Contents

List of Figures

Chapter 1

Theory

1.1 Introduction

The purpose of this report is to numerically solve the one-dimensional viscous Burgers equation with a forcing term:

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = \nu \frac{\partial^2 u}{\partial x^2} + f(x,t) \tag{1.1}$$

This equation is solved with a periodic condition in a domain of length L. The forcing term is also periodic with a random phase Φ (white noise) and acts on the wave numbers k=2 and 3. This nodal forcing term is

$$f(x,t) = \sqrt{\Delta t}\left(\cos(2k_0 x + \Phi_2) + \cos(3k_0 x + \Phi_3)\right) \tag{1.2}$$

with $k_0 = 2\pi/L$ and Δt is the increment in time used in the discretization.

Direct numerical simulations are performed using various high order finite difference and finite element methods.

1.1.1 Smagorinsky model (eddy viscosity)

Large-eddy simulations are also performed by modelling the subgrid terms with a Smagorinsky term (see [1] for further explanations about this topic). In this type of computation, the small spatial scales in the inertial and dissipation ranges are not resolved. Their effect are taken into account by an additional dissipation mechanism, here the turbulent eddy viscosity. The aim of this supplemental dissipation is to avoid the pile-up of energy near the cut-off wavenumber imposed by the coarse grid size.

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = \nu \frac{\partial^2 u}{\partial x^2} + f(x,t) + S(u) \tag{1.3}$$

where the source term for the unresolved scales is given by

$$S(u) = \frac{\partial}{\partial x}\left(\nu_e \frac{\partial u}{\partial x}\right) \tag{1.4}$$

and the eddy viscosity ν_e manifests itself in the following expression for the Smagorinsky model

$$\nu_e = (C_s \delta)^2 \left|\frac{\partial u}{\partial x}\right| \tag{1.5}$$

Herein δ is formally defined by the filter width and is usually set to the mesh size. The coefficient C_s is the Smagorinsky constant and has usually values between 0.1 and 1.

1.1.2 Dynamic Smagorinsky model (eddy viscosity)

As it will be highlighted later, the constant C_s is not single-valued and depends strongly on the grid size, the nature of the spatial discretization and on the viscosity. A dynamic self adaptive estimation of the constant C_s is also possible in the least square. Let us denote the averaging operator

$$\langle f \rangle = \frac{1}{L} \int_0^L f \, \mathrm{d}x \tag{1.6}$$

and $\tilde{f}$ a filter operation (defined in Section 1.1.3). Then the dynamic Smagorinsky constant is given by

$$(C_s \Delta x)^2 = -\frac{1}{2} \frac{\langle L\, M \rangle}{\langle M^2 \rangle} \tag{1.7}$$

where

$$L = \widetilde{u^2} - \tilde{u}\tilde{u} \tag{1.8}$$

$$M = \kappa^2 \frac{\partial \tilde{u}}{\partial x} \left| \frac{\partial \tilde{u}}{\partial x} \right| - \widetilde{\frac{\partial u}{\partial x} \left| \frac{\partial u}{\partial x} \right|} \tag{1.9}$$

and $\kappa = \tilde{\delta}/\delta$ is the filter ratio (taken equal to 2 in the present work). This way to compute the Smagorinsky constant ensures that the average error between the Burgers equation and the filtered Burgers equation is minimized in the least square sense.

1.1.3 Low-pass filter operators

Low-pass filters play a central role in LES where they are used to define the large scales. Low-pass binomial Padé filters and binomial filters from Maulik and San [1] are implemented in the present work. These smoothing functions are based on an averaging operator which requires information from a stencil whose size depends on the order of the chosen binomial filter. The implemented filters are

$$^{(1,1)}B := \tilde{f}_i = \frac{f_{i-1} + 2f_i + f_{i+1}}{4} \tag{1.10}$$

$$^{(2,1)}B := \tilde{f}_i = \frac{-f_{i-2} + 4f_{i-1} + 10f_i + 4f_{i+1} - f_{i+2}}{16} \tag{1.11}$$

$$^{(3,1)}B := \tilde{f}_i = \frac{f_{i-3} - 6f_{i-2} + 15f_{i-1} + 44f_i + 15f_{i+1} - 6f_{i+2} + f_{i+3}}{64} \tag{1.12}$$

$$^{(4,1)}B := \tilde{f}_i = \frac{-f_{i-4} + 8f_{i-3} - 28f_{i-2} + 56f_{i-1} + 186f_i + 56f_{i+1} - 28f_{i+2} + 8f_{i+3} - f_{i+4}}{256} \tag{1.13}$$

$$\text{Padé} \;:=\; \alpha \tilde{f}_{i-1} + \tilde{f}_i + \alpha \tilde{f}_{i+1} = \sum_{s=0}^{3} \frac{a_s}{2} \left(f_{i-s} + f_{i+s} \right) \tag{1.14}$$

$$a_0 = \frac{11 + 10\alpha}{16} \quad a_1 = \frac{15 + 34\alpha}{32} \quad a_2 = \frac{-3 + 6\alpha}{16} \quad a_3 = \frac{1 - 2\alpha}{32}$$

The parameter α in the Padé filter belongs to the range $]-0.5, 0.5[$ and controls the dissipation power of the smoothing filter. The transfer function $T(k)$ is positive within the stated range of α and this ensures the well-posedness of the LES model.

A Fourier analysis (see Section 1.6 for how to perform it) is carried out to study the behaviour of these smoothing functions in the wavenumber space. Using the standard modified wavenumber analysis, a transfer function $T(k)$ can be determined that correlates the Fourier coefficients of

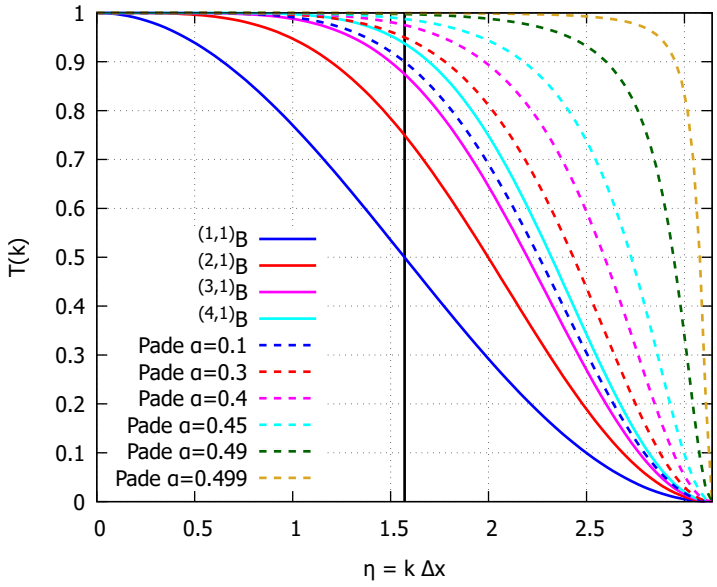


Figure 1.1: Transfer functions of the low-pass binomial and Padé filters used in the subgrid models. The vertical line represents the Fourier cut-off filter at $k\Delta x = \pi/2$.

the smoothed and unsmoothed variables as $\tilde{f} = T(k)f$. The transfer functions for the binomial filters ${}^{(p,1)}B$ are expressed as

$$T(k) = 1 - \left(1 - \frac{1}{2}\left(1 + \cos(k\Delta x)\right)\right)^p \tag{1.15}$$

The transfer function of the Padé filter is given by

$$T(k) = \frac{\sum_{s=0}^{3} a_s \cos(sk\Delta x)}{1 + 2\alpha\cos(k\Delta x)} \tag{1.16}$$

As shown on Fig. 1.1 all the filter function attenuates the effects of the highest wavenumber η completely. The transfer function of the Fourier cut-off filter is also shown for comparison purposes. This cut-off filter removes small scales with wavenumbers $k\Delta x \geq \pi/2$. Small stencil binomial filter functions dissipates more in the mid-range than big stencil filters. The Padé filter happens to attenuate the effects of the highest wavenumber and has low dissipation in the inertial zone. This effectiveness comes at the cost of the requirement to solve a tridiagonal system of equations. Note that the transfer function of the Padé filter with $\alpha = -0.5$ is equivalent to ${}^{(2,1)}B$, $\alpha = 0$ is equivalent to ${}^{(3,1)}B$ and $\alpha = 0.2$ is similar in shape to ${}^{(4,1)}B$.

1.2 Finite differences method

1.2.1 Energy conservative scheme of second order (Hc2)

Centered finite difference schemes of second order are used for the first and second derivatives. The second derivative is computed through

$$\frac{\partial^2 u}{\partial x^2}_{x=x_i} \simeq \frac{u_{i-1} - 2u_i + u_{i+1}}{\Delta x^2} \tag{1.17}$$

The nonlinear convective term is discretized by the skew-symmetric form Hc2 = (2Hd2 + Ha2)/3, which is a combination of two energy non-conservative schemes. These two forms are based on

respectively the divergence (Hd2) and advective (Ha2) formulation of the nonlinear term. Both these schemes are based on a second order centered finite difference scheme for the first derivative. These three schemes are given by

$$\text{Divergence form Hd2 :} \qquad \frac{1}{2}\frac{\partial u^2}{\partial x}_{x=x_i} \simeq \frac{u_{i+1}^2 - u_{i-1}^2}{4\Delta x} \tag{1.18}$$

$$\text{Advective form Ha2 :} \qquad \frac{1}{2}\frac{\partial u^2}{\partial x}_{x=x_i} = u\frac{\partial u}{\partial x}_{x=x_i} \simeq u_i\frac{u_{i+1} - u_{i-1}}{2\Delta x} \tag{1.19}$$

$$\text{Skew-symmetric form Hc2 :} \qquad \frac{1}{2}\frac{\partial u^2}{\partial x}_{x=x_i} \simeq \frac{(u_{i+1} + u_i + u_{i-1})(u_{i+1} - u_{i-1})}{6\Delta x} \tag{1.20}$$

It can be shown that both the Hd2 and Ha2 schemes are not energy conservative by multiplying the nonlinear term at node i by the nodal velocity value u_i at the same node and summing up all the contributions. The energy production by the nonlinear term for the energy dissipative form Hd2 is

$$\begin{aligned} E_{\text{Hd2}}^{\text{nonlinear}} &= ... + h\frac{u_{i-1}}{2}\frac{\partial u^2}{\partial x}_{x=x_{i-1}} + h\frac{u_i}{2}\frac{\partial u^2}{\partial x}_{x=x_i} + h\frac{u_{i+1}}{2}\frac{\partial u^2}{\partial x}_{x=x_{i+1}} + ... \\ &= \frac{h}{4\Delta x}\left(... + u_{i-1}(u_i^2 - u_{i-2}^2) + u_i(u_{i+1}^2 - u_{i-1}^2) + u_{i+1}(u_{i+2}^2 - u_i^2) + ...\right) \\ &= \frac{h}{4\Delta x}\left(... + u_i u_{i-1}(u_i - u_{i-1}) + u_{i+1}u_i(u_{i+1} - u_i) + ...\right) \end{aligned} \tag{1.21}$$

On the other hand, the energy production by the nonlinear term for the advective form Ha2 is

$$\begin{aligned} E_{\text{Ha2}}^{\text{nonlinear}} &= ... + h\frac{u_{i-1}}{2}\frac{\partial u^2}{\partial x}_{x=x_{i-1}} + h\frac{u_i}{2}\frac{\partial u^2}{\partial x}_{x=x_i} + h\frac{u_{i+1}}{2}\frac{\partial u^2}{\partial x}_{x=x_{i+1}} + ... \\ &= \frac{h}{2\Delta x}\left(... + u_{i-1}^2(u_i - u_{i-2}) + u_i^2(u_{i+1} - u_{i-1}) + u_{i+1}^2(u_{i+2} - u_i) + ...\right) \\ &= \frac{h}{2\Delta x}\left(... - u_i u_{i-1}(u_i - u_{i-1}) - u_{i+1}u_i(u_{i+1} - u_i) - ...\right) \end{aligned} \tag{1.22}$$

From that, it is clear that the energy production of the nonlinear term for the skew-symmetric form Hc2 = (2Hd2 + Ha2)/3 should be equal to zero.

The subgrid term defined in Eq. (1.4) is discretized by a second order centered finite difference scheme for the first derivatives:

$$S(u)_{x=x_i} = \frac{(C_s\Delta x)^2}{8\Delta x^3}\Big[\,|u_i - u_{i-2}|\,u_{i-2} - (|u_{i+2} - u_i| + |u_i - u_{i-2}|)\,u_i + |u_{i+2} - u_i|\,u_{i+2}\Big] \tag{1.23}$$

1.2.2 Energy conservative scheme of fourth order (Hc4)

Centered finite difference schemes of fourth order are used for the first and second derivatives. The first spatial derivative is discretized by

$$\frac{\partial u}{\partial x}_{x=x_i} \simeq \frac{u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}}{12\Delta x} \tag{1.24}$$

and the second derivative is given by

$$\frac{\partial^2 u}{\partial x^2}_{x=x_i} \simeq \frac{-u_{i-2} + 16u_{i-1} - 30u_i + 16u_{i+1} - u_{i+2}}{12\Delta x^2} \tag{1.25}$$

The nonlinear convective term is discretized by the skew-symmetric form Hc4 = (2Hd4 + Ha4)/3, which is a combination of two energy non-conservative schemes. These two forms are based on

	α	β	a	b	c
Optimal	0.5771439	0.0896406	1.3025166	0.99355	0.03750245
Order 10	0.5	1/20	17/12	101/150	0.01
Order 8	4/9	1/36	40/27	25/54	0
Order 6	1/3	0	$(9+\alpha-20\beta)/6$	$(-9+32\alpha+62\beta)/15$	$(1-3\alpha+12\beta)/10$
Order 4	0	0	$2(\alpha+2)/3$	$(4\alpha-1)/3$	0
Order 2	0	0	1	0	0

Table 1.1: Values of the coefficients α, β, a, b and c for the first derivative.

respectively the divergence (Hd4) and advective (Ha4) formulation of the nonlinear term and are given by

$$\text{Divergence form Hd4 :} \qquad \frac{1}{2}\frac{\partial u^2}{\partial x}_{x=x_i} \simeq \frac{u_{i-2}^2 - 8u_{i-1}^2 + 8u_{i+1}^2 - u_{i+2}^2}{24\Delta x} \tag{1.26}$$

$$\text{Advective form Ha4 :} \qquad \frac{1}{2}\frac{\partial u^2}{\partial x}_{x=x_i} = u\frac{\partial u}{\partial x}_{x=x_i} \simeq u_i\frac{u_{i-2} - 8u_{i-1} + 8u_{i+1} - u_{i+2}}{12\Delta x} \tag{1.27}$$

In the same way as it has been done in the previous section, it can be shown that both these schemes are not energy conservative by multiplying the nonlinear term at node i by the nodal velocity value u_i at the same node and summing up all the contributions. However, the errors in the kinetic energy associated to the divergence and advective schemes have similar forms. It is thus possible to combine these two schemes to generate a kinetic conserving scheme Hc4 = (2Hd4 + Ha4)/3.

The subgrid term defined in Eq. (1.4) is discretized by the fourth order centered finite difference scheme for the first derivative mentioned before:

$$S(u)_{x=x_i} = \frac{(C_s\Delta x)^2}{12\Delta x}\Big[\,|f_{i-2}|\,f_{i-2} - 8\,|f_{i-1}|\,f_{i-1} + 8\,|f_{i+1}|\,f_{i+1} - |f(x_{i+2})|\,f_{i+2}\Big] \tag{1.28}$$

where the function f is the derivative $\partial u/\partial x$ in Eq. (1.24) at the nodes corresponding to the subindices.

1.2.3 Compact finite difference schemes with spectral-like resolution

These schemes allow a better representation of the shorter length scales [2]. The first derivative at node i is computed through the resolution of a system of equations:

$$\beta\frac{\partial f}{\partial x}_{i-2} + \alpha\frac{\partial f}{\partial x}_{i-1} + \frac{\partial f}{\partial x}_{i} + \alpha\frac{\partial f}{\partial x}_{i+1} + \beta\frac{\partial f}{\partial x}_{i+2} = \frac{c}{6\Delta x}\left(f_{i+3} - f_{i-3}\right) + \frac{b}{4\Delta x}\left(f_{i+2} - f_{i-2}\right) + \frac{a}{2\Delta x}\left(f_{i+1} - f_{i-1}\right) \tag{1.29}$$

The coefficients α, β, a, b and c are derived by matching the Taylor series coefficients of various orders. Note that the second and fourth orders match the central schemes Hc2 and Hc4 introduced in the previous sections. As for the standard central schemes, a skew-symmetric discretization is used for the nonlinear term. This term is chosem equal to

$$u\frac{\partial u}{\partial x}_{\text{skew-symmetric}} = \frac{1}{3}\left(u\frac{\partial u}{\partial x} + \frac{\partial u^2}{\partial x}\right) \tag{1.30}$$

and it has been verified that the energy production of this term is nearly equal to 0 in the case of the forced Burgers equation, for all the orders given in Table 1.1.

	α	β	a	b	c
Optimal	0.50209266	0.05569169	0.21564935	1.723322	0.17659730
Order 10	334/899	43/1798	1065/1798	1038/899	79/1798
Order 8	344/1179	$(38\alpha-9)/214$	$(696-1191\alpha)/428$	$(2454\alpha-294)/535$	0
Order 6	2/11	0	$(6-9\alpha-12\beta)/4$	$(-3+24\alpha-6\beta)/5$	$(2-11\alpha+124\beta)/20$
Order 4	0	0	$4(1-\alpha)/3$	$(10\alpha-1)/3$	0
Order 2	0	0	1	0	0

Table 1.2: Values of the coefficients α, β, a, b and c for the second derivative.

The second derivative at node i is computed through the resolution of another matrix system:

$$\beta\frac{\partial^2 f}{\partial x^2}_{i-2} + \alpha\frac{\partial^2 f}{\partial x^2}_{i-1} + \frac{\partial^2 f}{\partial x^2}_{i} + \alpha\frac{\partial^2 f}{\partial x^2}_{i+1} + \beta\frac{\partial^2 f}{\partial x^2}_{i+2}$$
$$= \frac{c}{9\Delta x^2}\left(f_{i+3} - 2f_i + f_{i-3}\right) + \frac{b}{4\Delta x^2}\left(f_{i+2} - 2f_i + f_{i-2}\right) + \frac{a}{\Delta x^2}\left(f_{i+1} - 2f_i + f_{i-1}\right) \quad (1.31)$$

Note that the second and fourth orders match the central schemes Hc2 and Hc4 introduced in the previous sections.

The subgrid term defined in Eq. (1.4) is discretized by solving the system of equations for the first derivatives as mentioned before:

$$S(u)_{x=x_i} = \Big[A\Big]^{-1}(C_s\Delta x)^2\Big[\frac{c}{6\Delta x}\Big(|f_{i+3}|\,f_{i+3} - |f_{i-3}|\,f_{i-3}\Big) + \frac{b}{4\Delta x}\Big(|f_{i+2}|\,f_{i+2} - |f_{i-2}|\,f_{i-2}\Big)$$
$$+ \frac{a}{2\Delta x}\Big(|f_{i+1}|\,f_{i+1} - |f_{i-1}|\,f_{i-1}\Big)\Big] \quad (1.32)$$

where the function f is the derivative $\partial u/\partial x$ at the nodes corresponding to the subindices. This function is already computed during the set up of the nonlinear convective term. The additional computational requirement comes from the solving of the system of equations where $[A]$ is the pentadiagonal matrix in Eq. (1.29) containing the coefficients α and β.

1.2.4 Nonlinear discretization schemes

As discussed in [3], nonlinear discretization schemes are an interesting option to smooth out the numerical oscillations appearing around shocks or regions of high gradients. Several schemes exist and are based on the reduction of the order of the discretization near regions of high gradients.

The oldest operators are the upwind schemes which involves a discretization stencil biased in the direction determined by the sign of the convection speed. The main drawback of this method is the excessive numerical dissipation.

Another class of nonlinear operators are the Total Variation Diminishing (TVD) schemes. They are based on slope limiters which ensures the monotonicity of the solution at all times. The second order central discretization is switched to a first or second order upwind discretization in the region of high gradients. The main disadvantage is that the upwind scheme is also engaged near extrema, which leads to unnecessary dissipation.

ENO (Essential Non-Oscillary) and WENO (Weighted ENO) schemes were introduced to obtain a higher order of accuracy. These schemes basically select the smoothest interpolating polynomial from a hierarchy of candidates with varying stencil supports. Then, a smooth approximation of the derivatives can be constructed. As observed by [4], these schemes are linearly unstable when coupled with one or two-steps Runge-Kutta time integrators but are linearly stable for three or four-steps integrators. Moreover these schemes generate numerical small-scale

turbulence (see [5]) which pollutes the tail of the energy-spectrum and thus limit the resolution range. Improvements are observed when the filters are applied on the solution.

The class of Dynamic Finite Difference (DFD) schemes was introduced to achieve higher accuracy by minimizing the dispersion error and not through the smoothness of the solution as done by the other schemes.

