LMECA2660 - Homework: simulating convection

DEGROOFF Vincent – NOMA : 09341800

Friday, 18 march 2022

We consider the 1-D convection equation in the domain $x = [-L/2, L/2]$, with constant velocity c and periodic boundary conditions:

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = 0 \tag{1}$$

1 Stability of the temporal integration scheme

As seen during the lectures, since we deal with a periodic function, we can decompose it into modes of different wavenumbers k and different amplitudes A_k. The mode k is given by:

$$u_k(x,t) = A_k(t)\; e^{\imath kx}$$

When we insert this expression of $u_k(x,t)$ into equation (1), we obtain the "eigenvalue" of the continuous mode k, $\lambda = \imath ck$:

$$\frac{\mathrm{d}A_k(t)}{\mathrm{d}t} + [\imath ck]\, A_k(t) = 0$$

When we discretize the partial derivative ∂_x of equation (1) with the mode $u_k(x,t)$, we find a modified version of λ:

$$\frac{\mathrm{d}A_k(t)}{\mathrm{d}t} + \left[\imath ck\,\frac{k^*h}{kh}\right] A_k(t)$$

Hence, the eigenvalues of the schemes are given by:

$$\lambda_j \Delta t = \imath \frac{c\Delta t}{h}(k^*h) = \imath \cdot CFL \cdot (k^*h)$$

We want all the eigenvalues to be inside the marginal stability curve of RK4C. Since they are all purely imaginary, the condition is $|\lambda_j| < 2\sqrt{2}$ for $0 \leq k_j h \leq \pi$. We can then determine the CFL condition as: $CFL\ \max(k^*h) \leq 2\sqrt{2}$. Based on the results presented in Table 1, I chose $CFL = 1.0$.

Scheme	k^*h as a function of kh (from [1])	$\max \lvert k^*h \rvert$	CFL number
E2	$\sin(kh)$	1.	2.828
E4	$[8\sin(kh) - \sin(2kh)]/6$	1.372	2.061
E6	$[45\sin(kh) - 9\sin(2kh) + \sin(3kh)]/30$	1.586	1.783
I4	$[3\sin(kh)]/[2 + \cos(kh)]$	1.732	1.632
I6	$[28\sin(kh) + \sin(2kh)]/[18 + 12\cos(kh)]$	1.989	1.421

Table 1. CFL number of various finite-difference schemes using RK4C.

2 Validation of the periodic hypothesis

The initial condition is a Gaussian with standard deviation σ. We also recall its Fourier transform.

$$u(x,0) = U \exp\left(-\frac{x^2}{\sigma^2}\right)$$
$$\mathcal{F}\big(u(x,0)\big) = U\sqrt{\pi}\sigma \exp\left(-\frac{k^2\sigma^2}{4}\right)$$

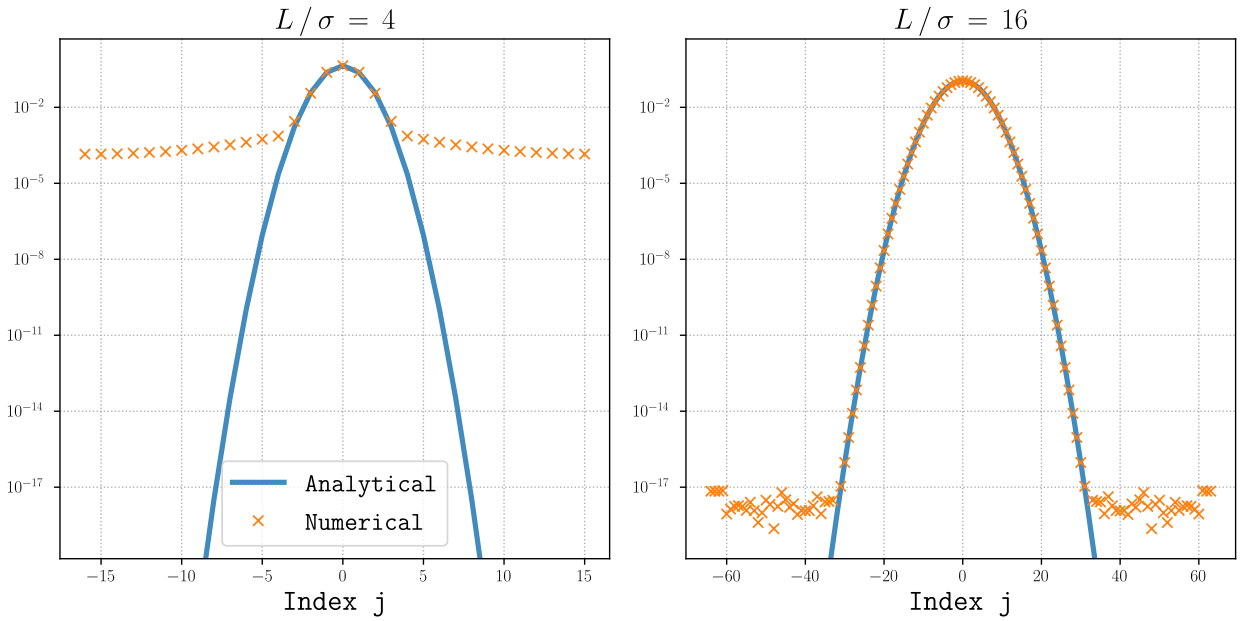

Figure 1. Amplitude of the Fourier transform and the FFT of the Gaussian initial function.

In both of the diagrams, the step size h is proportional to the width of the Gaussian: $h/\sigma = 1/8$. Hence, when $L/\sigma = 4$, we have $N = 32$ points, and when $L/\sigma = 16$, we have $N = 128$ points.

On one hand, in the left diagram of Figure 1 ($L/\sigma = 4$), the amplitude of the coefficients of the FFT and the Fourier transform in an infinite domain only coincide at low wavenumbers when $|j| \leq 3$. This represents approximately one fourth of the spectrum. At higher wavenumbers, they do not coincide while also being not negligible: $\approx 10^{-4}$.

On the other hand, in the right diagram of Figure 1 ($L/\sigma = 16$), the coefficients of the FFT have the right amplitude until they reach machine precision $\epsilon \approx 10^{-16}$. They coincide for $|j| \leq 31$: approximately half of the spectrum.

