minimal number of smoothing steps for which the iteration matrix is a strict contraction is greatly reduced with the dispersion correction. This striking result together with those from [9, 10, 16] show that the idea of using a real shift to reduce numerical dispersion is really promising. Finally, we would like to emphasize that all the results of this paper can be easily extended to the three-dimensional case without any additional difficulties.

Some interesting further work can be envisaged: first, the idea of the real shift could be applied to the sixth order scheme from [35]. According to the authors, the discrete wavenumber is not easily computable. Therefore, the asymptotic dispersion correction introduced in this paper can be appealing since only approximations as $G \to +\infty$ of the dispersion relation are needed (see e.g. Theorem 4.1) to compute the asymptotically optimal shifted wavenumber.

A further topic is the use of FD methods with reduced numerical dispersion in multigrid algorithms. We already studied numerically the behavior here, but a theoretical analysis would be of great interest, and is technically not easy. In addition, it is known that the Helmholtz operator with a wavenumber having a non-zero imaginary part ε is a good preconditioner for the original Helmholtz problem if $\varepsilon = O(k)$ [17, 12]. On the other hand, the multigrid algorithm has its best performance when the shift is large enough, that is $\varepsilon = O(k^2)$ [13, 11]. As a result, one should investigate the behavior of the multigrid method using a FD scheme with dispersion correction to see if this algorithm can have good performance for a shift smaller than the wavenumber squared.

Another further direction is the extension of the methods we introduced to Finite-Element methods. It could be indeed interesting to apply our techniques to some coercive variational formulations for the Helmholtz equation that have been introduced and studied in [14, 18, 25]. Finally, for some long-term perspectives, it could be very interesting to lift some key assumptions used throughout this paper like, for instance, considering non-uniform meshes and a spatially-varying, piecewise constant wavenumber.

7. Appendix. Certain technical calculations can be checked with symbolic computations. In this appendix, we give the corresponding commands which can directly be executed in Maple. In the Proof of Theorem 3.1, the Taylor expansion (3.2) can be obtained with

```
> S1:=f(x-h,y)+f(x+h,y)+f(x,y-h)+f(x,y+h);
> S2:=f(x-h,y-h)+f(x+h,y-h)+f(x-h,y+h)+f(x+h,y+h);
> S:=(4*a/h^2-k^2*b)*f(x,y)+((1-2*a)/h^2-(1/4)*k^2*c)*S1
  -((1-a)^2+(1/4)*k^2*(1-b-c))*S2;
> simplify(series(S,h));
```

The result allows one to get Equation (3.3) which gives the conditions on a, b, c for the FD scheme to be fourth order. Equation (3.4) can be obtained with

```
> f:=(x,y)->exp(I*k*(x*cos(theta)+y*sin(theta)));
> simplify(series(S,h,8));
```

The proof of Theorem 4.1 is based on the asymptotic expansion of $(x(\theta), y(\theta))$ solution to $F(\widetilde{k}_h, c, x(\theta), y(\theta), G) = 0$. The full asymptotic expansion of the function F as $G \to +\infty$ can be obtained with

```
> F:=(4*a/h^2-ktilde^2*b)+(2*((1-2*a)/h^2-(1/4)*ktilde^2*c))*(cos(h*x)+cos(h*y))
  -(2*((1-a)/ h^2+(1/4)*ktilde^2*(1-b-c)))*(cos(h*(x+y))+cos(h*(y-x)));
> a:=5/6; b:=5/6-(1/2)*c;
> h:=2*Pi/(k*G);
> y:=x*tan(theta);
> ktilde:=k0+k1/G+k2/G^2+k3/G^3+k4/G^4+k5/G^5+k6/G^6;
> c:=c0+c1/G+c2/G^2; x:=x0+x1/G+x2/G^2+x3/G^3+x4/G^4+x5/G^5+x6/G^6;
> collect(simplify(convert(asympt(F,G,8), polynom)),G);
```

The last result allows one to determine that

```
> k0:=k; x0:=k*cos(theta); x1:=0; k1:=0; x2:=0; x3:=0;
> k2:=0; k3:=0; c0:=8/45; x4:=0; k4:=-(1/30)*Pi^4*k;
> k5:=0; x5:=0; c1:=0;
```

To get x_6, one computes the sixth order term in the asymptotic expansion of F and sets it to zero,

```
> x6:=solve(limit(asympt(F,G,12)*G^6,G=infinity)=0,x6);
```