As demonstrated by [3], nonlinear operators do not ensure the conservation of kinetic energy: TVD and WENO operators are over-dissipative whereas the DFD scheme lacks dissipation. Moreover the nonlinear mechanism in these schemes leads to an erratic behaviour of the modified wavenumber in the entire wavenumber range $\eta \in [0, \eta_{\max}]$ (see section 1.6). This error on the modified wavenumber led to a severe reduction of accuracy in the low-wavenumber range (reduction to 1st-order). We see at this point that nonlinear operators offer mitigating results and this is why we will not put the stress on these methods in the following.

1.2.4.1 1st order UP1, 2nd order UP2 and 3rd order UP3 upwind schemes

Let's denote the transport velocity at the intermediate positions $i \pm 1/2$ by

$$u_{i+1/2} = \frac{u_i + u_{i+1}}{2} \qquad\qquad u_{i-1/2} = \frac{u_i + u_{i-1}}{2} \tag{1.33}$$

and $u_i^+ = \max(u_i, 0)$ and $u_i^- = \min(u_i, 0)$ are the positive and negative contributions of the transport velocity. The 1st, 2nd and 3rd order upwind discretization of the nonlinear term in the skew-symmetric form at position i are expressed respectively by

$$Skew_{x=x_i}^{UP1} = \frac{-u_{i-1}\left(u_i^+ + u_{i-1/2}^+\right) + u_i\left(u_i^+ + u_{i+1/2}^+ - u_i^- - u_{i-1/2}^-\right) + u_{i+1}\left(u_i^- + u_{i+1/2}^-\right)}{3\Delta x} \tag{1.34}$$

$$\begin{aligned} Skew_{x=x_i}^{UP2} = &\frac{\left(u_i^+ + u_{i+1/2}^+\right)(3u_i - u_{i-1}) + \left(u_i^- + u_{i+1/2}^-\right)(3u_{i+1} - u_{i+2})}{6\Delta x} \\ &- \frac{\left(u_i^+ + u_{i-1/2}^+\right)(3u_{i-1} - u_{i-2}) + \left(u_i^- + u_{i-1/2}^-\right)(3u_i - u_{i+1})}{6\Delta x} \end{aligned} \tag{1.35}$$

$$\begin{aligned} Skew_{x=x_i}^{UP3} = &\frac{\left(u_i^+ + u_{i+1/2}^+\right)(2u_{i+1} + 5u_i - u_{i-1}) + \left(u_i^- + u_{i+1/2}^-\right)(2u_i + 5u_{i+1} - u_{i+2})}{18\Delta x} \\ &- \frac{\left(u_i^+ + u_{i-1/2}^+\right)(2u_i + 5u_{i-1} - u_{i-2}) + \left(u_i^- + u_{i-1/2}^-\right)(2u_{i-1} + 5u_i - u_{i+1})}{18\Delta x} \end{aligned} \tag{1.36}$$

With some developments, it can be shown that the upwind schemes are equivalent to a combination of a standard central discretization and a dissipative correction term that ensures stability.

1.2.4.2 Central Dynamic Finite Difference (DFD) scheme

$$\begin{aligned} Skew_{x=x_i} = u_i &\frac{(u_{i+1} + c_{i+1}(u_{i+2} - 2u_{i+1} + u_i)) - (u_{i-1} - c_{i-1}(u_i - 2u_{i-1} + u_{i-2}))}{6\Delta x} \\ &+ \frac{\left(u_{i+1}^2 + c_{i+1}(u_{i+2}^2 - 2u_{i+1}^2 + u_i^2)\right) - \left(u_{i-1}^2 - c_{i-1}(u_i^2 - 2u_{i-1}^2 + u_{i-2}^2)\right)}{6\Delta x} \end{aligned} \tag{1.37}$$

and the dynamic coefficient c is determined by the transported scalar u:

$$c_i = -\frac{1}{6}\left(1 + f\max\left(\min\left(\frac{u_{i+2} - 4u_{i+1} + 6u_i - 4u_{i-1} + u_{i-2}}{u_{i+1} - 2u_i + u_{i-1}}, 0\right), -3\right)\right)^{-1} \tag{1.38}$$

and the parameter $f = 0.2$ is obtained by calibrating the scheme to obtain a maximum accuracy for a field with prescribed inertial range. The DFD scheme is a combination of a central discretization and a dispersive correction term.

1.2.4.3 Total Variation Diminishing (TVD) scheme

$$Skew_{x=x_i} = \frac{(u_i^+ + u_{i+1/2}^+)\left(u_i + \frac{\psi(r_i)}{2}(u_{i+1} - u_i)\right) - (u_i^+ + u_{i-1/2}^+)\left(u_{i-1} + \frac{\psi(r_{i-1})}{2}(u_i - u_{i-1})\right)}{3\Delta x} + \frac{(u_i^- + u_{i+1/2}^-)\left(u_{i+1} + \frac{\psi(1/r_{i+1})}{2}(u_{i+1} - u_i)\right) - (u_i^- + u_{i-1/2}^-)\left(u_i + \frac{\psi(1/r_i)}{2}(u_i - u_{i-1})\right)}{3\Delta x} \tag{1.39}$$

where the transport velocity at the intermediate positions $i \pm 1/2$ and the positive and negative contributions of the transport velocity have the same definition as for the upwind schemes. The ratio of consecutive difference in a node i is given by

$$r_i = \frac{u_i - u_{i-1}}{u_{i+1} - u_i} \tag{1.40}$$

and the slope limiter function is $\psi(r) = \max(\min(r, 2), 0)$. The scheme reduces to a 1st-order upwind scheme in the case of sharp gradients, oppositive slopes or zero gradient ($\psi(r) = 0$). The scheme switches to a 1st-order downwind scheme for $\psi(r) = 2$ and to a 2nd-order centre scheme in the case where $\psi(r) = 1$ (smooth solution, equal slopes). The scheme is equivalent to a 2nd-order upwind scheme when $\psi(r) = r$. The TVD scheme can be reinterpreted as a central scheme combined with a dissipative correction. The limiter function weakens the dissipative correction in comparison with a 2nd-order upwind scheme.

1.3 Galerkin Finite element method

The fundation of the Galerkin finite element formulation is the following. Equation. (1.3) is multiplied by an arbitrary test function $N_i(x)$ (smooth within each element) and is integrated over the whole domain. The overall integral can be decomposed into a summation of integrals over each element:

$$\begin{aligned} &\int_0^L N_i(x)\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) - \nu\frac{\partial^2 u}{\partial x^2} - f(x,t) - S(u)\right)\mathrm{d}x = 0 \\ \Leftrightarrow &\sum_e \int_e N_i(x)\left(\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) - \nu\frac{\partial^2 u}{\partial x^2} - f(x,t) - S(u)\right)\mathrm{d}x = 0 \end{aligned} \tag{1.41}$$

The summation is over all the elements of length Δx and the unknown u is interpolated by polynomials such that

$$u(x,t) = \sum_{j=1}^{N} u_j(t) N_j(x) \tag{1.42}$$

Note that the polynomials used for the interpolation are the same as the ones used to weight the integrated equation. The shape functions are defined such that

$$N_i(x_j) = \delta_{ij} \qquad \text{and} \qquad \sum_i N_i(x) = 1 \tag{1.43}$$

The discretization is inserted in the Galerkin finite element definition and the viscous terms are integrated by parts in order to reduce the second derivative to a first one. One gets

$$M_{ij}\frac{\mathrm{d}u_j}{\mathrm{d}t} = -C_{ijk}u_j(t)u_k(t) - K_{ij}u_j(t) + M_{ij}F_j(x,t) + S_i(t) \tag{1.44}$$

where

$$M_{ij} = \int_0^L N_i(x)N_j(x)\mathrm{d}x \tag{1.45}$$

$$K_{ij} = \nu\int_0^L \frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}N_j}{\mathrm{d}x}\mathrm{d}x \tag{1.46}$$

$$C_{ijk} = \int_0^L \frac{N_i}{2}\frac{\mathrm{d}(N_jN_k)}{\mathrm{d}x}\mathrm{d}x \tag{1.47}$$

$$S_i = (C_s\Delta x)^2\int_0^L \frac{\mathrm{d}N_i}{\mathrm{d}x}\left|\frac{\mathrm{d}N_j}{\mathrm{d}x}u_j\right|\left(\frac{\mathrm{d}N_k}{\mathrm{d}x}u_k\right)\mathrm{d}x \tag{1.48}$$

These different matrices depend only on the shape functions and can be analytically evaluated once and for all. These local matrices are evaluated for each element and inserted in the global system of equation for all the nodes. In the next sections, three types of shape functions are considered and their matrices computed. Note that the formulation used for the nonlinear term is not important for the finite element method. Indeed both the divergence and advective formulations introduced in the finite difference methods provide the same discretization for the finite element method.

A major disadvantage for the finite element methods in comparison with the finite diffences and volumes is that a system of equation must be solved in order to compute the solution at the next time step. Due to the local nature of the shape functions, the global mass matrix constructed by taking into account of all the elements is sparse. It is sometimes useful to circumvent this necessity of solving a system of equations by 'lumping' the mass matrix in order to obtain a pure diagonal lumped mass matrix. The error induced in the solution is negligible as shown later in the results.

1.3.1 Linear Lagrange finite element P1

The interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i(t)N_1(x) + u_{i+1}(t)N_2(x) \tag{1.49}$$

and the shape functions are

$$N_1(x) = 1 - \xi$$
$$N_2(x) = \xi$$

with $0 \leq \xi = x/\Delta x \leq 1$ the nondimensional spatial coordinate

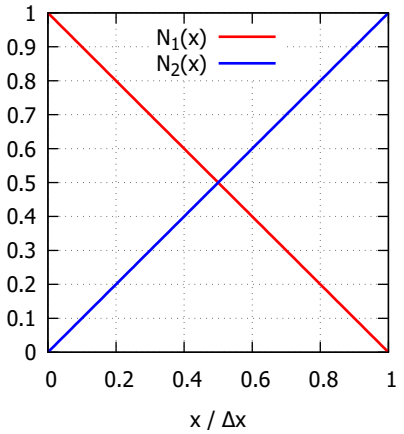


Figure 1.2: Shape functions for the linear Lagrange element

The transposed global unknown vector is

$$u^t = (\quad \cdots \quad u_{i-1} \quad u_i \quad u_{i+1} \quad \cdots \quad)^t \tag{1.50}$$

The local matrices for the element bounded by the nodes i and $i+1$ are given by

$$M_{ij} = \frac{\Delta x}{6} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{1.51}$$

$$K_{ij} = \frac{\nu}{\Delta x} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{1.52}$$

The nonlinear convective term over the element is

$$C_{ijk} u_j u_k = \frac{1}{6} \begin{pmatrix} -2u_i^2 + u_i u_{i+1} + u_{i+1}^2 \\ -u_i^2 - u_i u_{i+1} + 2u_{i+1}^2 \end{pmatrix} \tag{1.53}$$

and the subgrid source term over the element is equal to

$$S_i = C_s^2 \begin{pmatrix} -(u_{i+1} - u_i)\,|u_{i+1} - u_i| \\ (u_{i+1} - u_i)\,|u_{i+1} - u_i| \end{pmatrix} \tag{1.54}$$

1.3.1.1 Lumped mass matrix

The direct lumping technique is used. This technique states that the diagonal components of the lumped mass matrix are equal to the sum of the components on each line ($M_{ii}^L = \sum_j M_{ij}$). The sum over all the diagonal terms of the lumped matrix shoud be equal to the length of the element ($\sum_i M_{ii}^L = \Delta x$). The lumped mass matrix is thus given by

$$M_{ij} = \Delta x \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \tag{1.55}$$

Finally a major observation is that the Lagrange P1 finite element method with lumped mass matrix provides exactly the same discretization as the conservative finite difference scheme Hc2. The spatially discretized equation at node i for these two particular schemes is

$$\frac{\partial u_i}{\partial t} = -\frac{u_{i+1}^2 + u_i(u_{i+1} - u_{i-1}) - u_{i-1}^2}{6\Delta x} + \nu \frac{u_{i-1} - 2u_i + u_{i+1}}{\Delta x^2} + F_i(t) \tag{1.56}$$

1.3.2 Cubic Lagrange finite element P3

The interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i(t)N_1(x) + u_{i+\frac{1}{3}}(t)N_2(x) + u_{i+\frac{2}{3}}(t)N_3(x) + u_{i+1}(t)N_4(x) \tag{1.57}$$

Two intermediate nodes are inserted at $x_{i+\frac{1}{3}} = x_i + \Delta x/3$ and $x_{i+\frac{2}{3}} = x_i + 2\Delta x/3$ in order to build the third order polynomials. For the same number of elements as for the linear Lagrange element, the size of the system of equations is multiplied by three for the cubic Lagrange element.

The shape functions are

$$N_1(x) = \frac{1}{2}(1-\xi)(2-3\xi)(1-3\xi)$$
$$N_2(x) = \frac{9}{2}\xi(1-\xi)(2-3\xi)$$
$$N_3(x) = \frac{9}{2}\xi(1-\xi)(3\xi-1)$$
$$N_4(x) = \frac{1}{2}\xi(2-3\xi)(1-3\xi)$$

with $0 \leq \xi = x/\Delta x \leq 1$ the nondimensional spatial coordinate

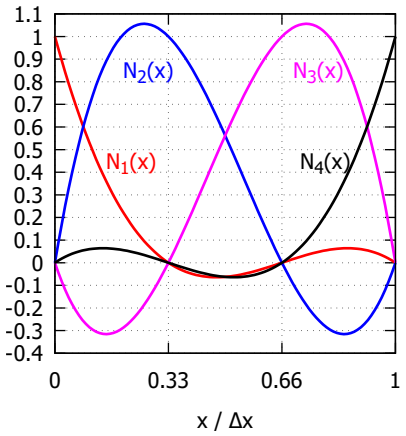


Figure 1.3: Shape functions for the cubic Lagrange element

The transposed global unknown vector is

$$u^t = \begin{pmatrix} \cdots & u_{i-1} & u_{i-\frac{2}{3}} & u_{i-\frac{1}{3}} & u_i & u_{i+\frac{1}{3}} & u_{i+\frac{2}{3}} & u_{i+1} & \cdots \end{pmatrix}^t \tag{1.58}$$

The local matrices for the element bounded by the nodes i and $i+1$ are given by

$$M_{ij} = \frac{\Delta x}{1680}\begin{pmatrix} 128 & 99 & -36 & 19 \\ 99 & 648 & -81 & -36 \\ -36 & -81 & 648 & 99 \\ 19 & -36 & 99 & 128 \end{pmatrix} \tag{1.59}$$

$$K_{ij} = \frac{\nu}{40\Delta x}\begin{pmatrix} 148 & -189 & 54 & -13 \\ -189 & 432 & -297 & 54 \\ 54 & -297 & 432 & -189 \\ -13 & 54 & -189 & 148 \end{pmatrix} \tag{1.60}$$

and the nonlinear convective term over the element for the lines corresponding respectively to the lines of the unknowns u_i, $u_{i+\frac{1}{3}}$, $u_{i+\frac{2}{3}}$ and u_{i+1}

$$C_{ijk}u_ju_k|_{x=x_i} = \frac{1}{6720}\Big(-2240u_i^2 + 1611u_iu_{i+\frac{1}{3}} + 2025u_{i+\frac{1}{3}}^2 - 630u_iu_{i+\frac{2}{3}} - 1053u_{i+\frac{1}{3}}u_{i+\frac{2}{3}} - 162u_{i+\frac{2}{3}}^2 + 139u_iu_{i+1} + 180u_{i+\frac{1}{3}}u_{i+1} - 9u_{i+\frac{2}{3}}u_{i+1} + 139u_{i+1}^2\Big) \tag{1.61}$$

$$C_{ijk}u_ju_k|_{x=x_{i+\frac{1}{3}}} = \frac{3}{2240}\Big(-179u_i^2 - 225u_iu_{i+\frac{1}{3}} + 72u_iu_{i+\frac{2}{3}} + 243u_{i+\frac{1}{3}}u_{i+\frac{2}{3}} + 243u_{i+\frac{2}{3}}^2 - 21u_iu_{i+1} - 18u_{i+\frac{1}{3}}u_{i+1} - 45u_{i+\frac{2}{3}}u_{i+1} - 70u_{i+1}^2\Big) \tag{1.62}$$

$$C_{ijk}u_j u_k|_{x=x_{i+\frac{2}{3}}} = \frac{3}{2240}\Big(70u_i^2 + 45u_i u_{i+\frac{1}{3}} + 18u_i u_{i+\frac{2}{3}} + 21u_i u_{i+1} - 243u_{i+\frac{1}{3}}^2 - 243u_{i+\frac{1}{3}}u_{i+\frac{2}{3}} - 72u_{i+\frac{1}{3}}u_{i+1} + 225u_{i+\frac{2}{3}}u_{i+1} + 179u_{i+1}^2\Big) \quad (1.63)$$

$$C_{ijk}u_j u_k|_{x=x_{i+1}} = \frac{1}{6720}\Big(-139u_i^2 + 9u_i u_{i+\frac{1}{3}} - 180u_i u_{i+\frac{2}{3}} - 139u_i u_{i+1} + 162u_{i+\frac{1}{3}}^2 + 1053u_{i+\frac{1}{3}}u_{i+\frac{2}{3}} + 630u_{i+\frac{1}{3}}u_{i+1} - 2025u_{i+\frac{2}{3}}^2 - 1611u_{i+\frac{2}{3}}u_{i+1} + 2240u_{i+1}^2\Big) \quad (1.64)$$

The subgrid source term is numerically integrated over the element because of the complexity of the integrand:

$$\begin{aligned} S_i &= -(C_s\Delta x)^2 \int_0^{\Delta x} \frac{\partial N_i}{\partial x}\frac{\partial u}{\partial x}\left|\frac{\partial u}{\partial x}\right| \mathrm{d}x \\ &= -(C_s\Delta x)^2 \sum_{j=1}^{NG} w_j \frac{\partial N_i(x_j)}{\partial x}\frac{\partial u(x_j)}{\partial x}\left|\frac{\partial u(x_j)}{\partial x}\right| \end{aligned} \quad (1.65)$$

The number of Gauss integration points NG is fixed to 5 and the weights w_j and point coordinates x_j are

$$w_j = (0.236927 \quad 0.478629 \quad 0.568889) \quad (1.66)$$

$$x_j = \frac{\Delta x}{2}(\xi_j + 1) \quad (1.67)$$

$$\xi_j = (\pm 0.906180 \quad \pm 0.538469 \quad 0.000000) \quad (1.68)$$

1.3.2.1 Lumped mass matrix

The direct lumping technique is used. This technique states that the diagonal components of the lumped mass matrix are equal to the sum of the components on each line ($M_{ii}^L = \sum_j M_{ij}$). The sum over all the diagonal terms of the lumped matrix shoud be equal to the length of the element ($\sum_i M_{ii}^L = \Delta x$). The lumped mass matrix is thus given by

$$M_{ij} = \frac{\Delta x}{1680}\begin{pmatrix} 210 & 0 & 0 & 0 \\ 0 & 630 & 0 & 0 \\ 0 & 0 & 630 & 0 \\ 0 & 0 & 0 & 210 \end{pmatrix} \quad (1.69)$$

1.3.3 Cubic Hermite finite element H3

The interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i(t)H_1^0(x) + \frac{\partial u_i(t)}{\partial x}H_1^1(x) + u_{i+1}(t)H_2^0(x) + \frac{\partial u_{i+1}(t)}{\partial x}H_2^1(x) \quad (1.70)$$

The interpolation by the Hermite third order polynomials is based on the knowledge of the solution and the first derivative of the solution at the nodes. This Hermite interpolation ensures the continuity of the solution up to the first spatial derivative through the H_i^1 functions. The Lagrange interpolation do not ensure the continuity of the derivative of the solution. For the same number of elements as for the linear Lagrange element, the size of the system of equations

is multiplied by two for the cubic Hermite element. The conditions that the Hermite polynomials must respect are

$$H_i^0(x_j) = \delta_{ij} \qquad \frac{\partial H_i^0(x_j)}{\partial x} = 0 \qquad H_i^1(x_j) = 0 \qquad \frac{\partial H_i^1(x_j)}{\partial x} = \delta_{ij} \tag{1.71}$$

The shape functions are

$$\begin{aligned} H_1^0(x_j) &= 1 - 3\xi^2 + 2\xi^3 \\ H_1^1(x_j) &= \Delta x \left(\xi - 2\xi^2 + \xi^3\right) \\ H_2^0(x_j) &= 3\xi^2 - 2\xi^3 \\ H_2^1(x_j) &= \Delta x \left(-\xi^2 + \xi^3\right) \end{aligned}$$

with $0 \leq \xi = x/\Delta x \leq 1$ the nondimensional spatial coordinate

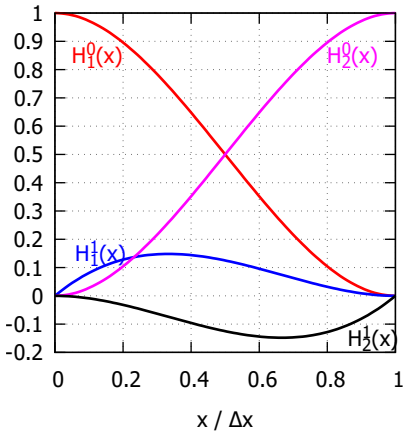


Figure 1.4: Shape functions for the cubic Hermite element

The transposed global unknown vector is

$$u^t = \begin{pmatrix} \cdots & u_{i-1} & \dfrac{\partial u_{i-1}}{\partial x} & u_i & \dfrac{\partial u_i}{\partial x} & u_{i+1} & \dfrac{\partial u_{i+1}}{\partial x} & \cdots \end{pmatrix}^t \tag{1.72}$$