In view of this, it is clear that in our further analysis we will take a large domain compared to the width of the Gaussian: $L/\sigma = 16$.

3 Results obtained with a uniform grid

In this section, we will consider a grid with a constant Δx from $-L/2$ to $L/2$. In the following simulations, if U is not explicitly mentioned, it is supposed to be set to 1.

3.1 Comparison of analytical and numerical solutions

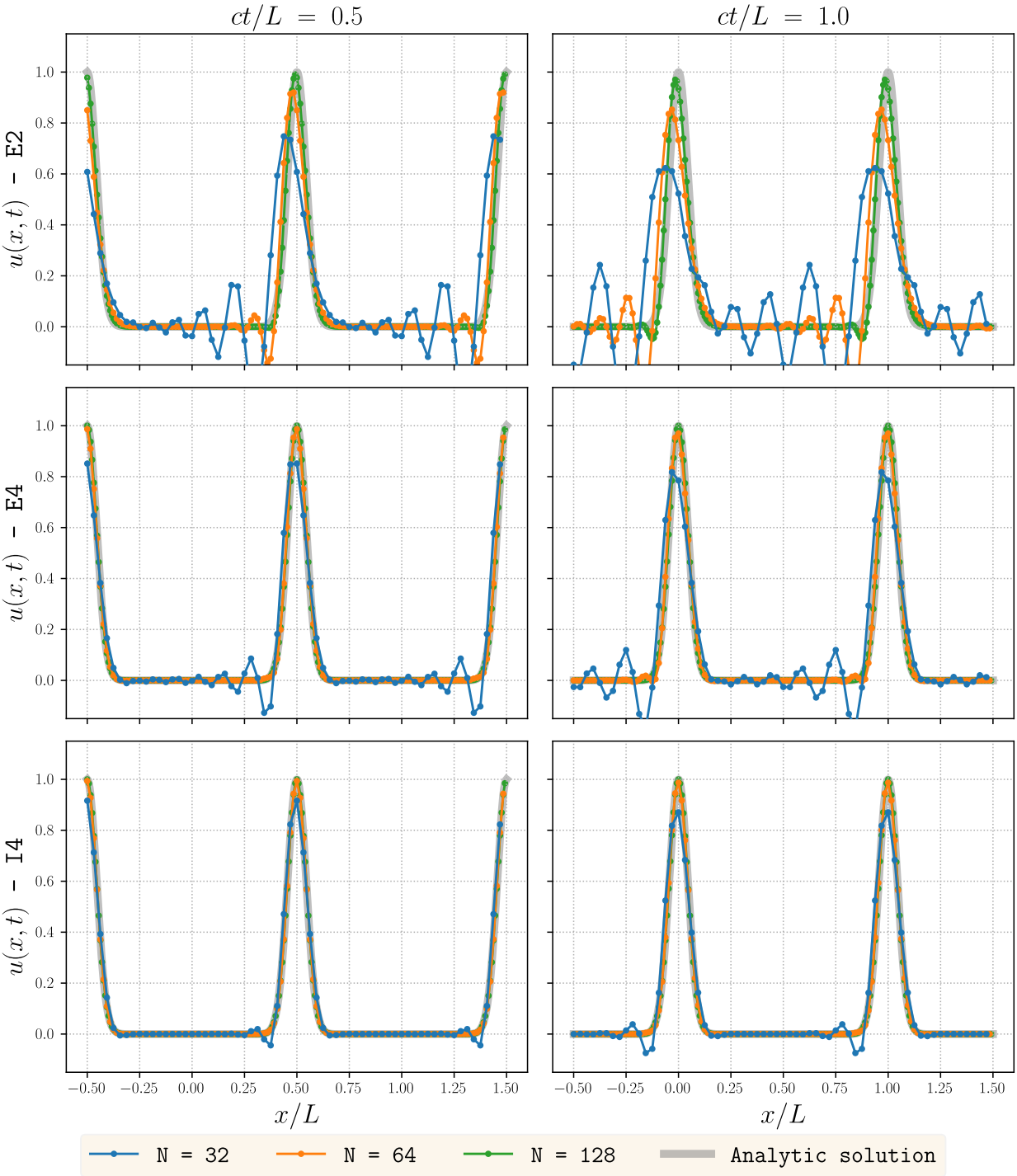

Figure 2. Solutions of the transport equation. Each row corresponds to one of the schemes E2, E4, I4. Each column shows the solution at a specific time. Then the colors correspond to the quality of the discretization.

As expected, in Figure 2, the scheme E2 gives the worst results, followed by E4 and then I4 provides the best approximation. This follows from the leading term of the truncation error T of each of these schemes [1, p. 5]:

$$T_{E2} = -\frac{1}{6}(kh)^2 \qquad T_{E4} = -\frac{1}{30}(kh)^4 \qquad T_{I4} = -\frac{1}{180}(kh)^4$$

For all the schemes, the large step size $h = \sigma/2$ ($N = 32$) gives bad results, particularly at the left of the peak of the Gaussian. In the case of the E2 scheme, still with $N = 32$, the signal appears "destroyed" already at $ct/L = 0.5$.

More generally, the larger the step size (the lower the N), the bigger is the dissipation (decrease of amplitude) and the dispersion (difference of velocity). Practically, in order to obtain results with sufficient quality, we can use the E2 scheme with at least $N = 128$, and the E4 and I4 schemes with at least $N = 64$.