The distance between the discrete and continuous dispersions relations and its sixth order term involved in Equation (4.5) are then obtained as

```
> d:=simplify(sqrt(1+tan(theta)^2)*x*csgn(1/(cos(theta))));
> d6:=collect((limit((d-k)*G^6,G=infinity))/k,k);
```

The function $H(Y, k_6, c_2) = d_6(\arccos(Y), k_6, c_2)$, its derivative, its critical points and the value of H at these points that are needed to get Equation (4.6) can be obtained with

```
> theta:=arccos(Y); H:=d6; CritPoints:=solve(diff(H,Y)=0,Y);
> for L to 7 do
    Y:=CritPoints[L]; collect(simplify(H),k); Y:=’Y’;
  end do;
```

In the proof of Theorem 5.1, one can get the expression of the function f involved in Equation (5.1) based on F defined above by

```
> F:=expand(F); B:=solve(%,cos(h*y));
> h:=2*Pi/(ktilde*G); cos(2*Pi*x/(ktilde*G)):=X;
> f:=collect(simplify(B),[X,G]);
```

The value of $f(-1, \widetilde{G}), f(+1, \widetilde{G})$ and the solution to $f(1, \widetilde{G}_{1,-}) = -1$, $f(-1, \widetilde{G}_{-1,\pm}) = \pm 1$ can be obtained by

```
> X:=1; fp1:=collect(B,[G,Pi]); X:=’X’; X:=-1;
> fm1:=collect(B,[G,Pi]); X:=’X’:
> Gp1:=solve(fp1=-1,G); Gm1:=solve(fm1=1,G);
> Gmm1:=solve(fm1=-1,G);
```

REFERENCES

[1] Ainsworth, M. (2004). Discrete dispersion relation for hp-version finite element approximation at high wave number. SIAM Journal on Numerical Analysis, 42(2), 553-575.

[2] Anguelov, R., & Lubuma, J. M. S. (2001). Contributions to the mathematics of the nonstandard finite difference method and applications. Numerical Methods for Partial Differential Equations: An International Journal, 17(5), 518-543.

[3] Babuska, I. M., & Sauter, S. A. (1997). Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?. SIAM Journal on Numerical Analysis, 34(6), 2392-2423.

[4] Barucq, H., Calandra, H., Pham, H., & Tordeux, S. (2017). A study of the Numerical Dispersion for the Continuous Galerkin discretization of the one-dimensional Helmholtz equation (Doctoral dissertation, Inria Bordeaux Sud-Ouest; Magique 3D).

[5] Bramble, J. H., Kwak, D. Y., & Pasciak, J. E. (1994). Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems. SIAM journal on Numerical Analysis, 31(6), 1746-1763.

[6] Chen, Z., Cheng, D., Feng, W., & Wu, T. (2013). An optimal 9-point finite difference scheme for the Helmholtz equation with PML. International Journal of Numerical Analysis & Modeling, 10(2), 389-410.

[7] Chen, Z., Cheng, D., & Wu, T. (2012). A dispersion minimizing finite difference scheme and preconditioned solver for the 3D Helmholtz equation. Journal of Computational Physics, 231(24), 8152-8175.

[8] Cheng, D., Tan, X., & Zeng, T. (2017). A dispersion minimizing finite difference scheme for the Helmholtz equation based on point-weighting. Computers & Mathematics with Applications, 73(11), 2345-2359.

[9] Cocquet, P. H., Gander, M. J., & Xiang, X. (2019). Dispersion correction for Helmholtz in 1D with piecewise constant wavenumber, accepted for publication in Domain Decomposition Methods in Science and Engineering XXV, LNCSE, Springer-Verlag.

[10] Cocquet, P. H., Gander, M. J., & Xiang, X. (2018). A finite difference method with optimized dispersion correction for the Helmholtz equation, Domain Decomposition Methods in Science and Engineering XXIV, LNCSE, Springer-Verlag, 205-213.

[11] Cocquet, P. H., & Gander, M. J. (2016). On the minimal shift in the shifted Laplacian preconditioner for multigrid to work. In Domain Decomposition Methods in Science and Engineering XXII. Springer International Publishing, 137-145.

[12] Cocquet, P. H., & Gander, M. J. (2018). Analysis of the shifted Helmholtz expansion preconditioner for the Helmholtz equation. Domain Decomposition Methods in Science and Engineering XXIV, LNCSE, Springer-Verlag, 195-204.

[13] Cocquet, P. H., & Gander, M. J. (2017). How Large a Shift is Needed in the Shifted Helmholtz Preconditioner for its Effective Inversion by Multigrid?. SIAM Journal on Scientific Computing, 39(2), A438-A478.