The local matrices for the element bounded by the nodes i and $i+1$ are given by

$$M_{ij} = \frac{\Delta x}{420} \begin{pmatrix} 156 & 22\Delta x & 54 & -13\Delta x \\ 22\Delta x & 4\Delta x^2 & 13\Delta x & -3\Delta x^2 \\ 54 & 13\Delta x & 156 & -22\Delta x \\ -13\Delta x & -3\Delta x^2 & -22\Delta x & 4\Delta x^2 \end{pmatrix} \tag{1.73}$$

$$K_{ij} = \frac{\nu}{30\Delta x} \begin{pmatrix} 36 & 3\Delta x & -36 & 3\Delta x \\ 3\Delta x & 4\Delta x^2 & -3\Delta x & -\Delta x^2 \\ -36 & -3\Delta x & 36 & -3\Delta x \\ 3\Delta x & -\Delta x^2 & -3\Delta x & 4\Delta x^2 \end{pmatrix} \tag{1.74}$$

and the nonlinear convective term over the element for the lines corresponding respectively to the lines of the unknowns u_i, $\partial u_i/\partial x$, u_{i+1} and $\partial u_{i+1}/\partial x$

$$\begin{aligned} C_{ijk}u_j u_k|_{u_i} = \frac{1}{840}\Bigg[&-280u_i^2 + 140u_i u_{i+1} + 140u_{i+1}^2 + 50u_i\frac{\partial u_i}{\partial x}\Delta x - 50u_{i+1}\frac{\partial u_{i+1}}{\partial x}\Delta x - 34u_i\frac{\partial u_{i+1}}{\partial x}\Delta x \\ &+ 34u_{i+1}\frac{\partial u_i}{\partial x}\Delta x + 5\left(\frac{\partial u_i}{\partial x}\Delta x\right)^2 - 8\frac{\partial u_i}{\partial x}\frac{\partial u_{i+1}}{\partial x}\Delta x^2 + 5\left(\frac{\partial u_{i+1}}{\partial x}\Delta x\right)^2\Bigg] \end{aligned} \tag{1.75}$$

$$\begin{aligned} C_{ijk}u_j u_k|_{u_i'} = \frac{\Delta x}{840}\Bigg[&-50u_i^2 + 16u_i u_{i+1} + 34u_{i+1}^2 - 5u_i\frac{\partial u_i}{\partial x}\Delta x - 11u_{i+1}\frac{\partial u_{i+1}}{\partial x}\Delta x \\ &- 3u_i\frac{\partial u_{i+1}}{\partial x}\Delta x + 5u_{i+1}\frac{\partial u_i}{\partial x}\Delta x - \frac{\partial u_i}{\partial x}\frac{\partial u_{i+1}}{\partial x}\Delta x^2 + \left(\frac{\partial u_{i+1}}{\partial x}\Delta x\right)^2\Bigg] \end{aligned} \tag{1.76}$$

$$C_{ijk}u_ju_k|_{u_{i+1}} = \frac{1}{840}\Bigg[-140u_i^2-140u_iu_{i+1}+280u_{i+1}^2-50u_i\frac{\partial u_i}{\partial x}\Delta x+50u_{i+1}\frac{\partial u_{i+1}}{\partial x}\Delta x+34u_i\frac{\partial u_{i+1}}{\partial x}\Delta x$$
$$- 34u_{i+1}\frac{\partial u_i}{\partial x}\Delta x - 5\left(\frac{\partial u_i}{\partial x}\Delta x\right)^2 + 8\frac{\partial u_i}{\partial x}\frac{\partial u_{i+1}}{\partial x}\Delta x^2 - 5\left(\frac{\partial u_{i+1}}{\partial x}\Delta x\right)^2\Bigg] \quad (1.77)$$

$$C_{ijk}u_ju_k|_{u'_{i+1}} = \frac{\Delta x}{840}\Bigg[34u_i^2 + 16u_iu_{i+1} - 50u_{i+1}^2 + 11u_i\frac{\partial u_i}{\partial x}\Delta x + 5u_{i+1}\frac{\partial u_{i+1}}{\partial x}\Delta x$$
$$- 5u_i\frac{\partial u_{i+1}}{\partial x}\Delta x + 3u_{i+1}\frac{\partial u_i}{\partial x}\Delta x + \left(\frac{\partial u_i}{\partial x}\Delta x\right)^2 - \frac{\partial u_i}{\partial x}\frac{\partial u_{i+1}}{\partial x}\Delta x^2\Bigg] \quad (1.78)$$

As done in the case of the cubic Lagrange P3 element, the subgrid term $S(u)$ is approximated by a numerical integration over 5 Gauss points.

1.3.3.1 Lumped mass matrix

A variant of the direct lumping technique is used. The mixing of scales in the original matrix Eq.(1.73) does not allow the simple summation ($M_{ii}^L = \sum_j M_{ij}$) as stated for the Lagrange P3 element. Here the sum takes into account the elements linked with the considered row: for a row linked with the unknown u, the sum takes into account the first and third rows of Eq.(1.73). A row linked with the unknown $\partial u/\partial x$ takes into account the second and fourth columns. The lumped matrix for the cubic Hermite elements reads:

$$M_{ij} = \frac{\Delta x}{420}\begin{pmatrix} 210 & 0 & 0 & 0 \\ 0 & \Delta x^2 & 0 & 0 \\ 0 & 0 & 210 & 0 \\ 0 & 0 & 0 & \Delta x^2 \end{pmatrix} \quad (1.79)$$

1.3.4 Fifth order Hermite finite element H5

The interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i(t)H_1^0(x) + \frac{\partial u_i(t)}{\partial x}H_1^1(x) + \frac{\partial^2 u_i(t)}{\partial x^2}H_1^2(x)$$
$$+ u_{i+1}(t)H_2^0(x) + \frac{\partial u_{i+1}(t)}{\partial x}H_2^1(x) + \frac{\partial^2 u_{i+1}(t)}{\partial x^2}H_2^2(x) \quad (1.80)$$

The interpolation by the fifth order Hermite polynomials is based on the knowledge of the solution, its first derivative and second derivatives at the nodes. This Hermite interpolation ensures the continuity of the solution up to the second spatial derivative through the H_i^1 and H_i^2 shape functions. The Lagrange interpolation do not ensure the continuity of the derivative of the solution. For the same number of elements as for the linear Lagrange element, the size of the system of equations is multiplied by three for the fifth order Hermite element. The conditions that the Hermite polynomials must respect are

$$\begin{array}{lll} H_i^0(x_j) = \delta_{ij} & \dfrac{\partial H_i^0(x_j)}{\partial x} = 0 & \dfrac{\partial^2 H_i^0(x_j)}{\partial x^2} = 0 \\ H_i^1(x_j) = 0 & \dfrac{\partial H_i^1(x_j)}{\partial x} = \delta_{ij} & \dfrac{\partial^2 H_i^1(x_j)}{\partial x^2} = 0 \\ H_i^2(x_j) = 0 & \dfrac{\partial H_i^2(x_j)}{\partial x} = 0 & \dfrac{\partial^2 H_i^2(x_j)}{\partial x^2} = \delta_{ij} \end{array} \quad (1.81)$$

The shape functions are

$$
\begin{aligned}
H_1^0(x_j) &= 1 - 10\xi^3 + 15\xi^4 - 6\xi^5 \\
H_1^1(x_j) &= \Delta x \left(\xi - 6\xi^3 + 8\xi^4 - 3\xi^5\right) \\
H_1^2(x_j) &= \frac{\Delta x^2}{2}\left(\xi^2 - 3\xi^3 + 3\xi^4 - \xi^5\right) \\
H_2^0(x_j) &= 10\xi^3 - 15\xi^4 + 6\xi^5 \\
H_2^1(x_j) &= \Delta x \left(-4\xi^3 + 7\xi^4 - 3\xi^5\right) \\
H_2^2(x_j) &= \frac{\Delta x^2}{2}\left(\xi^3 - 2\xi^4 + \xi^5\right)
\end{aligned}
$$

with $0 \leq \xi = x/\Delta x \leq 1$ the nondimensional spatial coordinate

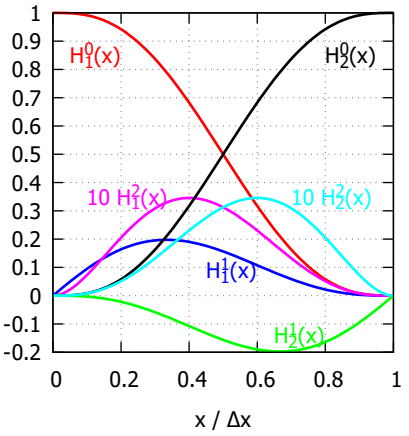


Figure 1.5: Shape functions for the fifth order Hermite element

The transposed global unknown vector is

$$
u^t = \left(\cdots \quad u_{i-1} \quad \frac{\partial u_{i-1}}{\partial x} \quad \frac{\partial^2 u_{i-1}}{\partial x^2} \quad u_i \quad \frac{\partial u_i}{\partial x} \quad \frac{\partial^2 u_i}{\partial x^2} \quad u_{i+1} \quad \frac{\partial u_{i+1}}{\partial x} \quad \frac{\partial^2 u_{i+1}}{\partial x^2} \quad \cdots \right)^t \tag{1.82}
$$

The local matrices for the element bounded by the nodes i and $i+1$ are given by

$$
M_{ij} = \frac{\Delta x}{55440}\begin{pmatrix}
21720 & 3732\Delta x & 281\Delta x^2 & 6000 & -1812\Delta x & 181\Delta x^2 \\
3732\Delta x & 832\Delta x^2 & 69\Delta x^3 & 1812\Delta x & -532\Delta x^2 & 52\Delta x^3 \\
281\Delta x^2 & 69\Delta x^3 & 6\Delta x^4 & 181\Delta x^2 & -52\Delta x^3 & 5\Delta x^4 \\
6000 & 1812\Delta x & 181\Delta x^2 & 21720 & -3732\Delta x & 281\Delta x^2 \\
-1812\Delta x & -532\Delta x^2 & -52\Delta x^3 & -3732\Delta x & 832\Delta x^2 & -69\Delta x^3 \\
181\Delta x^2 & 52\Delta x^3 & 5\Delta x^4 & 281\Delta x^2 & -69\Delta x^3 & 6\Delta x^4
\end{pmatrix} \tag{1.83}
$$

$$
K_{ij} = \frac{\nu}{1260\Delta x}\begin{pmatrix}
1800 & 270\Delta x & 15\Delta x^2 & -1800 & 270\Delta x & -15\Delta x^2 \\
270\Delta x & 288\Delta x^2 & 21\Delta x^3 & -270\Delta x & -18\Delta x^2 & 6\Delta x^3 \\
15\Delta x^2 & 21\Delta x^3 & 2\Delta x^4 & -15\Delta x^2 & -6\Delta x^3 & \Delta x^4 \\
-1800 & -270\Delta x & -15\Delta x^2 & 1800 & -270\Delta x & 15\Delta x^2 \\
270\Delta x & -18\Delta x^2 & -6\Delta x^3 & -270\Delta x & 288\Delta x^2 & -21\Delta x^3 \\
-15\Delta x^2 & 6\Delta x^3 & \Delta x^4 & 15\Delta x^2 & -21\Delta x^3 & 2\Delta x^4
\end{pmatrix} \tag{1.84}
$$

1.3.4.1 Lumped mass matrix

A variant of the direct lumping technique is used. The mixing of scales in the original matrix Eq.(1.83) does not allow the simple summation ($M_{ii}^L = \sum_j M_{ij}$) as stated for the Lagrange P3 element. Here the sum takes into account the elements linked with the considered row: for a row linked with the unknown u, the sum takes into account the first and fourth rows of Eq.(1.83). A row linked with the unknown $\partial u/\partial x$ takes into account the second and fifth columns. A row linked with the unknown $\partial^2 u/\partial x^2$ takes into account the third and sixth columns. The lumped matrix for the fifth order Hermite elements reads:

$$
M_{ij} = \frac{\Delta x}{55440}\begin{pmatrix}
27720 & 0 & 0 & 0 & 0 & 0 \\
0 & 300\Delta x^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 11\Delta x^4 & 0 & 0 & 0 \\
0 & 0 & 0 & 27720 & 0 & 0 \\
0 & 0 & 0 & 0 & 300\Delta x^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 11\Delta x^4
\end{pmatrix} \tag{1.85}
$$

1.4 Discontinuous Galerkin finite element method

The philosophy lying behind the discontinuous Galerkin finite element (DGFE) method is very similar to the continuous one presented in Section 1.3. The major difference is that the DGFE method does not impose the continuity of the solution at the boundary between two adjacent elements. Thus at one node, two values coexist for the solution, one at the left of the node (denoted by the superscript - in the following) and one at the right of the node (denoted by the superscript + in the following).

Because of the similarity between the two finite element methods, their numerical discretization is very similar. Hereunder the main differences are underlined for the advective and diffusive terms.

1.4.1 Advective term

The advective term in Eq. 1.3 is integrated by part in order to obtain a weak formulation. Moreover the following notation is introduced $f(u) = u^2/2$ in order to ease the definition of the flux:

$$\int_e N_i(x)\frac{\partial f(u)}{\partial x}\mathrm{d}x = -\int_e f(u)\frac{\partial N_i(x)}{\partial x}\mathrm{d}x + \left[N_i(x)f(u)\right]_i^{i+1} \tag{1.86}$$

The term between brackets is undetermined at this level because of the two-valued solution at the boundaries of the elements. Similarly to what is done in the finite volume method, a numerical flux $\widehat{f(u)}$ is defined and takes into account both values left and right of the boundaries. A second integration by part reintroduces the strong formulation:

$$\int_e N_i(x)\frac{\partial f(u)}{\partial x}\mathrm{d}x = \underbrace{\int_e N_i(x)\frac{\partial f(u)}{\partial x}\mathrm{d}x}_{C_{ijk}u_ju_k} + \underbrace{\left[N_i(x)\left(\widehat{f(u)} - f(u)\right)\right]_i^{i+1}}_{\Phi_{adv}} \tag{1.87}$$

The first term matches the definition 1.47 in the continuous finite element method. The second term in the brackets $f(u)$ is here evaluated at the internal parts of the element (u_i^+, u_{i+1}^-).

1.4.2 Diffusive term

The diffusive term in Eq. (1.3) is integrated by part in order to obtain a weak formulation. If one naively follows what has been done for the advective term, inconsistency in the numerical scheme are introduced as explained by Shu [6]. Instead the modification introduced by Baumann and Oden [7] is used:

$$\begin{aligned}\nu\int_e N_i(x)\frac{\partial^2 u}{\partial x^2}\mathrm{d}x = &\underbrace{-\nu\int_e \frac{\partial N_i(x)}{\partial x}\frac{\partial u}{\partial x}\mathrm{d}x}_{K_{ij}u_j} \\ &\underbrace{+\nu\left[N_i(x)\widehat{\widehat{\frac{\partial u}{\partial x}}}\right]_i^{i+1} + \frac{\nu}{2}\frac{\partial N_i(x_{i+1}^-)}{\partial x}(u_{i+1}^+ - u_{i+1}^-) + \frac{\nu}{2}\frac{\partial N_i(x_i^+)}{\partial x}(u_i^+ - u_i^-)}_{\Phi_{diff}}\end{aligned} \tag{1.88}$$

The term in brackets in the diffusive flux is evaluated by another numerical flux function $\widehat{\widehat{g(u)}}$ different than the one used in the advective term.

1.4.3 General formulation of the DGFE

Finally the spatially discretized formulation of the DGFE is expressed by

$$M_{ij}\frac{\mathrm{d}u_j}{\mathrm{d}t} = -C_{ijk}u_j(t)u_k(t) - \Phi_{adv} - K_{ij}u_j(t) + \Phi_{diff} + M_{ij}F_j(x,t) + S_i(t) \tag{1.89}$$

In comparison with the equivalent one for the continuous formulation (1.44), one sees that the two methods differs by the advective and diffusive flux terms at the boundaries of the elements. These fluxes are of primary importance because they create a connection between the elements by linking the values left and right of each node.

1.4.4 Definition of the numerical fluxes

The Lax-Friedrichs flux is chosen for the numerical flux $\widehat{f(u)}$ used in the advective term:

$$\widehat{f(u)} = \frac{1}{2}\left(f(u^-) + f(u^+) - \alpha(u^+ - u^-)\right), \qquad \alpha = \max_u\left(\left|f'(u)\right|\right) \tag{1.90}$$

where $f'(u)$ is the derivative of $f(u)$ in relation with u (i.e. $f'(u) = u$ here). Thus $\alpha = \max(|u^-|, |u^+|)$.

Concerning the numerical flux $\widehat{\widehat{g(u)}}$ used in the diffusive term, a central flux is chosen because of the lack of upwinding mechanism in a heat equation:

$$\widehat{\widehat{g(u)}} = \frac{1}{2}\left(g(u^-) + g(u^+)\right) \tag{1.91}$$

1.4.5 Slope limiter

minmod 1

Artificial numerical oscillations can appear around discontinuities in the solution, because the projection onto the finite element basis functions are approximate. To diminish these undesirable oscillations, we apply a slope limiter after every time step (post-processing step) in order to obtain provable total variation stability. In the present study the minmod limiter [8] has been selected. Let us denote the element average of the solution as

$$\bar{u}_i = \frac{1}{\Delta x}\int_{x_i^+}^{x_{i+1}^-} u\mathrm{d}x \tag{1.92}$$

and further denote

$$\tilde{u} = u_{i+1}^- - \bar{u}_i, \qquad \tilde{\tilde{u}} = \bar{u}_i - u_i^+ \tag{1.93}$$

The slope limiter should not change the average $\bar{u}_i$ but may change $\tilde{u}$ and/or $\tilde{\tilde{u}}$. In particular the minmod limiter changes $\tilde{u}$ and $\tilde{\tilde{u}}$ into

$$\tilde{u}^{(mod)} = m(\tilde{u},\, \Delta_+\bar{u}_i,\, \Delta_-\bar{u}_i), \qquad \tilde{\tilde{u}}^{(mod)} = m(\tilde{\tilde{u}},\, \Delta_+\bar{u}_i,\, \Delta_-\bar{u}_i) \tag{1.94}$$

where

$$\Delta_+\bar{u}_i = \bar{u}_{i+1} - \bar{u}_i, \qquad \Delta_-\bar{u}_i = \bar{u}_i - \bar{u}_{i-1} \tag{1.95}$$

and the minmod function m is defined by

$$m(a_1, ..., a_l) = \begin{cases} s\min(|a_1|, ..., |a_l|), & \text{if } s = sign(a_1) = ... = sign(a_l) \\ 0 & \text{otherwise} \end{cases} \tag{1.96}$$

The limited function $u^{(mod)}$ is then recovered to maintain the old element average $\bar{u}_i$ and the new point values $\tilde{u}^{(mod)}$ and $\tilde{\tilde{u}}^{(mod)}$, that is

$$u_{i+1}^{-\,(mod)} = \bar{u}_i + \tilde{u}^{(mod)}, \qquad u_i^{+(mod)} = \bar{u}_i - \tilde{\tilde{u}}^{(mod)} \tag{1.97}$$

minmod 2

The use of a simple slope limiter has the disadvantage that it destroys the high-order accuracy in smooth regions. As underlined by Hesthaven [9] some improvement can be reached by limiting the application of the slope limiter to the elements where oscillations are detected. A smooth region is detected where $|u^{(mod)} - u| \leq \epsilon$ and the unaltered solution u is kept. This improvement is not sufficient around local extrema. One way to address this is to relax the condition on the decay of the total variation and require that the total variation of the mean is just bounded, called the TVBM condition. This can be achieved by slightly modifying the definition of the minmod function as

$$\check{m}(a_1, a_2, ..., a_l) = m(a_1, a_2 + M\Delta x^2 sign(a_2), ..., a_l + M\Delta x^2 sign(a_l)) \tag{1.98}$$

where M is a constant that should be an upper bound on the second derivative at the local extrema. This is naturally not easy to estimate *a priori*. Too small a value of M implies higher local dissipation and order reduction, whereas too high a value of M reintroduces the oscillations.