3.2 Evolution of the diagnostics with time

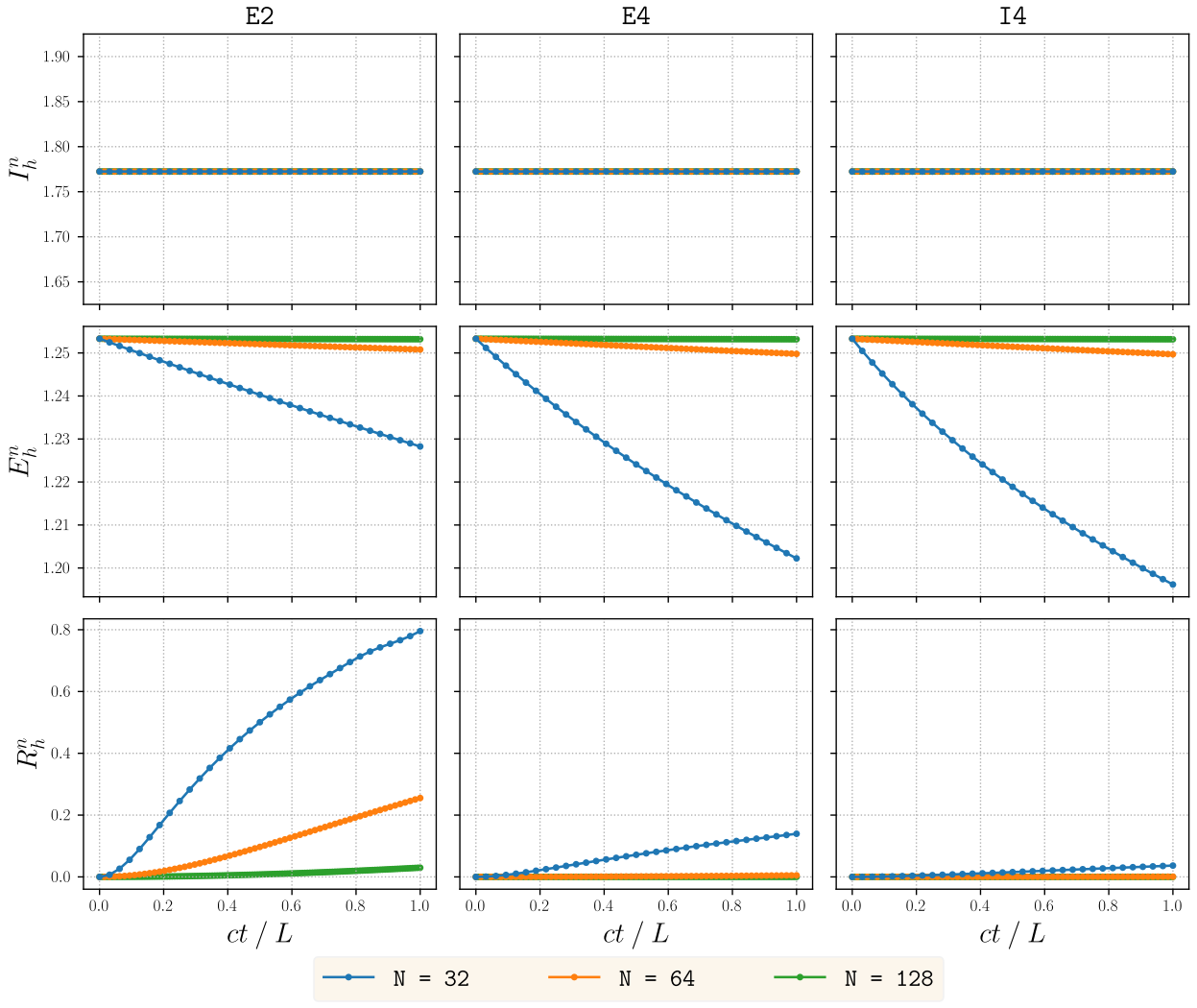

Figure 3. Presentation of the three diagnostic I_h, E_h and R_h as functions of a dimensionless time ct/L. The tree schemes are arranged in column and the three diagnostic in rows.

We recall the diagnostics: I for the integral, E for the energy, and R for the error. In the following, u_i^n indicates the discrete approximation of $u(x,t)$ at position $x_i = ih$ and time $t = n\Delta t$:

$$I_h^n = \frac{h}{\sigma U}\sum_{i=0}^{N-1} u_i^n \qquad E_h^n = \frac{h}{\sigma U^2}\sum_{i=0}^{N-1} (u_i^n)^2 \qquad R_h^n = \frac{h}{\sigma U^2}\sum_{i=0}^{N-1} \left[u_i^n - u\left(x_i, t^n\right)\right]^2$$

The first global diagnostic is the discrete version of the integral of the function $u(x,t)$ from $x = -b = -L/2$ to $x = b = L/2$. We see in Figure 3 that it is constant. We can also show that it is constant and find its value in the continuous case:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{-b}^{b} u\,\mathrm{d}x = \int_{-b}^{b} \frac{\partial u}{\partial t}\,\mathrm{d}x = -c\int_{-b}^{b} \frac{\partial u}{\partial x}\,\mathrm{d}x = -c\left[u\left(\frac{L}{2}\right) - u\left(-\frac{L}{2}\right)\right] = 0$$

$$\frac{1}{\sigma U}\int_{-b}^{b} u\,\mathrm{d}x = 2\int_{0}^{\frac{L}{2}\frac{16}{L}} \exp\left(-x^2\right)\mathrm{d}x = \sqrt{\pi}\,\mathrm{erf}(8) \approx 1.772$$

The second diagnostic concerns the energy contained in u. For all schemes, the worse the quality of the grid (the smaller the N), the faster is the decrease of energy. Even if it may appear counter-intuitive, the order of the scheme has the opposite effect. For all N, the lower the truncation error (the better the scheme), the faster is the decrease of energy. [1 year later] This is because the higher order schemes curves extend higher on the imaginary axis of the $\lambda\Delta t$ plane (cf. Table 1), which means that they are more dissipated by RK4C (look at Figure 10 for the marginal stability curve).

Finally, the third diagnostic measures the global squared error. This time, the results are not surprising. The error (1) increases over time, (2) decreases as the truncation error of the scheme decreases, and (3) decreases as we refine the grid (N increasing).