[14] Diwan, G. C., Moiola, A., & Spence, E. A. (2019). Can coercive formulations lead to fast and accurate solution of the Helmholtz equation?. Journal of Computational and Applied Mathematics, 352, 110-131.

[15] Ernst, O. G., & Gander, M. J. (2012). Why it is difficult to solve Helmholtz problems with classical iterative methods. In Numerical Analysis of Multiscale Problems. Springer Berlin Heidelberg. 325-363.

[16] Ernst, O. G., & Gander, M. J. (2013). Multigrid methods for Helmholtz problems: A convergent scheme in 1D using standard components. Direct and Inverse Problems in Wave Propagation and Applications. De Gruyer, 135-186.

[17] Gander, M.J., Graham, I.G., & Spence, E.A. (2015). Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: what is the largest shift for which wavenumber-independent convergence is guaranteed?. Numerische Mathematik, 131(3), 567-614.

[18] Ganesh, M., & Morgenstern, C. (2019). A coercive heterogeneous media Helmholtz model: formulation, wavenumber-explicit analysis, and preconditioned high-order FEM. Numerical Algorithms, 1-47.

[19] Harari, I., & Hughes, T. J. (1991). Finite element methods for the Helmholtz equation in an exterior domain: model problems. Computer Methods in Applied Mechanics and Engineering, 87(1), 59-96.

[20] Harari, I., & Turkel, E. (1995). Accurate finite difference methods for time-harmonic wave propagation. Journal of Computational Physics, 119(2), 252-270.

[21] W. Hackbusch. Multi-grid methods and applications, 2. printing. - Berlin [u.a.] : Springer, 2003. - XIV, 377 p. (Springer series in computational mathematics ; 4), ISBN 978-3-540-12761-1 ISBN 3-540-12761-5 ISBN 0-387-12761-5.

[22] Ihlenburg, F., & Babuska, I. (1995). Finite element solution of the Helmholtz equation with high wave number Part I: The h-version of the FEM. Computers & Mathematics with Applications, 30(9), 9-37.

[23] Jo, C. H., Shin, C., & Suh, J. H. (1996). An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator. Geophysics, 61(2), 529-537.

[24] Lambe, L. A., Luczak, R., & Nehrbass, J. W. (2003). A new finite difference method for the Helmholtz equation using symbolic computation. International Journal of Computational Engineering Science, 4(01), 121-144.

[25] Moiola, A., & Spence, E. A. (2014). Is the Helmholtz equation really sign-indefinite?. Siam Review, 56(2), 274-312.

[26] Melenk, J. M., & Sauter, S. (2011). Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation. SIAM Journal on Numerical Analysis, 49(3), 1210-1243.

[27] Nicolaides, R. A. (1978). On multigrid convergence in the indefinite case. Mathematics of Computation, 32(144), 1082-1086.

[28] Nadukandi, P., O$\tilde{n}$ate, E., & Garcia, J. (2011). A fourth-order compact scheme for the Helmholtz equation: Alpha-interpolation of FEM and FDM stencils. International Journal for Numerical Methods in Engineering, 86(1), 18-46.

[29] Reusken, A. (2009). Introduction to multigrid methods for elliptic boundary value problems, *Multiscale Simulation Methods in Molecular Sciences*, p. 467-506.

[30] Stolk, C. C. (2016). A dispersion minimizing scheme for the 3-D Helmholtz equation based on ray theory. Journal of computational Physics, 314, 618-646.

[31] Stolk, C. C., Ahmed, M., & Bhowmik, S. K. (2014). A multigrid method for the Helmholtz equation with optimized coarse grid corrections. SIAM Journal on Scientific Computing, 36(6), A2819-A2841.

[32] Sutmann, G. (2007). Compact finite difference schemes of sixth order for the Helmholtz equation, Journal of Computational and Applied Mathematics, 203, 15-31.

[33] Turkel, E., Gordon, D., Gordon, R., & Tsynkov, S. (2013). Compact 2D and 3D sixth order schemes for the Helmholtz equation with variable wave number. Journal of Computational Physics, 232(1), 272-287.

[34] Wu, T. (2017). A dispersion minimizing compact finite difference scheme for the 2D Helmholtz equation. Journal of Computational and Applied Mathematics, 311, 497-512.

[35] Wu, T., & Xu, R. (2018). An optimal compact sixth-order finite difference scheme for the Helmholtz equation. Computers & Mathematics with Applications, 75(7), 2520-2537.

[36] Zhu, L., & Wu, H. (2013). Preasymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part II: hp version. SIAM Journal on Numerical Analysis, 51(3), 1828-1852.