1.4.6 Linear Lagrange finite element P1

The interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i^+(t)N_1(x) + u_{i+1}^-(t)N_2(x) \tag{1.99}$$

where the shape functions $N_1(x)$ and $N_2(x)$ are identical to Fig. 1.2. The transposed global unknown vector is

$$u^t = \begin{pmatrix} \cdots & u_{i-1}^+ & u_i^- & u_i^+ & u_{i+1}^- & \cdots \end{pmatrix}^t \tag{1.100}$$

The matrices M_{ij} and K_{ij} are identical to Section 1.3.1. The nonlinear advective term over the element reads

$$C_{ijk}u_j u_k = \frac{1}{6}\begin{pmatrix} -2{u_i^+}^2 + u_i^+ u_{i+1}^- + {u_{i+1}^-}^2 \\ -{u_i^+}^2 - u_i^+ u_{i+1}^- + 2{u_{i+1}^-}^2 \end{pmatrix} \tag{1.101}$$

Following the definitions of the numerical fluxes, one gets the expressions of the advective and diffusive fluxes within the element:

$$\begin{aligned} \Phi_{adv} &= \frac{1}{2}\begin{pmatrix} \frac{{u_i^+}^2 - {u_i^-}^2}{2} + \max(|u_i^-|, |u_i^+|)(u_i^+ - u_i^-) \\ \frac{{u_{i+1}^+}^2 - {u_{i+1}^-}^2}{2} - \max(|u_{i+1}^-|, |u_{i+1}^+|)(u_{i+1}^+ - u_{i+1}^-) \end{pmatrix} \\ \Phi_{diff} &= \frac{\nu}{2\Delta x}\begin{pmatrix} u_{i-1}^+ - u_{i+1}^+ \\ u_{i+2}^- - u_i^- \end{pmatrix} \end{aligned} \tag{1.102}$$

1.4.7 Cubic Hermite finite element H3

One could argue that the main advantage of this element (continuity of the derivative across the elements) is lost when used in a discontinuous finite element method. However as we will see later, the results obtained by this element are quite excellent. Another advantage of the Hermite H3 element is still kept: the spatial third order is reached by only doubling the size of

the matrices (contrary to the tripling induced by the cubic Lagrange L3 element). Similarly to Section 1.3.3, the interpolation inside an element bounded by the nodes i and $i+1$ is

$$u(x,t) = u_i^+(t)H_1^0(x) + \frac{\partial u_i^+(t)}{\partial x}H_1^1(x) + u_{i+1}^-(t)H_2^0(x) + \frac{\partial u_{i+1}^-(t)}{\partial x}H_2^1(x) \tag{1.103}$$

where the shape functions H_1^0, $H_1^1(x)$, $H_2^0(x)$ and H_2^1 are identical to Fig. 1.4. The transposed global unknown vector is

$$u^t = \begin{pmatrix} \cdots & u_{i-1}^+ & \dfrac{\partial u_{i-1}^+}{\partial x} & u_i^- & \dfrac{\partial u_i^-}{\partial x} & u_i^+ & \dfrac{\partial u_i^+}{\partial x} & u_{i+1}^- & \dfrac{\partial u_{i+1}^-}{\partial x} & \cdots \end{pmatrix}^t \tag{1.104}$$

The matrices M_{ij} and K_{ij} are identical to Section 1.3.1. The nonlinear advective term over the element for the lines corresponding respectively to the lines of the unknowns u_i^+, $\partial u_i^+/\partial x$, u_{i+1}^- and $\partial u_{i+1}^-/\partial x$

$$\begin{aligned} C_{ijk}u_ju_k|_{u_i^+} = \frac{1}{840}\Bigg[& -280{u_i^+}^2 + 140u_i^+{u_{i+1}}^- + 140{u_{i+1}}^{-2} \\ & + 50u_i^+\frac{\partial u_i^+}{\partial x}\Delta x - 50{u_{i+1}}^-\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x - 34u_i^+\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x + 34{u_{i+1}}^-\frac{\partial u_i}{\partial x}\Delta x \\ & + 5\left(\frac{\partial u_i^+}{\partial x}\Delta x\right)^2 - 8\frac{\partial u_i^+}{\partial x}\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x^2 + 5\left(\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x\right)^2\Bigg] \end{aligned} \tag{1.105}$$

$$\begin{aligned} C_{ijk}u_ju_k|_{\partial u_i^+/\partial x} = \frac{\Delta x}{840}\Bigg[& -50{u_i^+}^2 + 16u_i^+{u_{i+1}}^- + 34{u_{i+1}}^{-2} \\ & - 5u_i^+\frac{\partial u_i^+}{\partial x}\Delta x - 11{u_{i+1}}^-\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x - 3u_i^+\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x + 5{u_{i+1}}^-\frac{\partial u_i^+}{\partial x}\Delta x \\ & - \frac{\partial u_i^+}{\partial x}\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x^2 + \left(\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x\right)^2\Bigg] \end{aligned} \tag{1.106}$$

$$\begin{aligned} C_{ijk}u_ju_k|_{u_{i+1}^-} = \frac{1}{840}\Bigg[& -140{u_i^+}^2 - 140u_i^+{u_{i+1}}^- + 280{u_{i+1}}^{-2} \\ & - 50u_i^+\frac{\partial u_i^+}{\partial x}\Delta x + 50{u_{i+1}}^-\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x + 34u_i^+\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x - 34{u_{i+1}}^-\frac{\partial u_i}{\partial x}\Delta x \\ & - 5\left(\frac{\partial u_i^+}{\partial x}\Delta x\right)^2 + 8\frac{\partial u_i^+}{\partial x}\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x^2 - 5\left(\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x\right)^2\Bigg] \end{aligned} \tag{1.107}$$

$$\begin{aligned} C_{ijk}u_ju_k|_{\partial u_{i+1}^-/\partial x} = \frac{\Delta x}{840}\Bigg[& 34{u_i^+}^2 + 16u_i^+{u_{i+1}}^- - 50{u_{i+1}}^{-2} \\ & + 11u_i^+\frac{\partial u_i^+}{\partial x}\Delta x + 5{u_{i+1}}^-\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x - 5u_i^+\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x + 3{u_{i+1}}^-\frac{\partial u_i^+}{\partial x}\Delta x \\ & + \left(\frac{\partial u_i}{\partial x}\Delta x\right)^2 - \frac{\partial u_i^+}{\partial x}\frac{\partial {u_{i+1}}^-}{\partial x}\Delta x^2\Bigg] \end{aligned} \tag{1.108}$$

For the advective flux within the element, only the H_1^0 and H_2^0 shape functions are non-zero at the boundaries of the element. Thus the flux array is similar to the one obtained for the Lagrange P1 element, with the additional empty lines for the unknowns $\partial u_i^+/\partial x$ and $\partial u_{i+1}^-/\partial x$:

$$\Phi_{adv} = \begin{pmatrix} \Phi_{adv}|u_i^+ \\ \Phi_{adv}|_{\partial u_i^+/\partial x} \\ \Phi_{adv}|u_{i+1}^- \\ \Phi_{adv}|_{\partial u_{i+1}^-/\partial x} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \frac{{u_i^+}^2 - {u_i^-}^2}{2} + \max(|u_i^-|, |u_i^+|)(u_i^+ - u_i^-) \\ 0 \\ \frac{{u_{i+1}^+}^2 - {u_{i+1}^-}^2}{2} - \max(|u_{i+1}^-|, |u_{i+1}^+|)(u_{i+1}^+ - u_{i+1}^-) \\ 0 \end{pmatrix} \tag{1.109}$$

The diffusive flux within the element can be evaluated by considering the conditions (1.71) that the shape functions must satisfy. The major difference with respect to the Lagrange P1 element is that du/dx is now an unknown of the problem.

$$\Phi_{diff} = \begin{pmatrix} \Phi_{diff}|u_i^+ \\ \Phi_{diff}|_{\partial u_i^+/\partial x} \\ \Phi_{diff}|u_{i+1}^- \\ \Phi_{diff}|_{\partial u_{i+1}^-/\partial x} \end{pmatrix} = \frac{\nu}{2} \begin{pmatrix} -\frac{\partial {u_i}^-}{\partial x} - \frac{\partial {u_i}^+}{\partial x} \\ u_i^+ - u_i^- \\ \frac{\partial {u_{i+1}}^-}{\partial x} + \frac{\partial {u_{i+1}}^+}{\partial x} \\ u_{i+1}^+ - u_{i+1}^- \end{pmatrix} \tag{1.110}$$

In comparison with the Lagrange P1 element, the diffusive flux hereabove offers a more compact radius.

1.5 Time integration

1.5.1 Standard Runge-Kutta method

The integration in time used in this project is the well-known fourth-order Runge-Kutta method with four steps RK4. This method allows to reach a high accuracy in time while having a limit on the Courant-Friedrich-Lewy number (defined in Eq. (1.111)) which is higher than other standard methods. This limit is identical for the purely advective and purely viscous forms of the Burgers equation.

$$\mathrm{CFL} = \frac{\max(u)\Delta t}{\Delta x} \lesssim 2.8 \tag{1.111}$$

In the case of a generic equation $\partial u/\partial t = f(u,t)$, the RK4 time integrator provides the next time step solution through

$$\begin{aligned} k_1 &= \Delta t f(u^n, t^n) \\ k_2 &= \Delta t f(u^n + \frac{k_1}{2}, t + \frac{\Delta t}{2}) \\ k_3 &= \Delta t f(u^n + \frac{k_2}{2}, t + \frac{\Delta t}{2}) \\ k_4 &= \Delta t f(u^n + k_3, t + \Delta t) \\ u^{n+1} &= u^n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \end{aligned} \tag{1.112}$$

As demonstrated by Pereira and Pereira [10], the RK4 shows a better stability region than the RK6 scheme if central differences are used in spatial discretization. Moreover the RK4 scheme shows the best spectral resolution among other classical time integrators such as the Adams-Bashforth, Quickest third-order and Leapfrog schemes.

1.5.2 Strong-stability-preserving Runge-Kutta method SSP-RK

Strong-stability-preserving (SSP) time discretization methods have a nonlinear stability property that makes them particularly suitable for the integration of hyperbolic conservation laws where discontinuous behaviour is present. Indeed they guarantee that no additional oscillations are introduced as part of the time-integration process. A third-order SSP-KR method is given by Gottlieb [11]:

$$\begin{aligned} v^1 &= u^n + \Delta t f(u^n) \\ v^2 &= \frac{3}{4}u^n + \frac{1}{4}v^1 + \frac{\Delta t}{4}f(v^1) \\ u^{n+1} = v^3 &= \frac{1}{3}u^n + \frac{2}{3}v^2 + \frac{2}{3}\Delta t f(v^2) \end{aligned} \tag{1.113}$$

A five-stage fourth-order optimal scheme is derived by Spiteri [12]:

$$\begin{aligned} v^1 &= u^n + 0.39175222700392\Delta t f(u^n) \\ v^2 &= 0.44437049406734u^n + 0.55562950593266v^1 + 0.36841059262959\Delta t f(v^1) \\ v^3 &= 0.62010185138540u^n + 0.37989814861460v^2 + 0.25189177424738\Delta t f(v^2) \\ v^4 &= 0.17807995410773u^n + 0.82192004589227v^3 + 0.54497475021237\Delta t f(v^3) \\ u^{n+1} = v^5 &= 0.00683325884039u^n + 0.51723167208978v^2 + 0.12759831133288v^3 \\ &\quad + 34833675773694v^4 + 0.08460416338212\Delta t f(v^3) + 0.22600748319395\Delta t f(v^4) \end{aligned} \tag{1.114}$$

The additional work added by the fifth stage is partially offset by the $[40-80]\%$ improvements in the effective time-step restriction over the most popular fourth-order schemes currently in use.

1.6 Dispersive error

The study of the dispersive error is based on the linear advection equation

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = 0 \tag{1.115}$$

It is supposed that there exists an exact solution $u(x,t) = \hat{u}\exp(Ikx)$. Injecting this exact solution leads to the ordinary differential equation

$$\frac{\partial \hat{u}}{\partial t} = \lambda_{ex}\hat{u} \tag{1.116}$$

with $\lambda_{ex} = Ika$ the exact eigenvalue of the previous equation. The spatial discretization of the linear advection equation introduces a set of equations to be solved on each nodes in order to extract the discretized solution u^h:

$$\frac{\partial u^h}{\partial t} = Au^h \tag{1.117}$$

At the position x_m, the eigenvectors of the matrix A take the form $v_m^j = \exp(Ik_j x_m)$ for each wavenumber $k_j = 2\pi j/L$ with $j = 0 \rightarrow N-1$. For each eigenvector, an eigenvalue λ^j is associated. The exact eigenvalue λ_{ex} and the one coming from the discretization λ^j can then be compared on the basis of the numerical wave propagation speed $a^j = I\lambda^j/k_j = a_r + Ia_i$, which is generally a complex number. The comparison is then

$$\frac{\lambda^j}{\lambda_{ex}} = \frac{a^j}{a} = \frac{a_r^j}{a} + I\frac{a_i^j}{a} \tag{1.118}$$

In the following analysis, only the dispersive error $\epsilon_\Phi = a_r^j/a$ will be considered. The reason is that all the considered discretizations are symmetrical and do not introduce any imaginary part a_i^j. Thus in the present study, no numerical dissipation is studied. A wave of wavenumber k_j propagates at the phase velocity λ^j which is different from the exact phase velocity λ. The phase velocity is approximated through the spatial discretization. In general, the wave propagates more slowly than the exact solution.

1.6.1 Modified wavenumbers for the different spatial discretizations

Finite difference scheme Hc2

The discretized equation reads

$$\frac{\partial u_i}{\partial t} + a\frac{u_{i+1} - u_{i-1}}{2\Delta x} = 0 \tag{1.119}$$

With the asumption of a periodic solution, one can deduce the solution at nodes $i+1$ and $i-1$ from the solution at node i:

$$u_{i+1} = \hat{u}(t)\exp(Ik(x+\Delta x)) = \underbrace{\hat{u}(t)\exp(Ikx)}_{u_i}\exp(I\underbrace{k\Delta x}_{\eta}) = u_i\exp(I\eta) \tag{1.120a}$$

$$u_{i-1} = \hat{u}(t)\exp(Ik(x-\Delta x)) = \underbrace{\hat{u}(t)\exp(Ikx)}_{u_i}\exp(-I\underbrace{k\Delta x}_{\eta}) = u_i\exp(-I\eta) \tag{1.120b}$$

so that the discretized equation is now written as

$$\frac{\partial u_i}{\partial t} + \frac{Ia}{\Delta x}\sin(\eta)u_i = 0 \tag{1.121}$$

Concerning the time dependency, it is assumed that the solution takes the form $u = \hat{u}\exp(-IKt)$, so that

$$I\left(\frac{a}{\Delta x}\sin(\eta) - K\right)u_i = 0 \tag{1.122}$$

If we define $\eta_{num} = K\Delta x/a$ as the nondimensional modified numerical wavenumber that results from the discretization, one gets the expression

$$\eta_{num} = \sin(\eta) \tag{1.123}$$

with $\eta \in [0, \pi]$.

Finite difference scheme Hc4

Similarly to what has been done for the Hc2, one gets the modified wavenumber

$$\eta_{num} = \frac{8\sin(\eta) - \sin(2\eta)}{6} \tag{1.124}$$

with $\eta \in [0, \pi]$.

Compact finite difference schemes with spectral-like resolution

$$\eta_{num} = \frac{a\sin(\eta) + (b/2)\sin(2\eta) + (c/3)\sin(3\eta)}{1 + 2\alpha\cos(\eta) + 2\beta\cos(2\eta)} \tag{1.125}$$

with $\eta \in [0, \pi]$.

Finite element P1

The methodology is identical to the one explained for the finite difference scheme Hc2 but one major difference appears: the time derivatives are also applied on the nodes $i-1$ and $i+1$ because of the mass matrix M_{ij}. The spatially discretized equation is thus written

$$\frac{1}{6}\left(\frac{\partial u_{i-1}}{\partial t}+4\frac{\partial u_i}{\partial t}+\frac{\partial u_{i+1}}{\partial t}\right)+a\frac{u_{i+1}-u_{i-1}}{2\Delta x}=0 \tag{1.126}$$

These times derivatives can be reduced to the time derivative at node i through

$$\frac{\partial u_{i+1}}{\partial t}=\frac{\partial u_i}{\partial t}\exp(I\eta) \qquad\qquad \frac{\partial u_{i-1}}{\partial t}=\frac{\partial u_i}{\partial t}\exp(-I\eta) \tag{1.127}$$

The modified wavenumber is then equal to

$$\eta_{num}=\frac{3\sin(\eta)}{2+\cos(\eta)} \tag{1.128}$$

with $\eta\in[0,\pi]$.

Finite element P3

Similarly to what has been done for the element P1, one gets the modified wavenumber

$$\eta_{num}=21\frac{57\sin(\eta/3)-24\sin(2\eta/3)+7\sin(\eta)}{128+99\cos(\eta/3)-36\cos(2\eta/3)+19\cos(\eta)} \tag{1.129}$$

with $\eta\in[0,3\pi]$ for the original mass matrix and

$$\eta_{num}=\frac{57\sin(\eta/3)-24\sin(2\eta/3)+7\sin(\eta)}{10} \tag{1.130}$$

for the lumped mass matrix. As it will be shown later, the denominator of Eq. (1.129) has a zero in the range $\eta\in[0,3\pi]$. This leads to an infinite value of η_{num} at this zero. Lumping the mass matrix allows to get rid of this zero and thus to keep limited values for η_{num}.

Finite element H3

The discretized equation at node i for the unknowns u_i and $u'_i=\partial u_i/\partial x$ is written as

$$\frac{\Delta x}{420}\begin{pmatrix}54 & 13\Delta x & 312 & 0 & 54 & -13\Delta x\\ -13\Delta x & -3\Delta x^2 & 0 & 8\Delta x^2 & 13\Delta x & -3\Delta x^2\end{pmatrix}\frac{\partial}{\partial t}\begin{pmatrix}u_{i-1}\\ u'_{i-1}\\ u_i\\ u'_i\\ u_{i+1}\\ u'_{i+1}\end{pmatrix}$$
$$+\frac{a}{60}\begin{pmatrix}-30 & -6\Delta x & 0 & 12\Delta x & 30 & -6\Delta x\\ 6\Delta x & \Delta x^2 & -12\Delta x & 0 & 6\Delta x & -\Delta x^2\end{pmatrix}\begin{pmatrix}u_{i-1}\\ u'_{i-1}\\ u_i\\ u'_i\\ u_{i+1}\\ u'_{i+1}\end{pmatrix}=0 \tag{1.131}$$

Then, after using $u = \hat{u}\exp(-IKt)$ and Eqs. (1.120) and (1.127), one gets

$$\begin{pmatrix} I\left(\sin(\eta) - \dfrac{26+9\cos(\eta)}{35}\eta_{num}\right) & \dfrac{\Delta x(1-\cos(\eta))}{5} - \dfrac{13\Delta x\sin(\eta)}{210}\eta_{num} \\ \dfrac{13\Delta x\sin(\eta)}{210}\eta_{num} - \dfrac{\Delta x(1-\cos(\eta))}{5} & -I\left(\dfrac{\Delta x^2(4-3\cos(\eta))}{210}\eta_{num} + \dfrac{\Delta x^2 sin(\eta)}{30}\right) \end{pmatrix} \begin{pmatrix} u_i \\ u_i' \end{pmatrix} = 0 \tag{1.132}$$

The solution is non-trivial $(u_i, u_i') \neq 0$ if the determinant of the previous matrix is equal to zero. This leads to an equation of second order on η_{num}, which is solved for $\eta \in [0, 2\pi]$. The equation on η_{num} is given by

$$\big(-131 + 72\cos(\eta) - \cos(2\eta)\big)\eta_{num}^2 + \big(-384\sin(\eta) + 12\sin(2\eta)\big)\eta_{num} + 966 - 1008\cos(\eta) + 42\cos(2\eta) = 0 \tag{1.133}$$

As shown in Fig. 1.6a, one of the solution is negative and should be rejected, the other one corresponding to the physical situation. To each η corresponds two eigenvectors that can be reconstructed. The three first eigenvectors (for $K = 0$, 1 and 2) are plotted in Fig. 1.6b-1.6d. One sees that one of the two eigenvectors represents the physical behaviour whereas the second mode behaves like a parasite oscillation. This second mode reaches its maximum at the locations where the first mode has the steepest variation. As the modified wavenumber K increases, the amplitude of the first eigenvector diminishes down to reaching the amplitude of the second parasite eigenvector.

When the lumped mass matrix formulation is used (Eq. (1.79)), the idea behind the previous development remains and the solution to the quadratic equation on η_{num} is given by

$$\eta_{num} = -3\sin(\eta) \pm \frac{\sqrt{190\sin^2(\eta) + 420\,(1-\cos(\eta))}}{5} \tag{1.134}$$

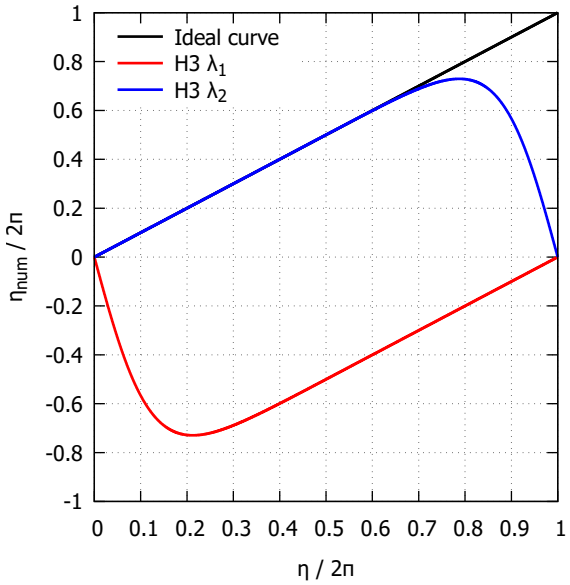
Only the solution with the + is kept because the other solution is negative over the whole range of η.