3.3 Order of convergence

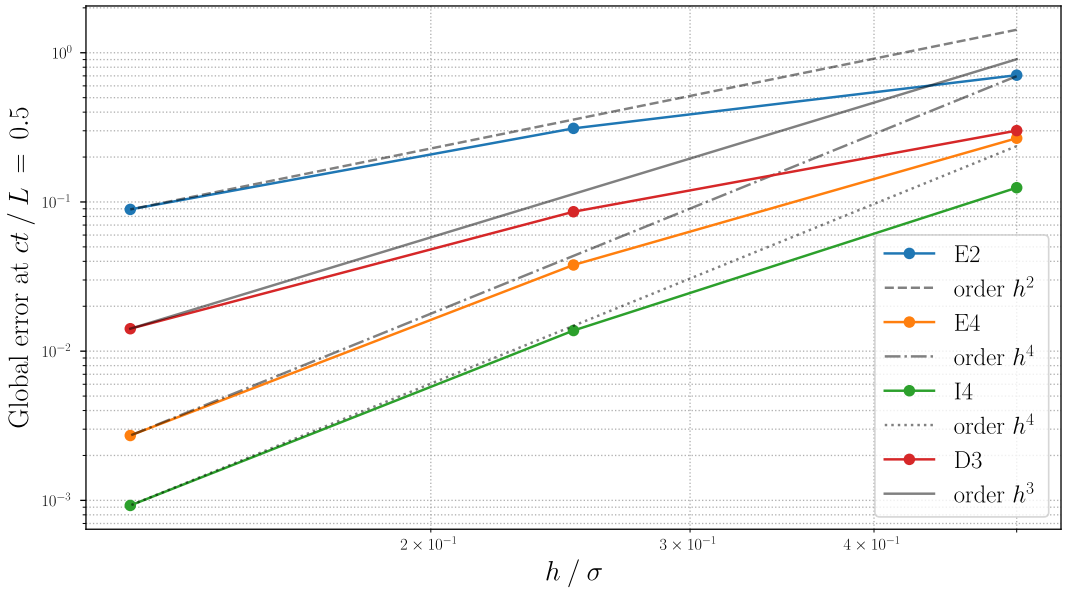

Figure 4. Error R_h at time $ct/L = 0.5$ as a function of h/σ for the different schemes.

The global error shown in Figure 4 is decreasing slightly faster than $\mathcal{O}(h^2)$ using the E2 scheme, and slightly faster than $\mathcal{O}(h^4)$ using the E4 and I4 schemes. As mentioned a few times before, the error made using I4 is smaller than the error made with E4, since their truncation errors differ by a factor of 6.

4 Results obtained with a non-uniform grid

We now consider a mapping, constant in time, from the "numerical space" $-\frac{L}{2} \leq \xi \leq \frac{L}{2}$ with a uniform grid to the "physical space" $-\frac{L}{2} \leq x \leq \frac{L}{2}$ with a non-uniform grid:

$$x = g(\xi) = \xi - a\frac{L}{2\pi}\sin\left(2\pi\frac{\xi}{L}\right) \qquad \text{and} \qquad \frac{\mathrm{d}x}{\mathrm{d}\xi} = g'(\xi) = 1 - a\cos\left(2\pi\frac{\xi}{L}\right)$$

We can use the chain rule to modify the transport equation (1).

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}x} = 0$$

$$\frac{\partial}{\partial t}\left(u\frac{\mathrm{d}x}{\mathrm{d}\xi}\right) + \frac{\partial (cu)}{\partial \xi} = 0 \qquad \text{since both } x = g(\xi) \text{ and } c \text{ do not depend on } t$$

$$\frac{\partial v}{\partial t} + \frac{\partial (bv)}{\partial \xi} = 0 \qquad \text{where } b(\xi) = \frac{c}{g'(\xi)} \quad \text{and} \quad v(\xi, t) = g'(\xi)\, u\big(x(\xi), t\big)$$

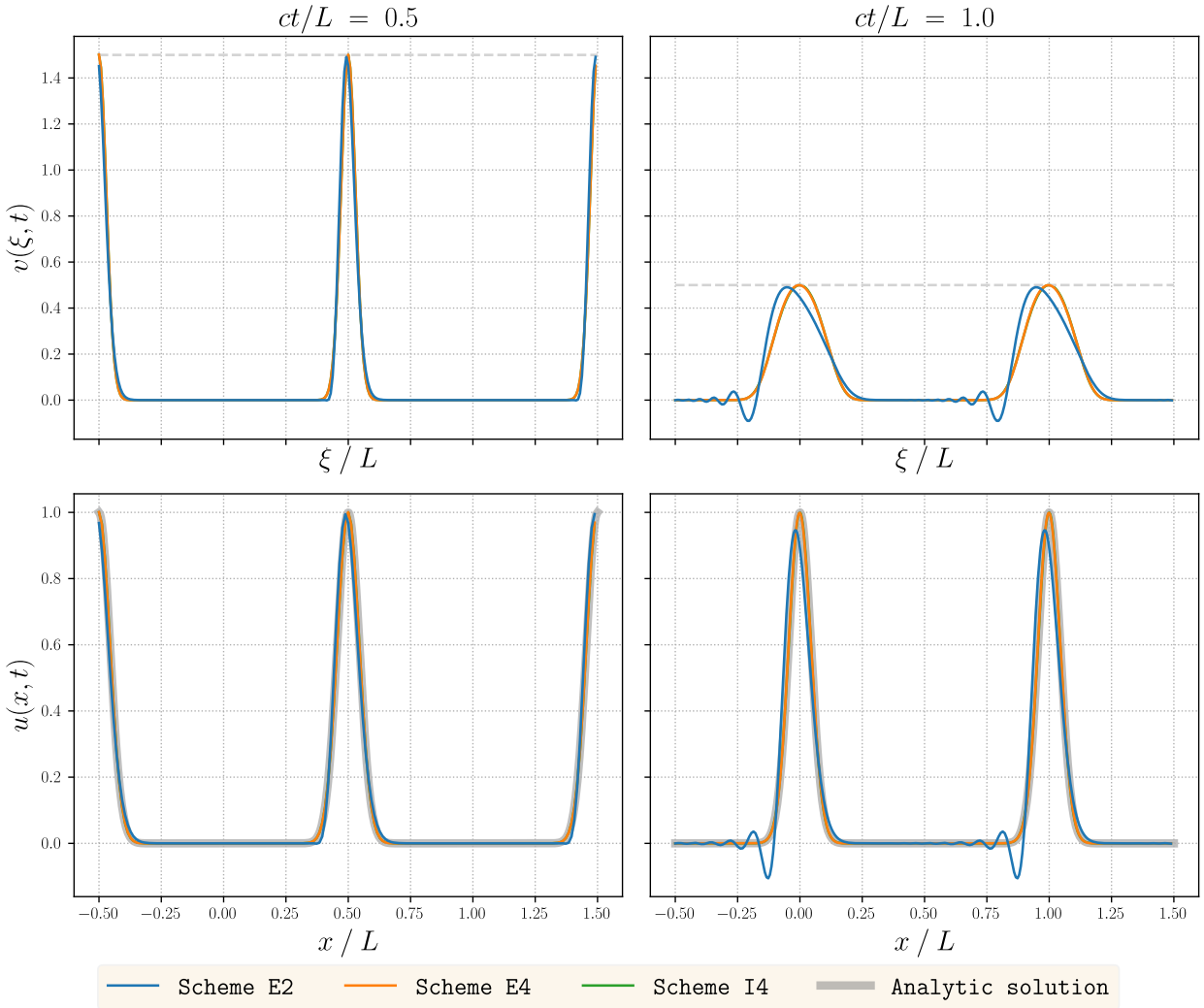

Figure 5. The first row shows the solution $v(\xi, t)$ in the numerical domain, while the second row shows the solution $u(x, t)$ in the physical domain. Each column shows the solution at a specific time.

When the peak of the Gaussian is located where the density of points ξ_i is high, it makes $v(\xi, t)$ more distributed and gives it a lower peak. This is what we observe in the right part of Figure 5 at $ct/L = 1$ when the Gaussian is located at $x = 0 \mod L$. The peak is enlarged since there are more points ξ under it, and it is flattened since $g'(\xi) < 1$ around $\xi = 0$.

The opposite happens at $ct/L = 0.5$, when the Gaussian is located at $L/2 \mod L$.

Once again, $E2$ gives worse results than $E4$ and $I4$, which are almost indistinguishable.

5 Wave packet

Now, we consider a different initial condition, with $k_p = \frac{2\pi}{L} \cdot 16$:

$$\tilde{u}(x, 0) = U \cos(k_p\, x) \exp\left(-\frac{x^2}{\sigma^2}\right) = \cos(k_p\, x) u(x, 0)$$

Using the properties of the Fourier Transform, we know that a product in the time domain is a convolution in the frequency domain, up to a scaling factor $\frac{1}{2\pi}$:

$$\begin{aligned}
\mathcal{F}\big(\tilde{u}(x,0)\big) &= \frac{1}{2\pi}\mathcal{F}\big(\cos(k_p\, x)\big) * \mathcal{F}\big(u(x,0)\big) \\
&= \frac{1}{2}\Big[\delta(k-k_p) + \delta(k+k_p)\Big] * \left[U\sqrt{\pi}\sigma \exp\left(-\frac{k^2\sigma^2}{4}\right)\right] \\
&= \frac{1}{2}U\sqrt{\pi}\sigma\left[\exp\left(-\frac{(k-k_p)^2\sigma^2}{4}\right) + \exp\left(-\frac{(k+k_p)^2\sigma^2}{4}\right)\right]
\end{aligned}$$

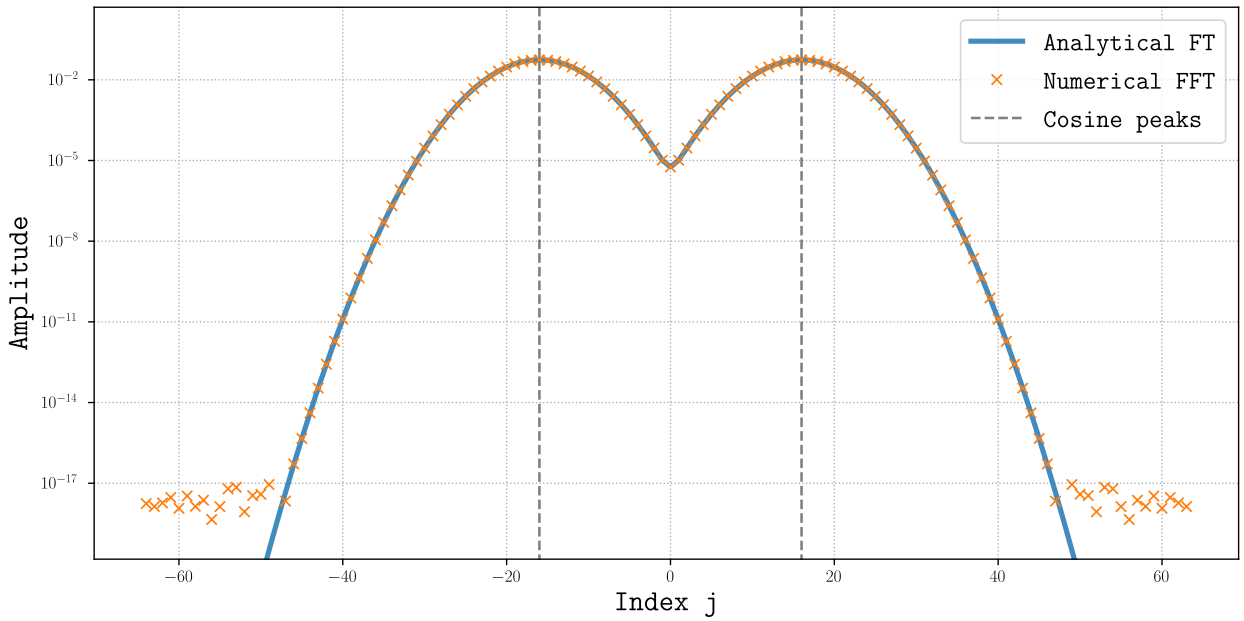

Figure 6. Amplitude of the Fourier transform and the FFT of the "wave packet" initial function.

As a result of the convolution, the Gaussian in the Fourier space is divided in two, as shown in Figure 6. Its maximum located at the wavenumber k_p of the cosine function. The signal is shifted towards higher wavenumbers compared to the first initial condition $u(x, 0)$ we considered.

Before showing the results of the simulation, we compute the group velocities c_g^* for each scheme as a function of the wavelength. As reminded in the homework statement, the group velocity satisfies the relation $c_g^* = c\,\frac{\mathrm{d}k^*}{\mathrm{d}k}$ and is equal to zero at $kh = k_m h$.

Therefore, the modes with wavelengths exceeding k_m have a negative group velocity. In order to keep a numerical method of quality, we need to ensure that h is small enough such that the wavenumbers stay below k_m.