Finite element H5

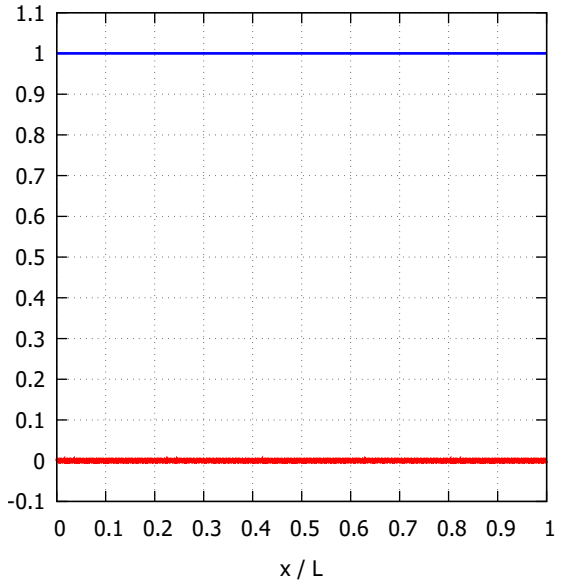
Similarly to what has been done for the Hermite H3 element, one can derive a cubic equation on η_{num} with $\eta \in [0, 3\pi]$.

$$\begin{aligned}\big(144116 + 81159\cos(\eta) + 1524\cos(2\eta) + \cos(3\eta)\big)\eta_{num}^3 \\ + \big(-630279\sin(\eta) - 25020\sin(2\eta) - 27\sin(3\eta)\big)\eta_{num}^2 \\ + \big(-4627800 - 4172130\cos(\eta) - 181080\cos(2\eta) - 270\cos(3\eta)\big)\eta_{num} \\ + 7988310\sin(\eta) + 495000\sin(2\eta) + 990\sin(3\eta) = 0\end{aligned} \tag{1.135}$$

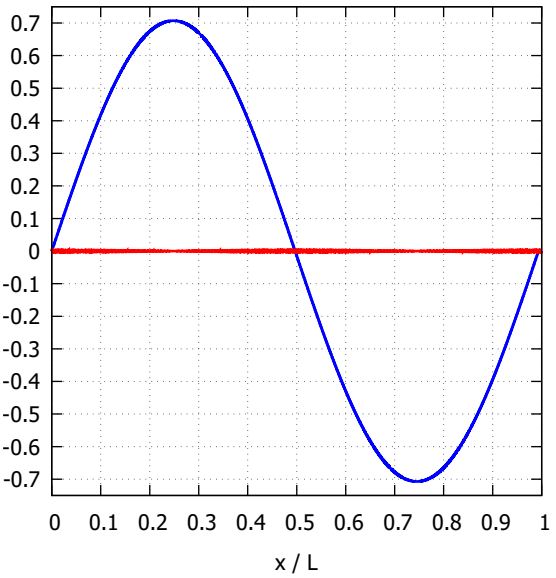
This cubic equation provides three solutions for η_{num} among which only one represents the physics and is thus acceptable (see Fig. 1.7a). To each η corresponds three eigenvectors that can be reconstructed. The three first eigenvectors (for $K = 0$, 1 and 2) are plotted in Fig. 1.7b-1.7d. One sees that one of the three eigenvectors represents the physical behaviour whereas the two other modes behave like a parasite oscillation. These two parasite modes reach their maximum at the locations where the first mode has the steepest variation. As the modified wavenumber K increases, the amplitude of the first eigenvector diminishes down to reaching the amplitude of the two other parasite eigenvectors.



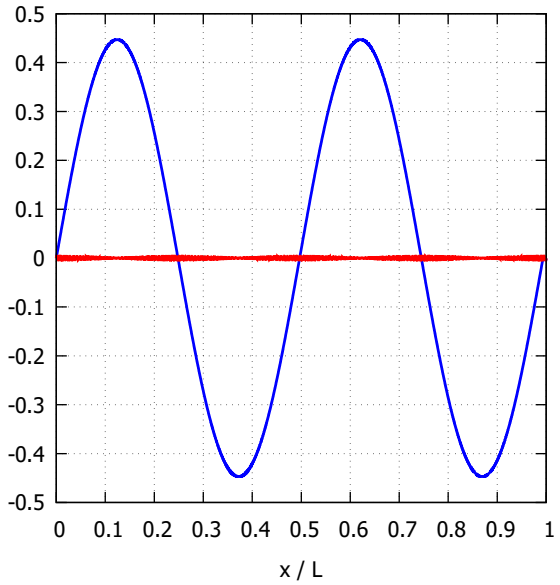
(a) Modified numerical wavenumber



(b) K=0

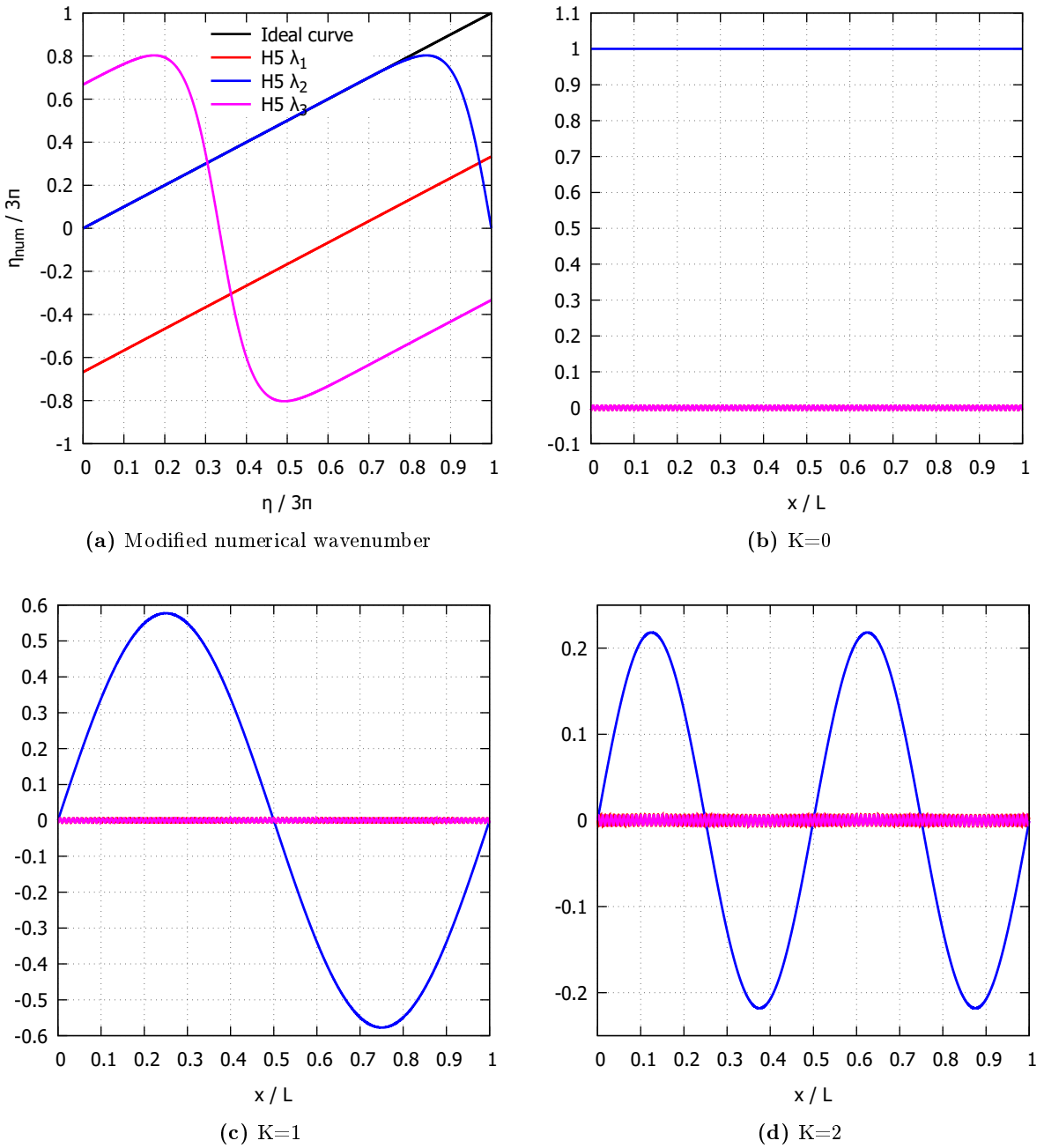


(c) K=1



(d) K=2

Figure 1.6: (a)Modified numerical wavenumber η_{num} as a function of η for the Hermite elements H3. (b)-(d) Eigenvectors of Hermite H3 for $K = 0$, 1 and 2.



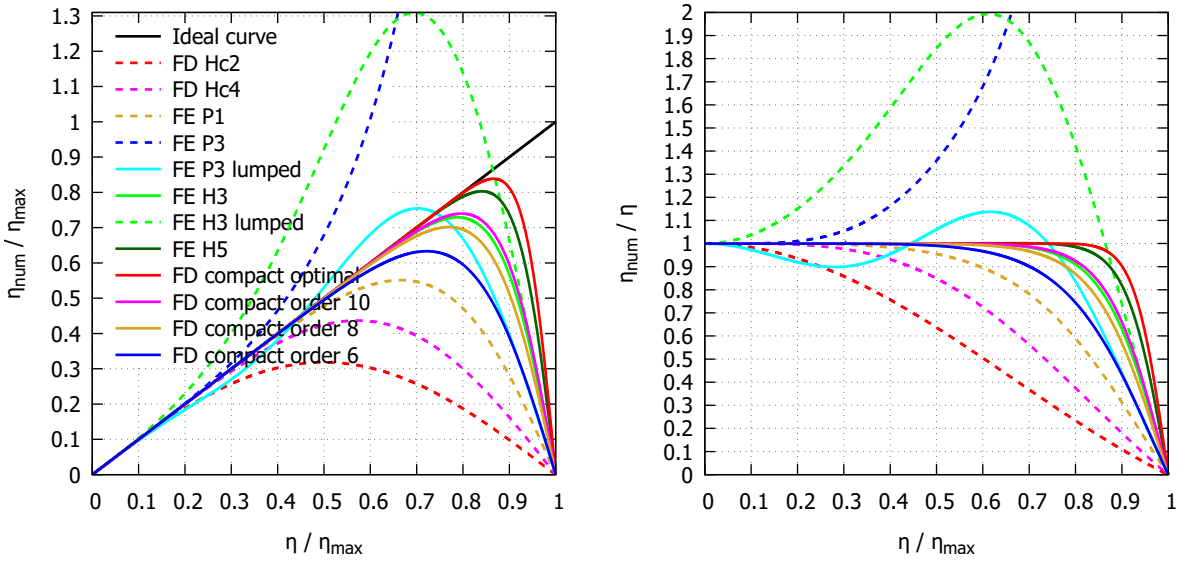
(a) Modified numerical wavenumber

(b) K=0

(c) K=1

(d) K=2

Figure 1.7: (a)Modified numerical wavenumber η_{num} as a function of η for the Hermite elements H5. (b)-(d) Eigenvectors of Hermite H5 for $K = 0$, 1 and 2 (the parasite modes have been magnified by a factor 100 for visibility).



(a) **(b)**

Figure 1.8: (left) Modified numerical wavenumber η_{num} as a function of η and (right) the phase velocity.

1.6.2 Result

The different modified wavenumbers η_{num} are plotted in Fig. 1.8a as a function of the theoretical wavenumber η. The ranges are nondimensionalized by the maximum values η_{max} taken by the spatial discretizations. Ideally the spatial discretization should capture the theoretical wavenumber η. In practice, the discretization introduces dispersive errors. A direct consequence is that the numerical scheme will capture a correct wavenumber until a given η_{div} where the numerical wavenumber diverges from the theoretical one. The ratio of phase velocity η_{num}/η is plotted in Fig. 1.8b. One clearly see that most of the schemes tend to slow down the propagation of waves as the ratio becomes smaller than 1. However the element P3 present an infinite value at $\eta_{div}/\eta_{max} \simeq 0.8838$ which forces the numerical wavenumber η_{num} to break off rapidly from the ideal curve. The waves propagate faster than the physical wave for $\eta_{num}/\eta_{max} < 0.8838$ whereas they propagate in the opposite direction to the physical waves for $\eta_{num}/\eta_{max} > 0.8838$. The lumped version of the element P3 shows a region where the waves propagate more slowly and then a region where they propagate faster than the physical waves. From $\eta_{num}/\eta_{max} \simeq 0.7458$, the behaviour of the numerical wavenumber is again similar to the one of other schemes, showing a slower propagation than the physical wave. The lumped version of the cubic Hermite element presents $\eta_{num} > \eta$ over most of the range of η. The compact finite difference schemes of high order show a very good behaviour in comparison with the high order finite element schemes H3 and H5.

Table 1.3 gives the values of η_{div} for which the relative error between numerical and theoretical wavenumbers is equal to 1%. It is clear that the standard finite differences schemes Hc2 and Hc4 present a lower η_{div} than the finite element schemes. Even the linear element P1 offers a higher η_{div} than the fourth order Hc4 scheme. The best results come from the two Hermite elements H3 and H5 and from the compact finite difference schemes of high order. Lumping the mass matrix of the finite element scheme P3 allows to remove the singularity but leads to an even smaller η_{div} than with the original mass matrix. However, the relative error of the lumped P3 scheme remains within an acceptable range ($\eta_{div}/\eta_{max} < 15\%$) up to a larger wavenumber. A contrario

	η_{max}	$\frac{\eta_{div}}{\eta_{max}}$
FD Hc2	π	0.0781
FD Hc4	π	0.2396
FE P1	π	0.3554
FE P3	3π	0.1968
FE P3 lumped	3π	0.0591
FE H3	2π	0.676
FE H3 lumped	2π	0.0513
FE H5	3π	0.7878
FD compact optimal	π	0.836
FD compact order 10	π	0.6817
FD compact order 8	π	0.6187
FD compact order 6	π	0.502

Table 1.3: Values of η_{div} for which the relative error between numerical and theoretical wavenumbers is equal to 1%.

the lumped cubic Hermite element shows very soon a poor capture of the wavenumber.

Chapter 2

Numerical results

2.1 Shock formation from a sinus wave

2.1.1 Physical behaviour

This section is concerned with the formation of a shock starting from the initial condition $u(x,0) = \sin(2\pi x/L)$. For this purpose, the forcing term is removed for the following results. The formation of the shock from a sine wave is linked with the nonlinear advective term. The maximum and minimum of the sine wave will tend to travel towards each other and join at the middle of the domain, where $u = 0$. Then, the viscosity effect will take over and damp the maxima of the shock. This standard test case is used as a validation for the different discretizations discussed in the previous section.

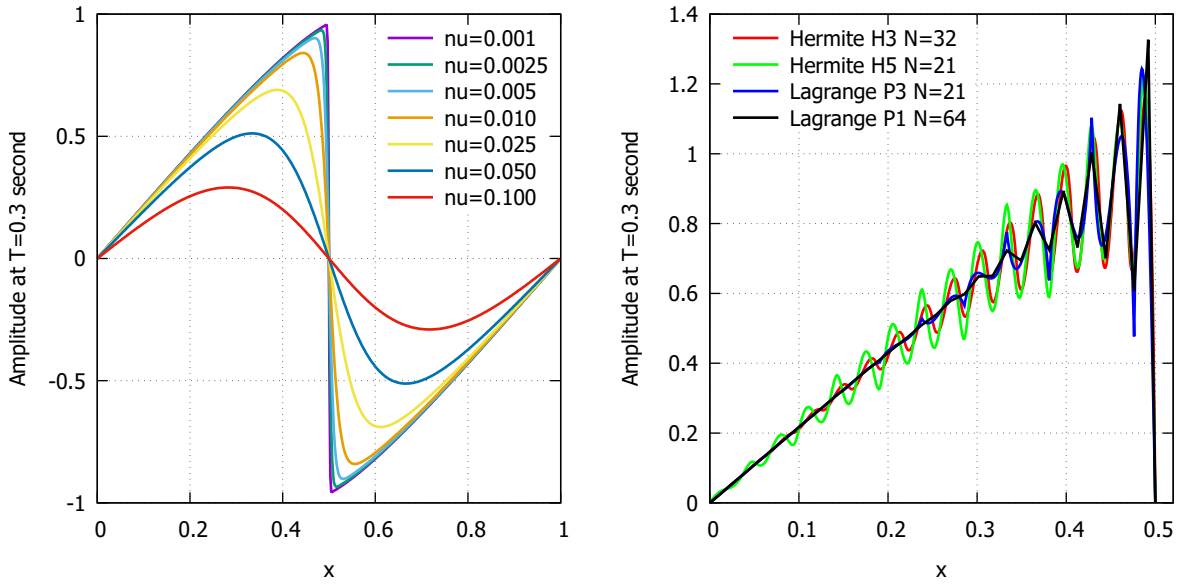
Figure 2.1a shows the amplitude of the solution for several values of the viscosity (from $\nu = 0.001$ to $0.1m^2/s$) at the same time $T = 0.3s$. It is clear that the shock forms only at the lower values of viscosity. For the bigger values of ν, the viscosity takes over the nonlinear effect and rapidly damps the sine wave.

In the case where N is too low, oscillations appear around the shock. This phenomenon is shown on Fig. 2.1b for the finite element Lagrange P1 and P3 and for the Hermite H3 and H5. The computations were performed at an identical problem size until $T = 0.3s$. The solution is re-interpolated inside the elements for the high order polynomials. The linear Lagrange P1 interpolation has the strongest oscillations but they are rapidly damped away from the shock. The same observation is made for the cubic Lagrange element. For this element, we clearly see that the derivatives of the solution are not conserved between the elements. The Hermite H3 and H5 elements ensure respectively the continuity up to the first and second derivatives of the solution across the elements. However, we see that the parasite oscillations propagate farther from the shock than the two other interpolations. For the Hermite H5 element, the damping of these oscillations is very weak so that they tend to propagate farther away of the shock than any other discretization.

Figure 2.1c shows the evolution of the difference between the two maxima as a function of time for different values of viscosity. In addition analytical expressions are also plotted. It is observed that the amplitude of the solution behaves like $\exp(-4t/10\nu)$ before the formation of the shock. After the formation of the shock, the amplitude of the solution behaves like $0.9/(t + 0.1)$. For $\nu = 0.1m^2/s$, the solution is governed by the initial evolution $\exp(-4t/10\nu)$ because no shock is created.

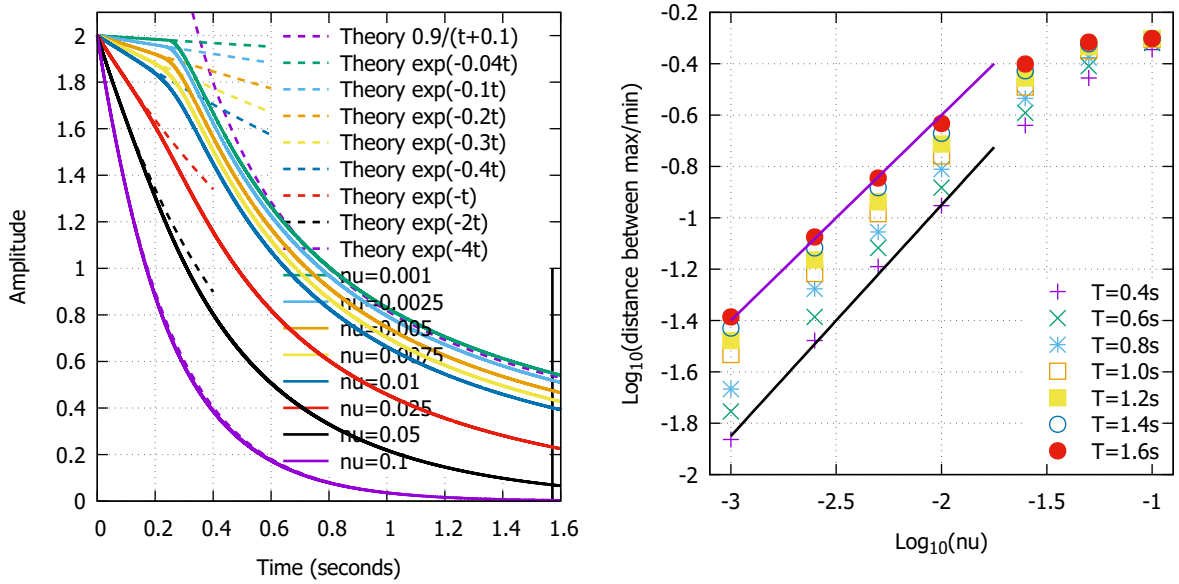
It can be proved that there exists a relation between the viscosity ν and the distance ϵ between the two maxima. This relation takes the form $\log(\epsilon) = \alpha \log(\nu) + \beta$ where α and β are constants. Figure 2.1d shows the distance between the two maxima as a function of the viscosity

at different times in logarithmic axes. We see that the theoretical relation is indeed respected for the lowest values of ν but not for the highest ones.