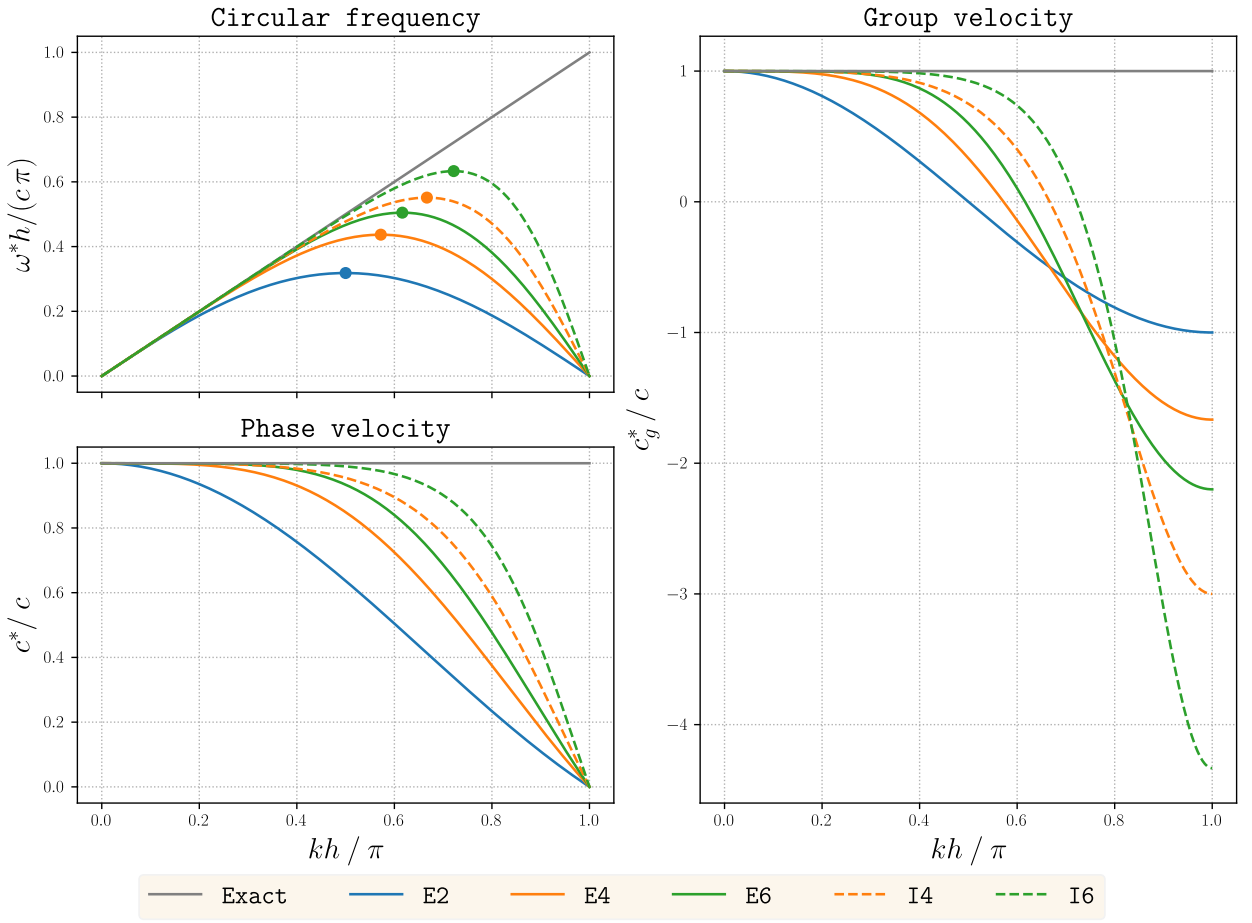

Figure 7. Circular frequency, phase and group velocities as functions of the wavenumber.

Scheme	c_g^* as a function of kh	$k_m h/\pi$	minimum value
E2	$\cos(kh)$	0.5	-1
E4	$\left[4\cos(kh) - \cos(2kh)\right]/3$	0.572	-1.666
E6	$\left[15\cos(kh) - 6\cos(2kh) + \cos(3kh)\right]/10$	0.616	-2.2
I4	$3\left[1 + 2\cos(kh)\right]/\left[2 + \cos(kh)\right]^2$	0.666	-3
I6	$\frac{\left[28\sin(kh) + \sin(2kh)\right]\sin(kh) + \left[3 + 2\cos(kh)\right]\left[14\cos(kh) + \cos(2kh)\right]}{3\left[3 + 2\cos(kh)\right]^2}$	0.721	-4.333

Table 2. Group velocity of the different schemes.

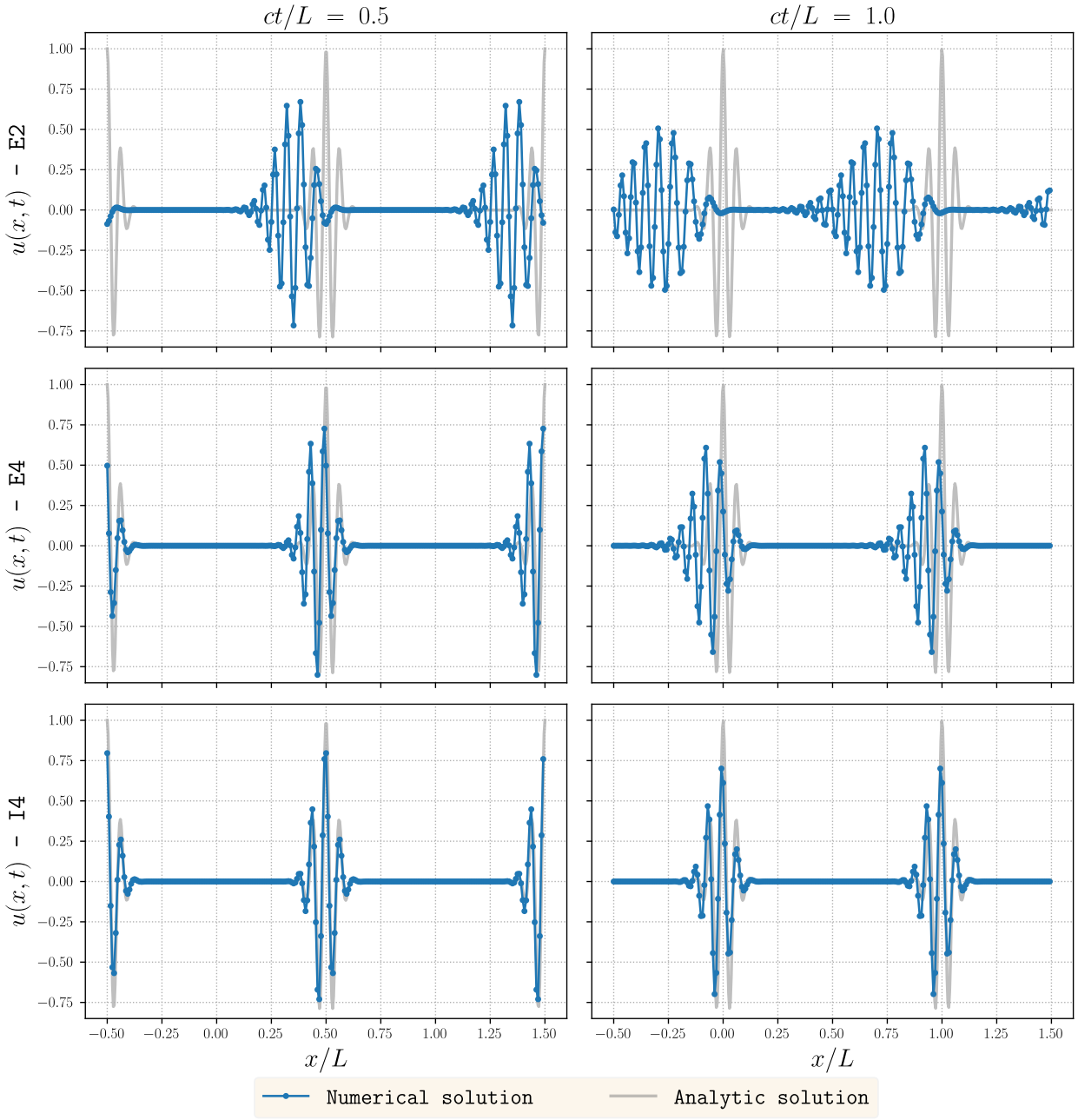

Figure 8. Solution of the transport equation with a wave packet as initial condition.

In the case of the E2 scheme, the principal mode with $j = 16$ has a group velocity of only 71% of the velocity c since $\cos(2\pi\frac{16}{128}) = 1/\sqrt{2} \approx 0.71$. This is the reason why it has only traveled until $x/L \approx 0.71$ after $ct/L = 1$ in the upper right diagram of Figure 8 while the true wave packet has traveled until $x = L$.

This effect is of course smaller for the other schemes, since they behave much better at this wavelength: $v_g^* = 94\%$ of c for E4 and $v_g^* = 98.8\%$ of c for I4.

6 Negative group velocity with the E6 scheme

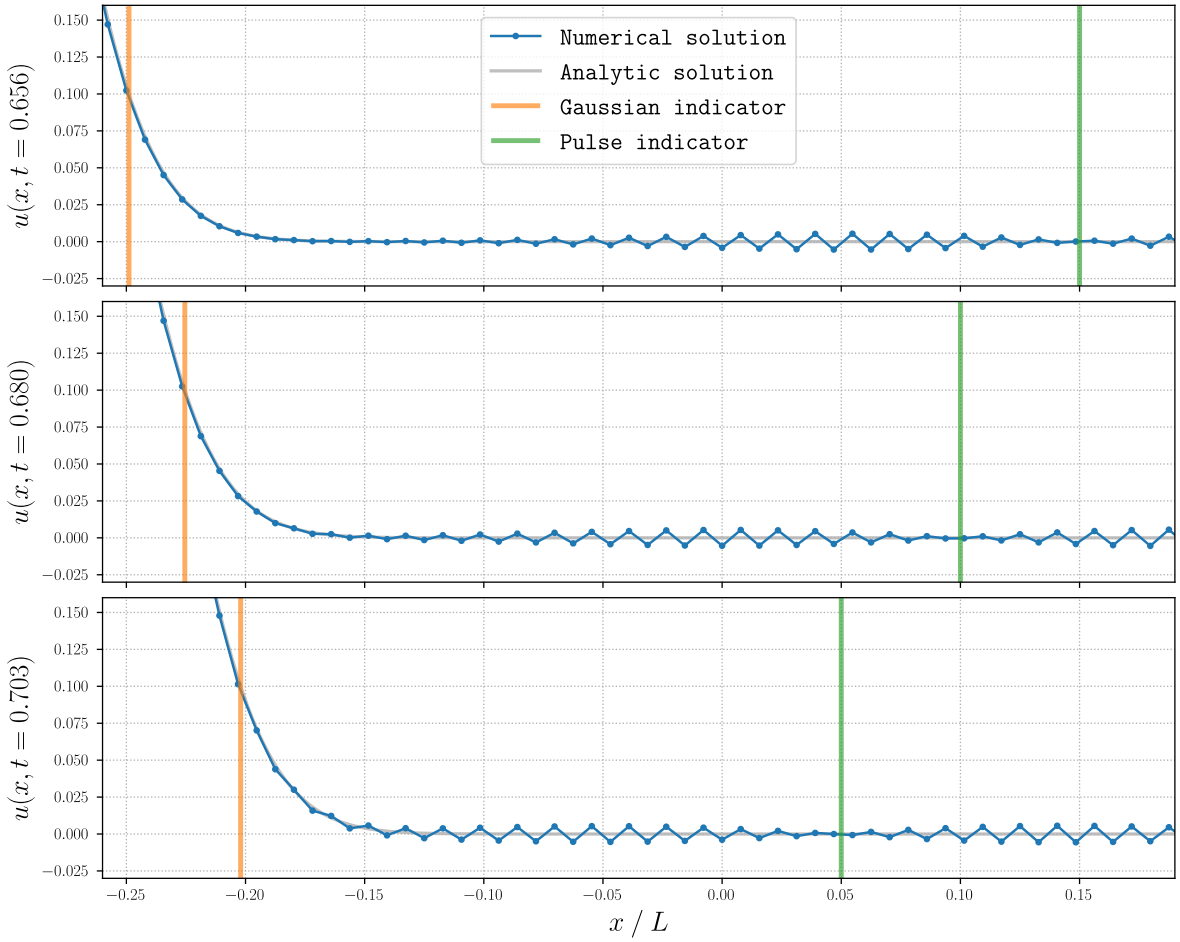

Figure 9. test of caption

In Figure 9, we observe that the solution contains a noise at the *flip-flop* mode. Since its wavenumber is maximal, its group velocity is minimal: $v_g^* = -2.2$ is given in Table 2 and in Figure 7 for the E6 scheme.