(a) Sinus wave for Lagrange P1 at T=0.3s and 512 elements

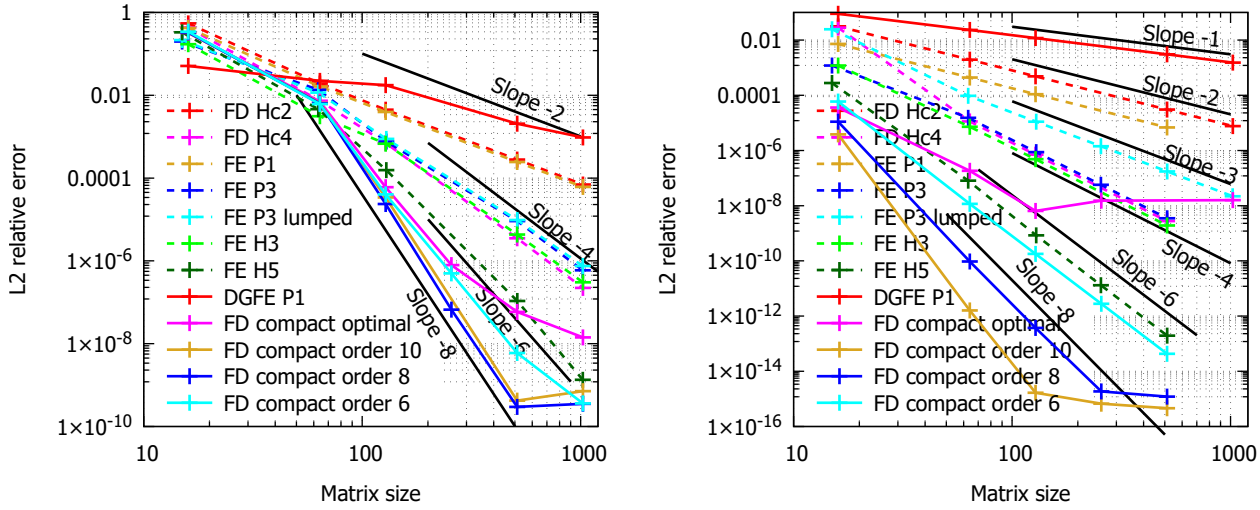
(b) Sinus wave for $\nu = 0.001$ at T=0.3s and reinterpolation for Lagrange P3 and Hermite



(c) Sinus wave, amplitude vs. time for Lagrange P1 and 512 elements

(d) Sinus wave, gap vs. viscosity for Lagrange P1 and 512 elements

Figure 2.1: Formation of a shock from a sinusoidal initial condition



(a) $\nu = 0.01 m^2/s$ **(b)** $\nu = 0.1 m^2/s$

Figure 2.2: Formation of a shock from a sinusoidal initial condition - Convergence curves for three values of the viscosity ($\nu = 0.01$, 0.1 and $1 m^2/s$).

2.1.2 Convergence study

The formation of the shock possesses an analytical solution in the case of an initial solution $u(x,0) = \sin(\pi x)$ where $0 \leq x \leq 1$. This solution is defined by the following expression:

$$u_{\text{th}}(x,t) = \frac{4\pi\nu \sum_{n=1}^{\infty} n I_n\left(\frac{1}{2\pi\nu}\right) \sin(n\pi x) \exp(-n^2\nu\pi^2 t)}{I_0\left(\frac{1}{2\pi\nu}\right) + 2\sum_{n=1}^{\infty} I_n\left(\frac{1}{2\pi\nu}\right) \cos(n\pi x) \exp(-n^2\nu\pi^2 t)} \tag{2.1}$$

where I_n denotes the modified Bessel function of the first kind and of order n. Because of the exponential growth of I_n, it is difficult to evaluate the analytical solution for a viscosity $\nu < 0.01 m^2/s$. Numerical computations were performed with the four finite element discretizations with an initial solution $u(x,0) = \sin(2\pi x)$ and $0 \leq x \leq 2$. The numerical solution is then compared with the analytical one in the domain $0 \leq x \leq 1$. The comparison is based on the L2 norm of the relative error

$$\epsilon_{L2}^{\text{rel}}(t) = \frac{\sqrt{\sum_i \left[u_{\text{num}}(x_i,t) - u_{\text{th}}(x_i,t)\right]^2}}{\sqrt{\sum_i u_{\text{th}}^2(x_i,t)}} \tag{2.2}$$

This relative error is plotted in Fig. 2.2 for 2 values of the viscosity ($\nu = 0.01$ and $0.1 m^2/s$). The computations were all performed at identical matrix sizes until the physical time $T = 1s$ is reached. At the highest matrix sizes, the slopes of the curves are equal to

- -1 for the DGFE Lagrange P1 element
- -2 for the finite element Lagrange P1 and finite difference Hc2 schemes
- -3 for the lumped version of the finite element Lagrange P3 scheme
- -4 for the finite element Lagrange P3 and Hermite H3 and the finite difference Hc4 schemes

	$\Delta t_{\max}$					
Viscosity	ν=0.01m^2/s		ν=0.1m^2/s		ν=1m^2/s	
Matrix size	512	1024	128	512	128	256
Hc2	1.05e-3	2.63e-4	1.67e-3	1.05e-4	1.67e-4	4.17e-5
Hc4	7.94e-4	1.96e-4	1.25e-3	7.94e-5	1.26e-4	3.17e-5
P1	3.57e-4	8.77e-5	5.56e-4	3.45e-5	5.56e-5	1.41e-5
P3	2.22e-4	5.62e-5	3.45e-4	2.22e-5	3.45e-5	8.93e-6
Lumped P3	5.26e-4	1.37e-4	8.33e-4	5.26e-5	8.33e-5	2.17e-5
H3	4e-4	1e-4	6.45e-4	4e-5	6.45e-5	1.61e-5
H5	4e-4	1.02e-4	6.45e-4	4.08e-5	6.45e-5	1.63e-5

Table 2.1: Maximum value for the increment in time for the spatial discretization with an explicit four steps Runge-Kutta time integrator at identical matrix sizes.

- -6 for the finite element Hermite H5 scheme and the compact finite difference scheme of order 6
- -8 for the compact finite difference scheme of order 8
- -10 for the compact finite difference scheme of order 10, at least for $\nu = 0.1m^2/s$
- unknown for the compact optimal finite difference scheme with spectral-like resolution

It is also clear that all the high order schemes lose their advantage at low values of the viscosity and low numbers of elements. This is because of the presence of parasite oscillations around the shock that appears in the middle of the domain. It is also with this combination of low values for the viscosity and numbers of elements that the relative errors are the highest. At low values of viscosity, the high order schemes retrieve their rate of convergence at high number of elements, when the parasite oscillations disappear.

Depending on the value of the viscosity, the finite element P1 and the finite difference scheme Hc2 show nearly identical errors or an order of magnitude different. The same observation is made for the finite elements P3 and H3 and the finite difference scheme Hc4.

2.1.3 A note on the maximum time step

Another interesting property of all the spatial discretization is their different limit on the increment in time which can be used before the explicit time integration scheme becomes unstable. However all the schemes present the same pattern: the limit on the increment in time is

$$\Delta t_{\max} = \text{constant} \frac{\Delta x^2}{\nu} \tag{2.3}$$

The maximum increments in time for each spatial discretization are given in Table 2.1. The results were based on the computations performed in the previous section. It must be underlined that the computations were performed at identical matrix sizes (the H5/P3 schemes need three times less nodes than the P1 scheme to have the same matrix size. The H3 scheme needs two times less nodes than the P1 scheme). The quadratic dependency of $\Delta t_{\max}$ on the mesh size Δx and the inverse dependency on the viscosity ν are clearly visible. The finite difference scheme Hc2 offers the largest increment in time whereas the finite element Lagrange P3 shows the most constraining increment in time. Lumping the mass matrix allows to multiply by a factor 2.4 this limit for the Lagrange P3 element. As a reminder, the lumped version of the P1 scheme is strictly equal to the Hc2 scheme and is thus not used here.

If the computations are performed at identical mesh sizes Δx, then the constant in Eq. (2.3) can be evaluated for the different schemes and are given in Eq. (2.4). It is clear that, at identical mesh sizes Δx, the high order finite element schemes show a very constraining limit on the increment in time in comparison with the P1 element and the finite difference schemes. Lumping the mass matrix helps to increase this limit by multiplying it by a factor 2.4 for the P3 scheme.

$$\begin{aligned}
\Delta t^{\text{Hc2}}_{\max} &\simeq 0.687\frac{\Delta x^2}{\nu} & \Delta t^{\text{Hc4}}_{\max} &\simeq 0.517\frac{\Delta x^2}{\nu} & \Delta t^{\text{P1}}_{\max} &\simeq 0.231\frac{\Delta x^2}{\nu} \\
\Delta t^{\text{P3}}_{\max} &\simeq 0.016\frac{\Delta x^2}{\nu} & \Delta t^{\text{P3 lumped}}_{\max} &\simeq 0.039\frac{\Delta x^2}{\nu} & & \\
\Delta t^{\text{H3}}_{\max} &\simeq 0.065\frac{\Delta x^2}{\nu} & \Delta t^{\text{H5}}_{\max} &\simeq 0.03\frac{\Delta x^2}{\nu} & &
\end{aligned} \tag{2.4}$$

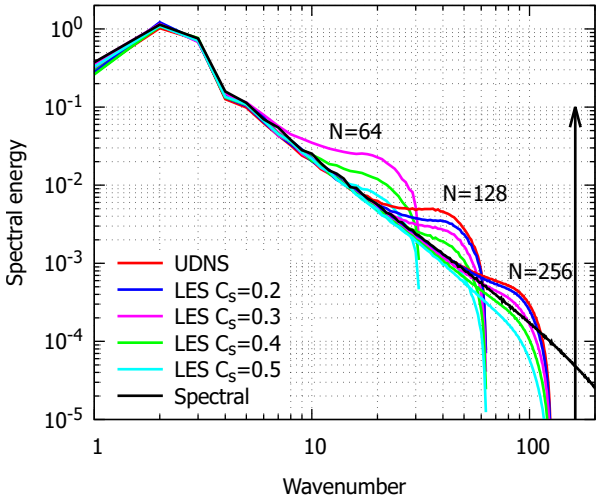
2.2 Turbulent flow

The forced Burgers equation (1.1) is solved through the numerical discretization given in the previous section. A random initial solution evolves towards the unsteady turbulent regime. This turbulent regime is obtained after 10 physical seconds according to the evolution of the kinetic energy. At this moment, the energy spectrum is computed and averaged in time over 200 physical seconds. These energy spectra are then compared with those coming from the spectral method, which is used as a reference. The first section will present the results for computations performed with an insufficient number of nodes to reach the dissipation range. In the second section, resolved DNS are performed. Concerning the cubic Lagrange and Hermite methods, the physical solution is reinterpolated within each element before computing the energy spectra. This ensures that the cubic representation of the solution is taken into account.

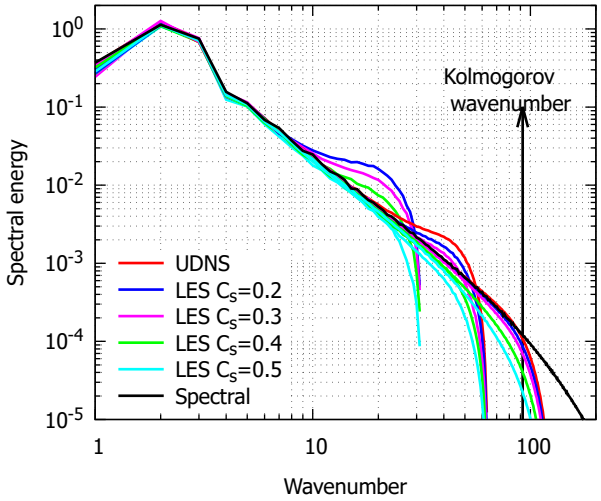
2.2.1 Under-resolved DNS (UDNS)

In the context of LES, a cutoff wavenumber is chosen in the inertial range at which the numerical result is filtered and the higher wavenumber are modelled. In this context, the best numerical scheme is the one which follow the inertial range the longer possible. In the following figures, the energy spectra are plotted as a function of the wavenumber for three values of the number of nodes (N = 64, 128 and 256) and two values for the viscosity ($\nu = 0.0035$ and $0.0075 m^2/s$). The Kolmogorov wavenumber is given as an indication ($k_\nu = 2\pi/L_\nu$ with the Kolmogorov length $L_\nu = LRe^{-3/4}$ and $Re = \text{mean}(u)L/\nu$). Note that the computations are performed at identical matrix size. The Lagrange P3/Hermite H5 and Hermite H3 methods require respectively three and two times less nodes than the linear Lagrange polynomial P1. The discontinuous (DGFE) Lagrange P1 element requires half the number of nodes than the equivalent continuous element. Moreover, for the DGFE P1 and H3, the solution at each node is averaged between the left and right values before the computation of the energy spectrum (this explains why the maximum captured wavenumber is half of the other continuous methods).

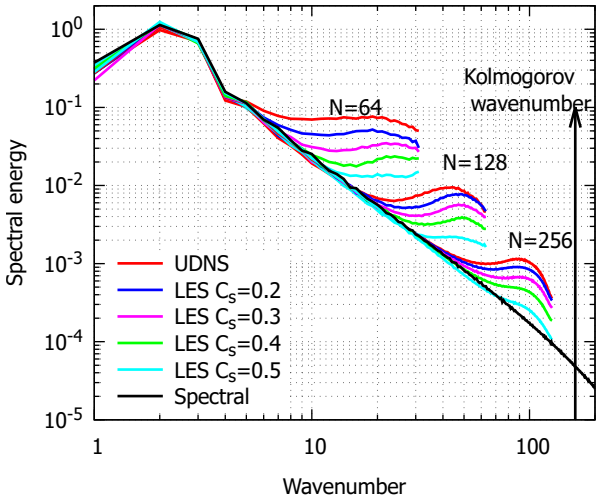
With $N = 64$, the second-order the divergence (Hd2) and energy conservative (Hc2) finite difference schemes and the DFD scheme show poor results in the inertial range. The energy spectra move off the reference curves rapidly to reach a maximum. The compact finite difference schemes also show poor performance at low viscosity ($\nu = 0.0035$). The excitation of the high frequencies is linked with the birth of high-frequency oscillations around the shocks in the physical solution. The fourth order energy conservative scheme Hc4 has comparable results as with the linear Lagrange finite element P1. In overall, the other spatial discretizations of higher order show better results at $N = 64$. However the high order Hermite polynomials show a consequent high-frequency excitation, for both the standard and lumped mass matrix formulations. This is



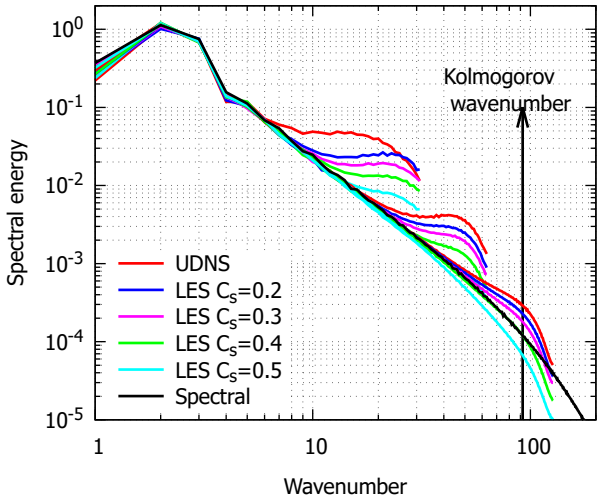
(a) Energy dissipative Hd2



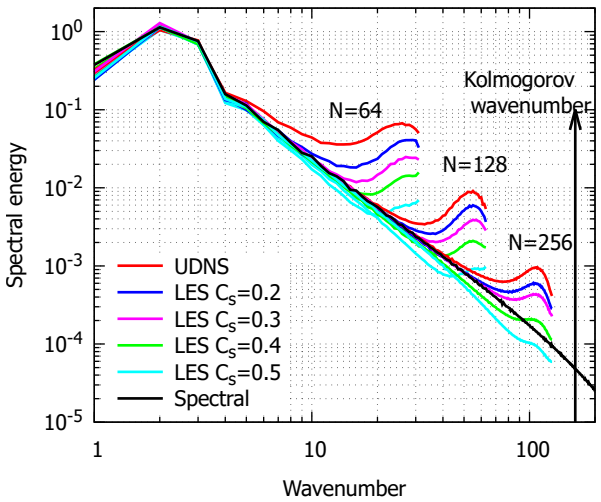
(b) Energy dissipative Hd2



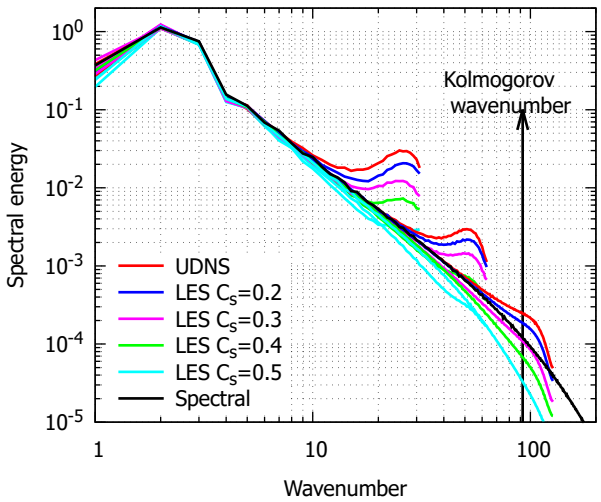
(c) Energy conservative Hc2



(d) Energy conservative Hc2

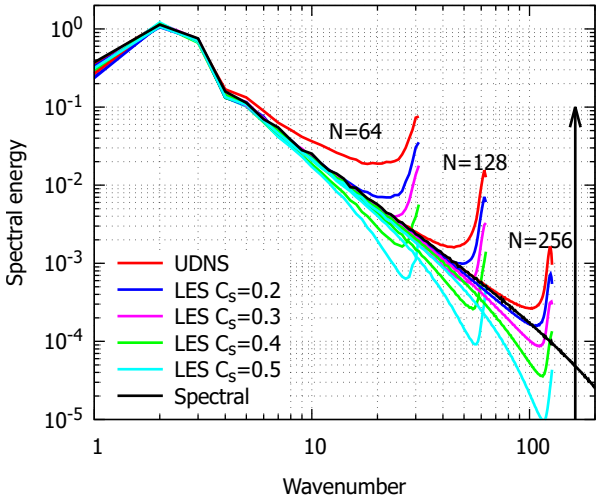


(e) Energy conservative Hc4

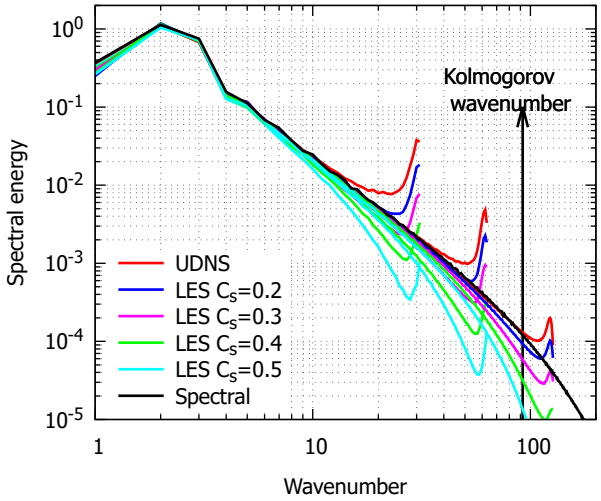


(f) Energy conservative Hc4

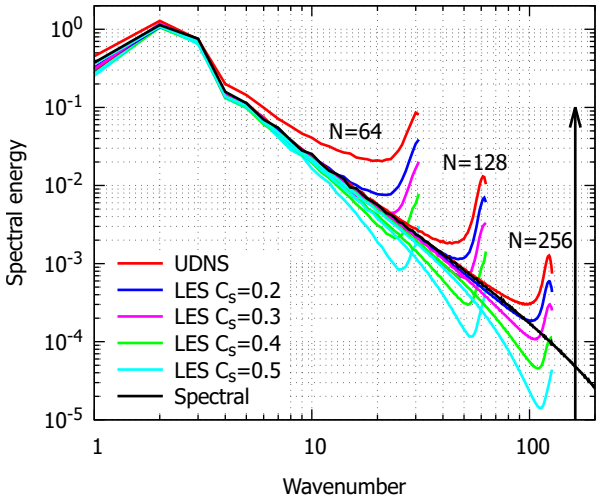
Figure 2.3: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



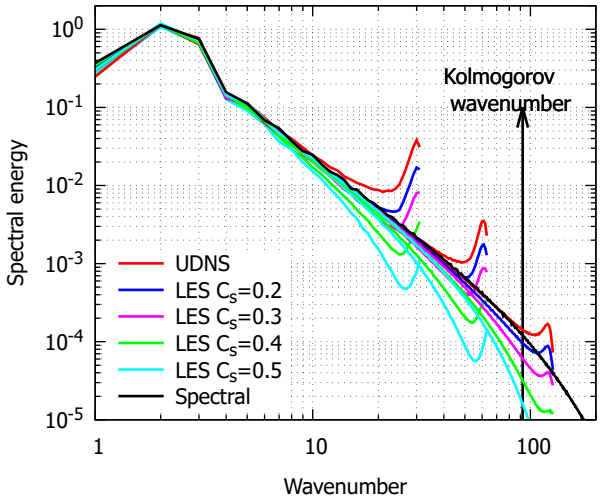
(a) Compact optimal



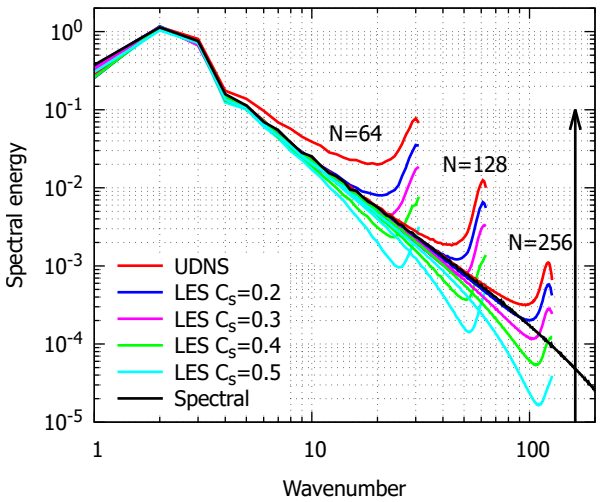
(b) Compact optimal



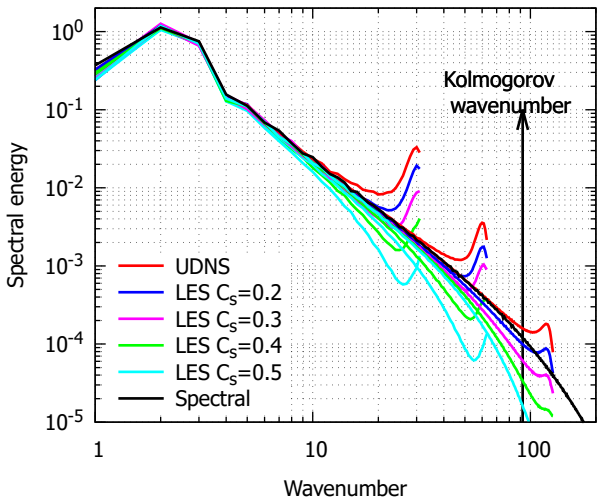
(c) Compact 10th order



(d) Compact 10th order

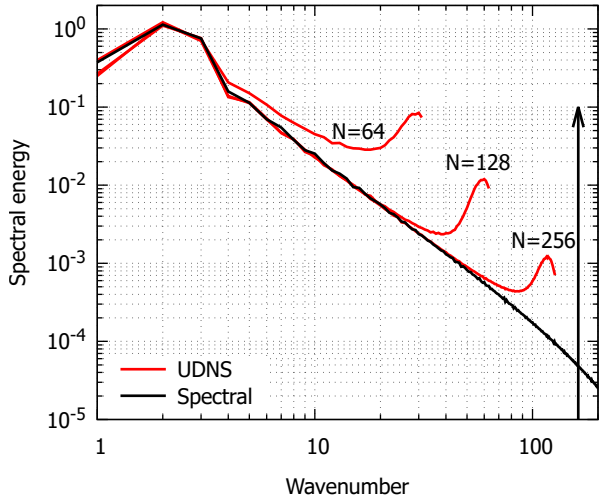


(e) Compact 8th order

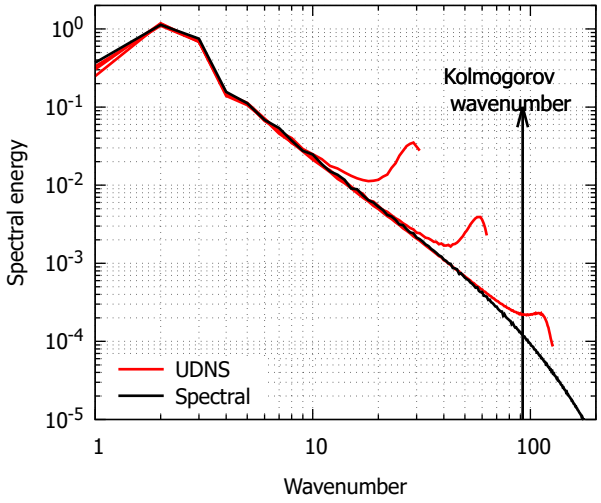


(f) Compact 8th order

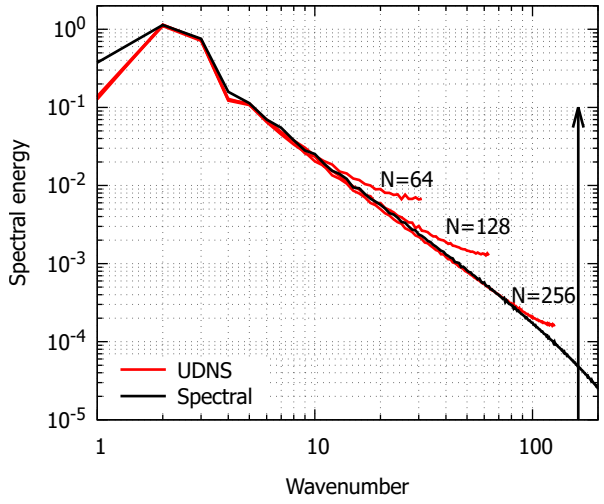
Figure 2.4: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



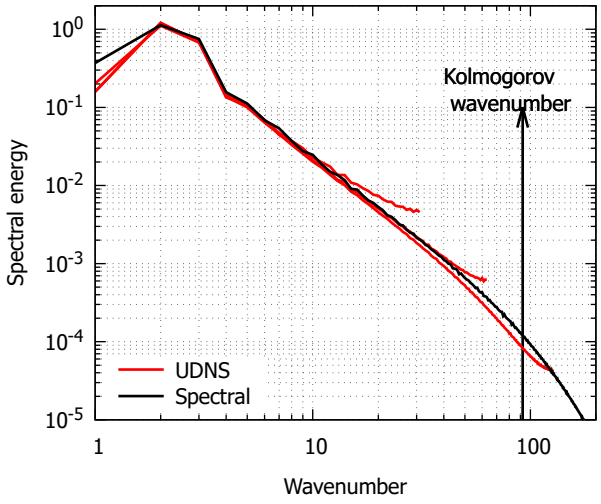
(a) Nonlinear DFD



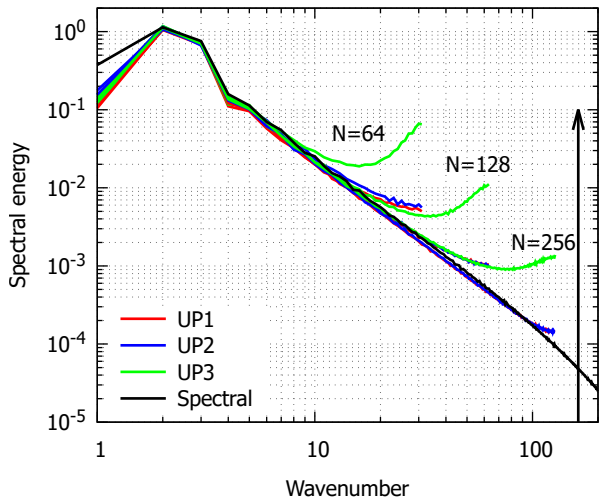
(b) Nonlinear DFD



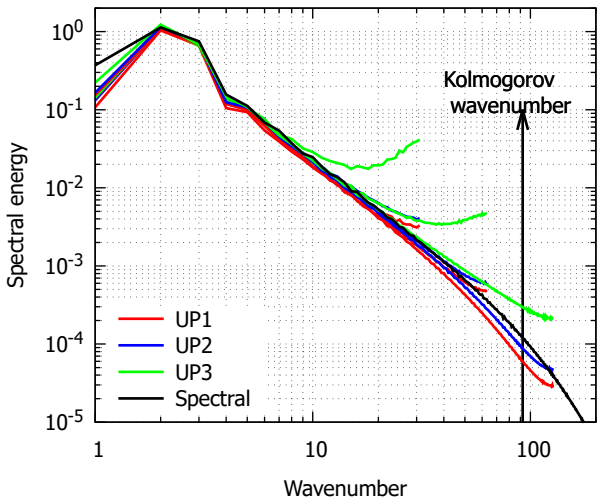
(c) Nonlinear TVD



(d) Nonlinear TVD

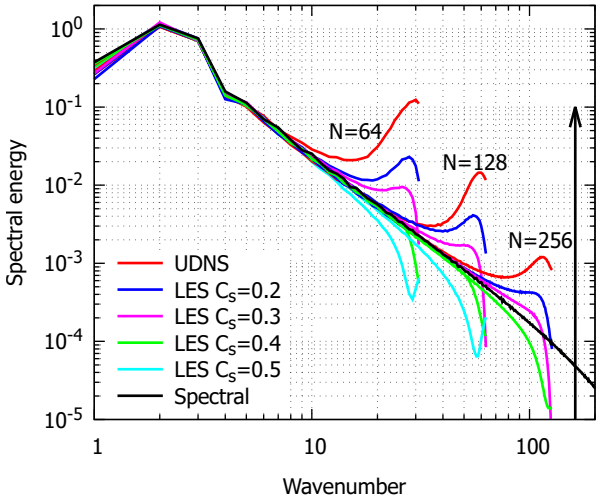


(e) Nonlinear Upwind

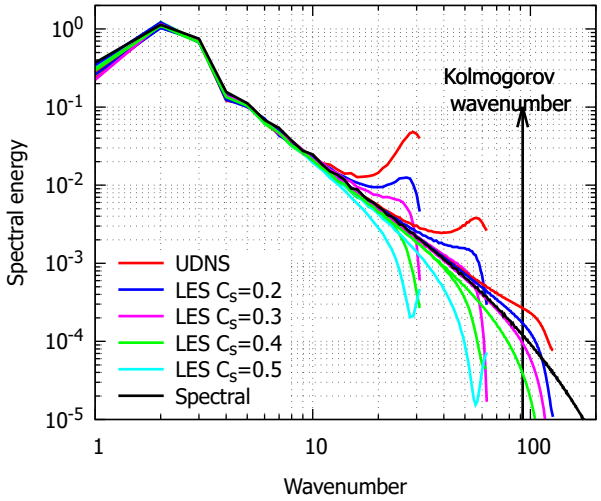


(f) Nonlinear Upwind

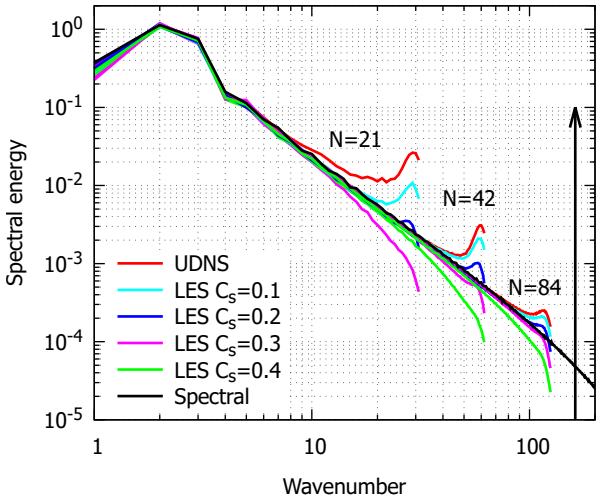
Figure 2.5: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



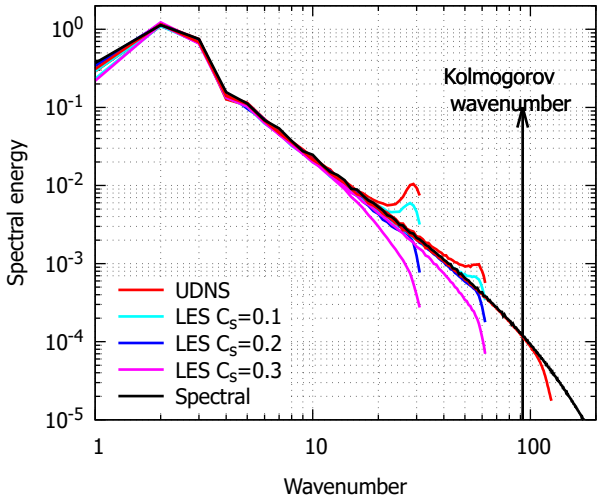
(a) Continuous Lagrange P1



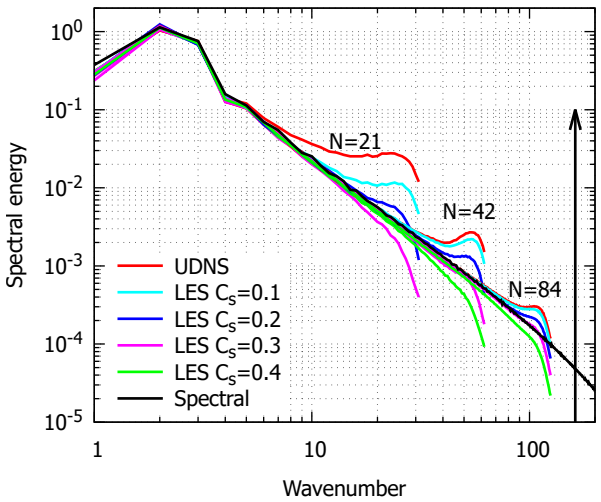
(b) Continuous Lagrange P1



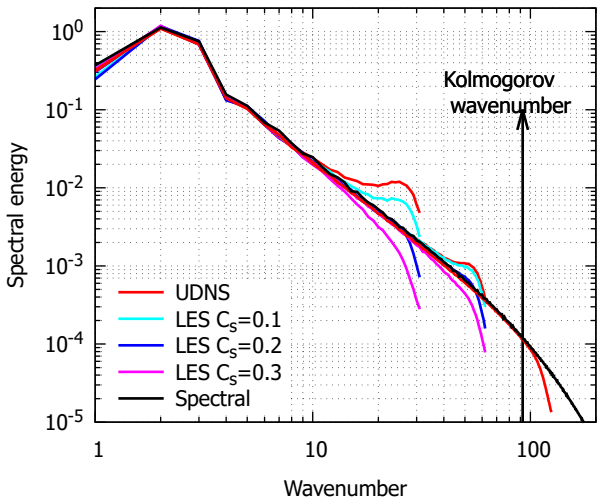
(c) Continuous Lagrange P3



(d) Continuous Lagrange P3

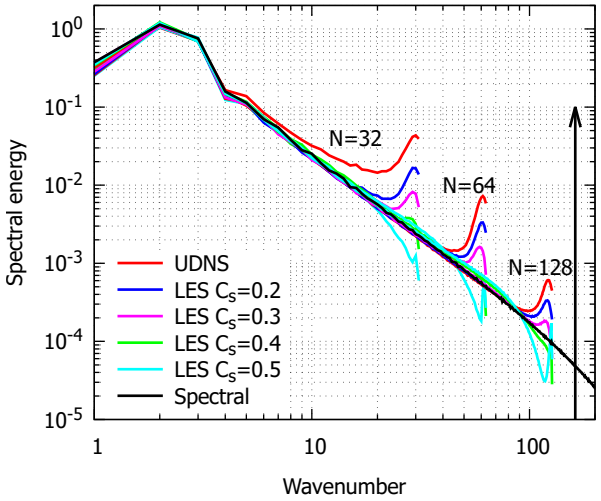


(e) Continuous Lagrange P3 lumped matrix

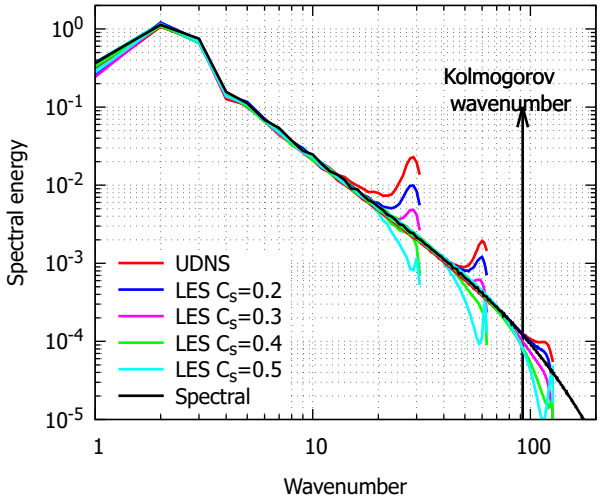


(f) Continuous Lagrange P3 lumped matrix

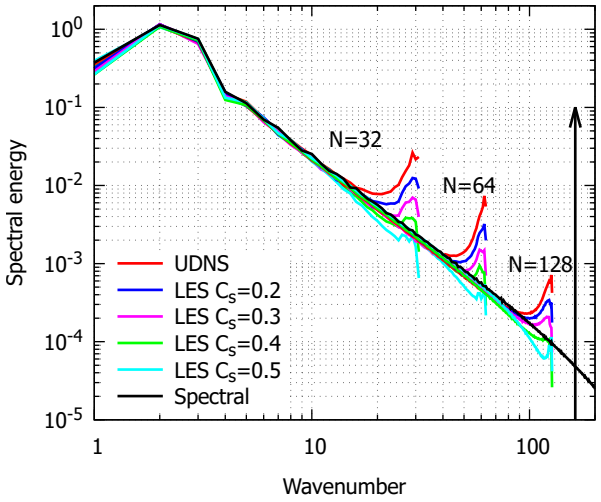
Figure 2.6: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