On one hand, the noise (indicated by the green line) travels a distance $\Delta x/L \approx -0.1 = 0.05 - 0.15$. On the other hand, the Gaussian has only traveled a distance $\Delta x/L \approx 0.047 = -0.202 - (-0.249)$. The ratio between both distances is ≈ 2.13 which is close to the 2.2 expected.

Without full certainty, we can safely assume that this phenomenon appears with the sixth-order schemes (E6 and I6) because we only use a fourth-order temporal scheme. In this case, the error is in $\mathcal{O}(\Delta t^4) = \mathcal{O}(\Delta x^4)$ since $\Delta x = c\Delta t/CFL$. In light of this, sixth-order schemes in space should not be used with RK4C.

References

[1] G. Winckelmans. *LMECA2660 - Finite differences convection diffusion [Course notes].* EPL, 2018.

[2] G. Winckelmans. *LMECA2660 - Time integration schemes [Course notes].* EPL, 2022.

7 Addendum 2023

The characteristics of the D3 scheme are enumerated below, with the $CFL = \frac{c\Delta t}{h}$:

$$\left.\frac{\partial u}{\partial x}\right|_i = \frac{u_{i-2} - 6u_{i-1} + 3u_i + 2u_{i+1}}{6h} + 2h^3\frac{\partial^4 u}{\partial x^4} + \mathcal{O}(h^4)$$

$$k^*h = \big[8\sin(kh) - \sin(2kh)\big]/6 + \imath\big[-3 + 4\cos(kh) - \cos(2kh)\big]/6$$

$$\lambda_j \Delta t = CFL\left[\frac{-3 + 4\cos(kh) - \cos(2kh)}{6} + \imath\,\frac{8\sin(kh) - \sin(2kh)}{6}\right]$$

The real part of the wavenumber k^*h is the same as the one of the E4 scheme. Waves will therefore be dispersed identically by both schemes. However, since it is a decentered scheme, waves will also be dissipated according to the nonzero imaginary part of k^*h. This phenomenon was not very visible with centered schemes since the only dissipation occurs due to the temporal discretization.

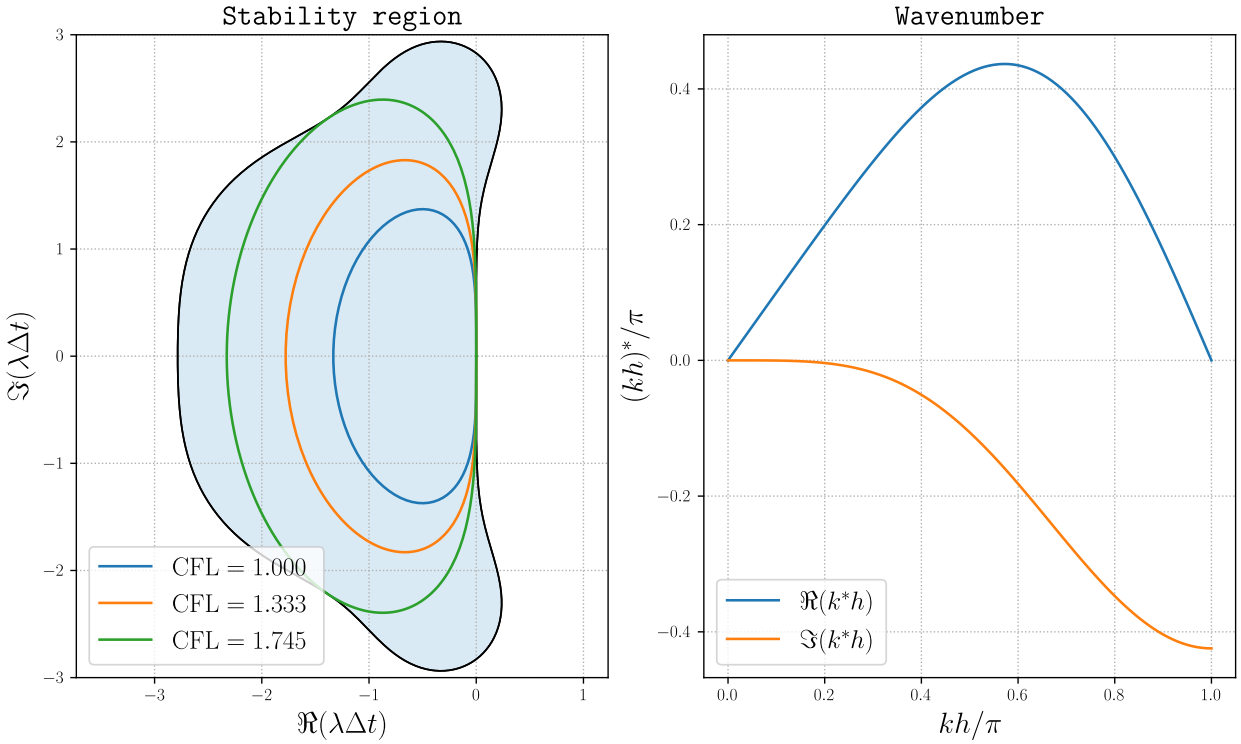

Figure 10. Information about the D3 scheme

Concerning Figure 1 and Figure 6, we want all the waves with nonzero amplitude to be *well treated*: the waves whose Fourier coefficients are larger than $\epsilon_{\text{machine}}$ should have wavenumbers $(kh)^*$ reasonably close to kh.

$$\frac{2\pi j_{\max}}{L} = kh \leq \frac{\pi}{2} \implies N \geq 4j_{\max}$$

This is why the wavepacket is not well enough resolved: $N = 128 < 200$.