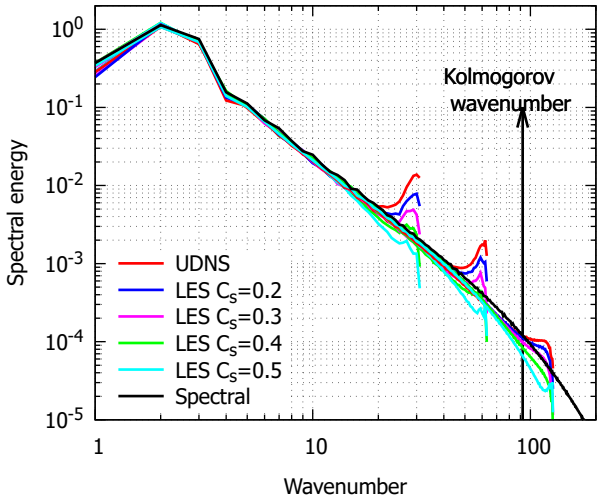
(a) Continuous Hermite H3



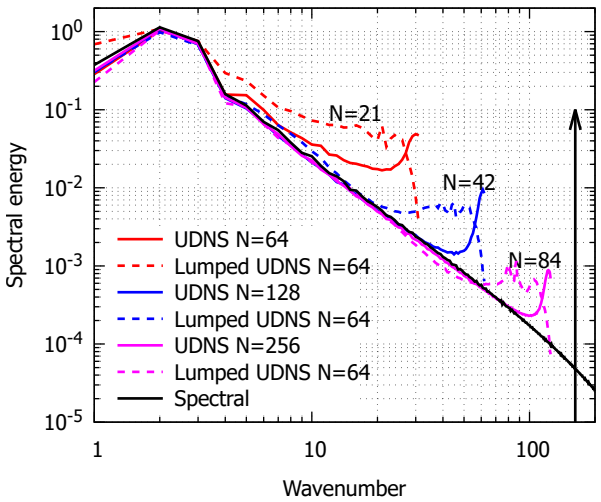
(b) Continuous Hermite H3



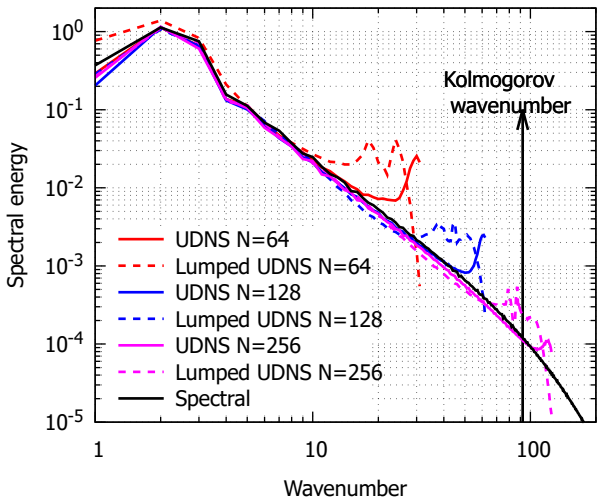
(c) Continuous Hermite H3 lumped matrix



(d) Continuous Hermite H3 lumped matrix

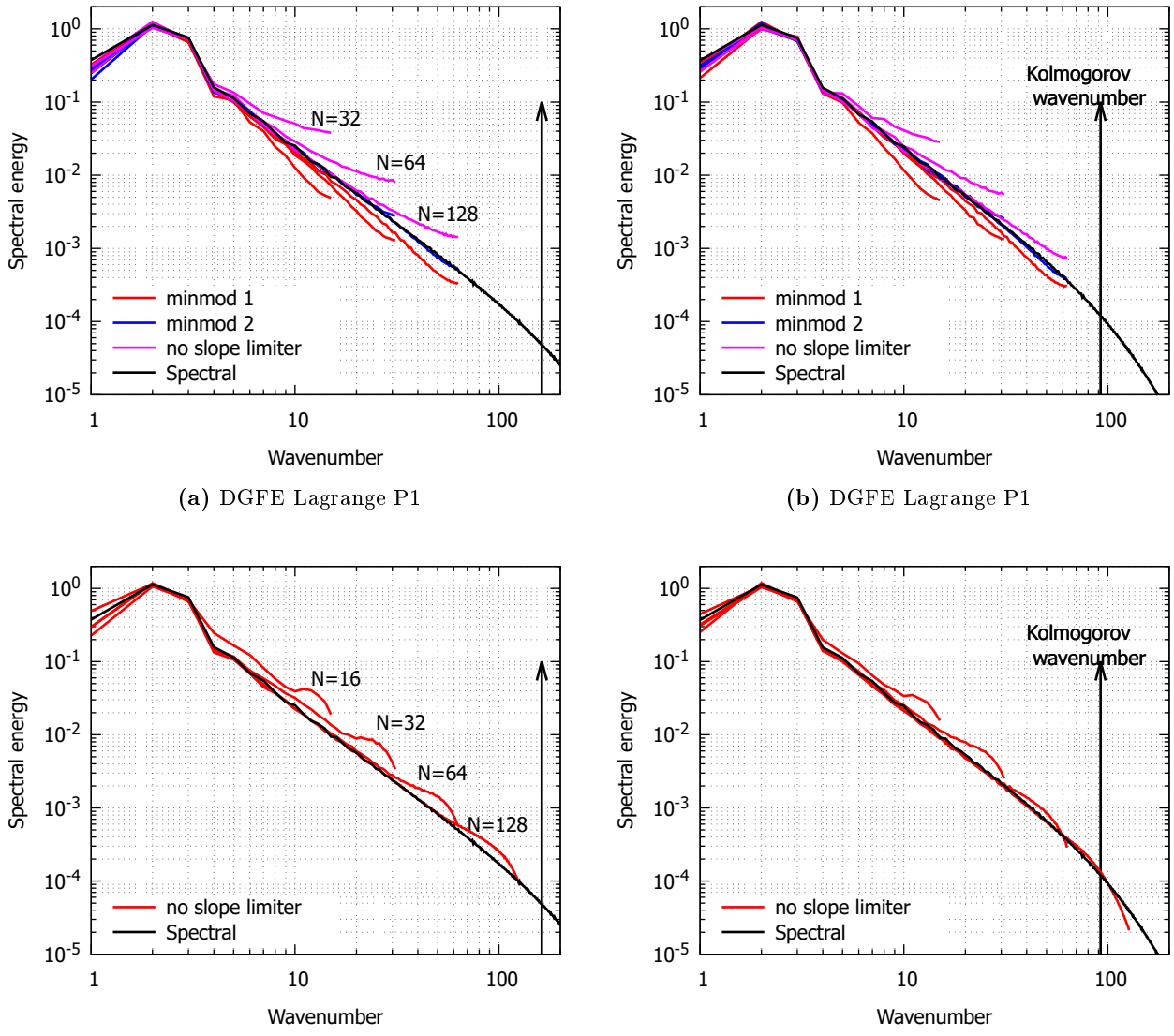


(e) Continuous Hermite H5



(f) Continuous Hermite H5

Figure 2.7: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



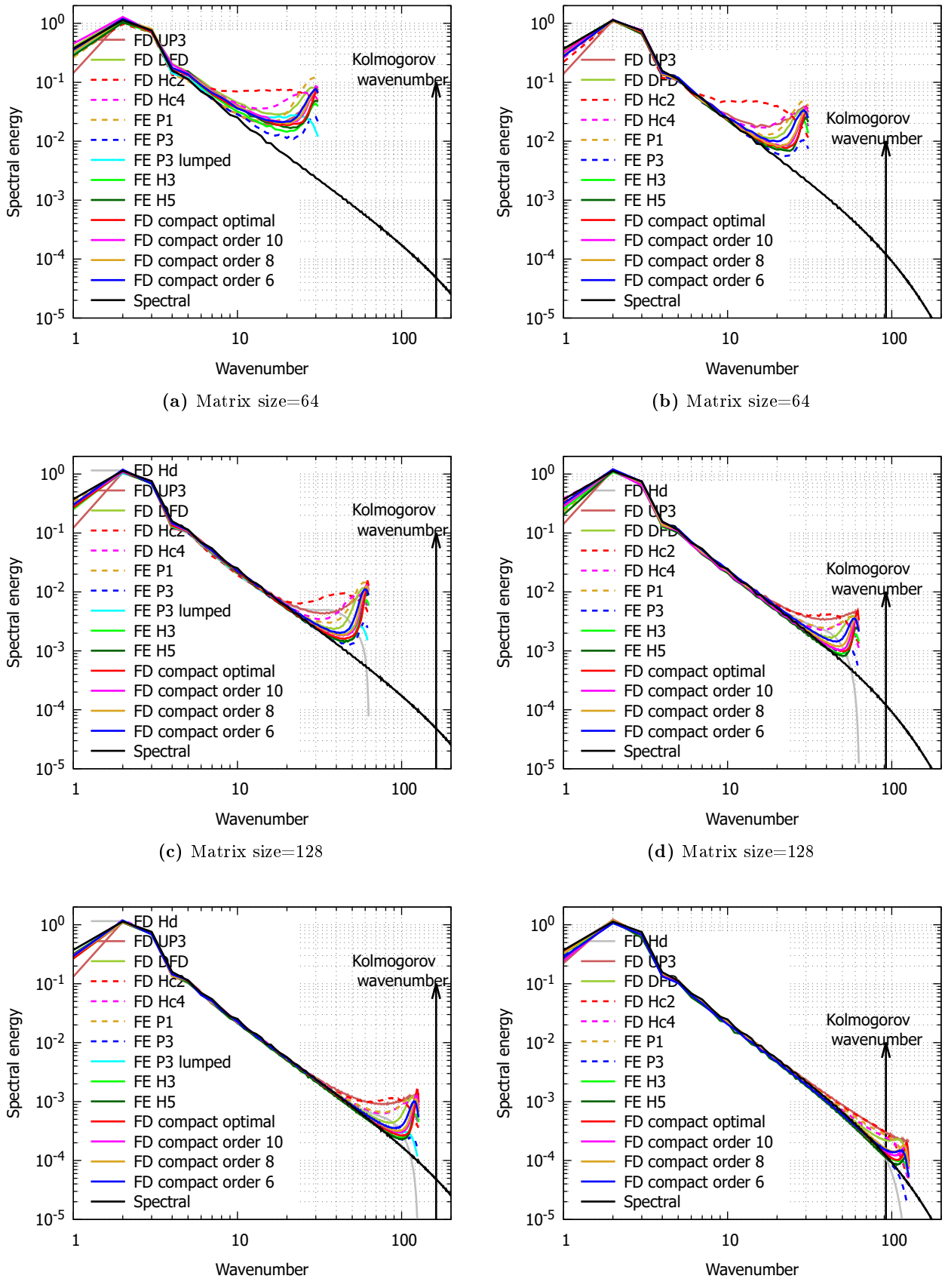
(a) DGFE Lagrange P1

(b) DGFE Lagrange P1

(c) DGFE Hermite H3

(d) DGFE Hermite H3

Figure 2.8: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$



(a) Matrix size=64

(b) Matrix size=64

(c) Matrix size=128

(d) Matrix size=128

(e) Matrix size=256

(f) Matrix size=256

Figure 2.9: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$

linked with their weak level of damping at these frequencies. Concerning the DGFE Lagrange P1 element, the importance of the choice of the slope limiter is highlighted in Fig. 2.8. Cancelling the slope limiter introduces numerical oscillations at high frequencies. The minmod 1 limiter displays a too high dissipation as it also acts on smooth regions. The minmod 2 limiter displays a nearly perfect matching with the reference curves. The main disadvantage of minmod 2 is the manual selection of the parameter M (see Section 1.4.5, $M = 10$ for $N = 32$, $M = 40$ for $N = 64$, $M = 100$ for $N = 128$). In overall the discontinuous methods without slope limiters display a better behaviour than their continuous counterpart without any subgrid term.

For $N = 128$ and 256, the inertial range is better captured by the different schemes. As expected the upwind UP3 and the TVD schemes show too much dissipation whereas the UP1 and DFD schemes show too few dissipation. Note the progressive damping of the maxima at high wavenumbers due to the disappearance of the oscillations around the shocks. This maximum even disappear for the energy dissipative scheme Hd2 with $\nu = 0.0075m^2/s$ and $N = 256$. The cubic Lagrange P3 and Hermite H3 and fifth order Hermite H5 finite elements offer the best performances with the break-off from the inertial range located at higher wavenumbers. The fifth order Hermite H5 element shows stronger peaks at high wavenumbers than the cubic Hermite H3 element. This is due to the weaker damping of the parasite oscillations for the H5 element, as seen in section 2.1. The cubic Lagrange P3 element fits the inertial range up to the same wavenumbers as the Hermite elements but shows weaker excitations of the high wavenumbers than the Hermite polynomials. In the configuration of subfigure 2.9f, the peak disappears and is replaced by an over-damping in comparison with the spectral method. The lumped version of the Hermite elements show a poor resolution of the inertial range, even at $N = 256$. The explanation lies in its over-prediction of the wave-number η observed in Fig. 1.8. The DGFE P1 element with the minmod 2 slope limiter displays the best matching with the reference curves.

Large-eddy computations (LES) were also performed with the constant eddy viscosity formulation. Results for different values of C_s are shown hereunder for some schemes. In overall the pile-up of energy close to grid cut-off is still present for the finite different methods. Obviously the dissipative scheme Hd2 succeeds in reducing this pile-up but the optimal value for C_s strongly depends on the grid size and on the viscosity. The best LES results come from the energy conservative schemes Hc2 and Hc4 and from the linear finite element method Lagrange P1.

The Smagorinsky constant has been dynamically calculated by the method explained in Section 1.1.2 and the results are given in Tables 2.2 and 2.3 for the two considered viscosities. The Smagorinsky constants have been averaged in time and their relative standard deviations, i.e. the ratio between the standard deviation and the average in time, are also given in percentage. The missing values for the energy dissipative scheme Hd2 are due to the presence of high peaks in the solution, which leads to the divergence of the computation. In overall one sees that an increase in the stencil of the filter decreases the constant C_s. Filters with small stencils are more dissipative than filters with big stencils. The optimal compact finite difference scheme shows the lowest values of C_s for small stencils but this comes at the cost of high energy peaks at high frequencies (see Figure 2.4). The linear Lagrange finite element method P1 shows relatively small relative standard deviations. As seen on Figure 2.6, this method succeeds in removing most of the energy peaks at high frequencies. The energy-conservative finite difference schemes Hc2 and Hc4 show values of C_s similar to the linear finite element method P1 but their relative standard deviation is much more variable. Moreover these schemes do not succeed in removing the energy peaks at high frequencies as shown on Figure 2.3.

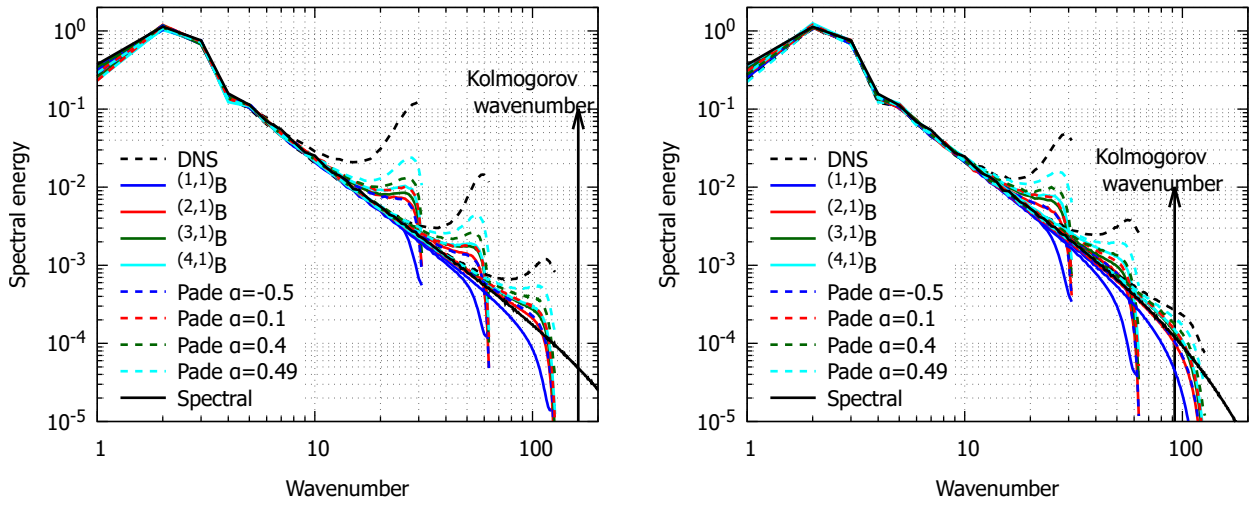
Figure 2.10 shows the influence of the type of low-pass filters on the dynamic Smagorinsky models (see Section 1.1.3). The ${}^{(1,1)}B$ binomial filter clearly over-damps the energy spectrum in the dissipative range. As stated in Section 1.1.3, the Padé filter with $\alpha = -0.5$ is identical

Discretization	Filters	$N = 64$	$N = 128$	$N = 256$
FD Hd2	${}^{(1,1)}B$	$0.3899 \pm 18.38\%$	$0.4060 \pm 8.44\%$	$0.3940 \pm 5.81\%$
	${}^{(2,1)}B$	$0.2424 \pm 33.12\%$	$0.2860 \pm 14.78\%$	$0.2722 \pm 10.56\%$
	${}^{(3,1)}B$	div	$0.2384 \pm 19.43\%$	$0.2142 \pm 15.13\%$
	${}^{(4,1)}B$	div	$0.2113 \pm 23.52\%$	$0.2007 \pm 16.79\%$
FD Hc2	${}^{(1,1)}B$	$0.4371 \pm 20.93\%$	$0.4477 \pm 11.78\%$	$0.4265 \pm 8.17\%$
	${}^{(2,1)}B$	$0.3176 \pm 33.97\%$	$0.3365 \pm 20.80\%$	$0.3167 \pm 15.03\%$
	${}^{(3,1)}B$	$0.2791 \pm 40.94\%$	$0.2889 \pm 27.76\%$	$0.2717 \pm 19.31\%$
	${}^{(4,1)}B$	$0.2351 \pm 46.92\%$	$0.2600 \pm 33.73\%$	$0.2461 \pm 22.94\%$
FD Hc4	${}^{(1,1)}B$	$0.4043 \pm 13.74\%$	$0.4041 \pm 7.66\%$	$0.3858 \pm 5.75\%$
	${}^{(2,1)}B$	$0.3049 \pm 20.41\%$	$0.3074 \pm 13.31\%$	$0.2840 \pm 10.54\%$
	${}^{(3,1)}B$	$0.2663 \pm 26.11\%$	$0.2701 \pm 16.16\%$	$0.2478 \pm 13.24\%$
	${}^{(4,1)}B$	$0.2428 \pm 29.83\%$	$0.2450 \pm 19.51\%$	$0.2242 \pm 15.65\%$
FD Compact	${}^{(1,1)}B$	$0.3799 \pm 8.36\%$	$0.3805 \pm 4.67\%$	$0.3681 \pm 3.99\%$
spectral	${}^{(2,1)}B$	$0.2877 \pm 9.60\%$	$0.2809 \pm 7.76\%$	$0.2651 \pm 6.92\%$
optimal	${}^{(3,1)}B$	$0.2521 \pm 11.09\%$	$0.2470 \pm 7.19\%$	$0.2302 \pm 7.75\%$
	${}^{(4,1)}B$	$0.2326 \pm 10.98\%$	$0.2274 \pm 7.76\%$	$0.2111 \pm 8.03\%$
FE P1	${}^{(1,1)}B$	$0.4102 \pm 9.02\%$	$0.4144 \pm 4.35\%$	$0.4015 \pm 3.75\%$
	${}^{(2,1)}B$	$0.3225 \pm 9.63\%$	$0.3184 \pm 6.81\%$	$0.2997 \pm 6.46\%$
	${}^{(3,1)}B$	$0.2929 \pm 10.10\%$	$0.2842 \pm 8.48\%$	$0.2653 \pm 8.38\%$
	${}^{(4,1)}B$	$0.2731 \pm 12.88\%$	$0.2670 \pm 8.90\%$	$0.2464 \pm 9.72\%$

Table 2.2: $\nu = 0.0035m^2/s$. Averaged values of the Smagorinsky constant C_s calculated by the dynamic method and its relative standard deviation in percent.

Discretization	Filters	$N = 64$	$N = 128$	$N = 256$
FD Hd2	${}^{(1,1)}B$	$0.3898 \pm 15.00\%$	$0.3898 \pm 7.46\%$	$0.3676 \pm 5.99\%$
	${}^{(2,1)}B$	$0.2739 \pm 20.78\%$	$0.2679 \pm 13.21\%$	$0.2347 \pm 12.94\%$
	${}^{(3,1)}B$	$0.2166 \pm 30.22\%$	$0.2203 \pm 17.64\%$	$0.2268 \pm 15.28\%$
	${}^{(4,1)}B$	$0.1977 \pm 30.76\%$	$0.1938 \pm 20.59\%$	$0.1597 \pm 21.71\%$
FD Hc2	${}^{(1,1)}B$	$0.4290 \pm 18.07\%$	$0.4204 \pm 9.98\%$	$0.3883 \pm 7.20\%$
	${}^{(2,1)}B$	$0.3207 \pm 28.27\%$	$0.3074 \pm 17.12\%$	$0.2660 \pm 15.17\%$
	${}^{(3,1)}B$	$0.2779 \pm 33.73\%$	$0.2679 \pm 21.61\%$	$0.2201 \pm 20.17\%$
	${}^{(4,1)}B$	$0.2528 \pm 37.49\%$	$0.2392 \pm 25.67\%$	$0.1959 \pm 24.51\%$
FD Hc4	${}^{(1,1)}B$	$0.3928 \pm 11.96\%$	$0.3798 \pm 7.39\%$	$0.3550 \pm 5.85\%$
	${}^{(2,1)}B$	$0.2992 \pm 17.22\%$	$0.2802 \pm 12.09\%$	$0.2405 \pm 12.71\%$
	${}^{(3,1)}B$	$0.2585 \pm 22.03\%$	$0.2411 \pm 14.75\%$	$0.2006 \pm 16.21\%$
FD Compact	${}^{(1,1)}B$	$0.3708 \pm 8.40\%$	$0.3649 \pm 4.95\%$	$0.3458 \pm 4.34\%$
spectral	${}^{(2,1)}B$	$0.2768 \pm 10.20\%$	$0.2608 \pm 8.47\%$	$0.2313 \pm 10.37\%$
optimal	${}^{(3,1)}B$	$0.2415 \pm 11.37\%$	$0.2265 \pm 9.00\%$	$0.1931 \pm 13.26\%$
	${}^{(4,1)}B$	$0.2242 \pm 11.35\%$	$0.2070 \pm 9.78\%$	$0.1745 \pm 12.14\%$
FE P1	${}^{(1,1)}B$	$0.4028 \pm 8.78\%$	$0.3978 \pm 8.47\%$	$0.3738 \pm 5.00\%$
	${}^{(2,1)}B$	$0.3103 \pm 11.07\%$	$0.2945 \pm 9.03\%$	$0.2556 \pm 12.28\%$
	${}^{(3,1)}B$	$0.2810 \pm 10.66\%$	$0.2614 \pm 10.03\%$	$0.2160 \pm 16.31\%$
	${}^{(4,1)}B$	$0.2611 \pm 12.67\%$	$0.2391 \pm 12.09\%$	$0.1912 \pm 19.32\%$

Table 2.3: $\nu = 0.0075m^2/s$. Averaged values of the Smagorinsky constant C_s calculated by the dynamic method and its relative standard deviation in percent.



(a) $\nu = 0.0035m^2/s$ **(b)** $\nu = 0.0075m^2/s$

Figure 2.10: Under-resolved DNS - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$. Influence of the type of low-pass filters on the dynamic Smagorinsky models. Lagrange P1 finite elements.

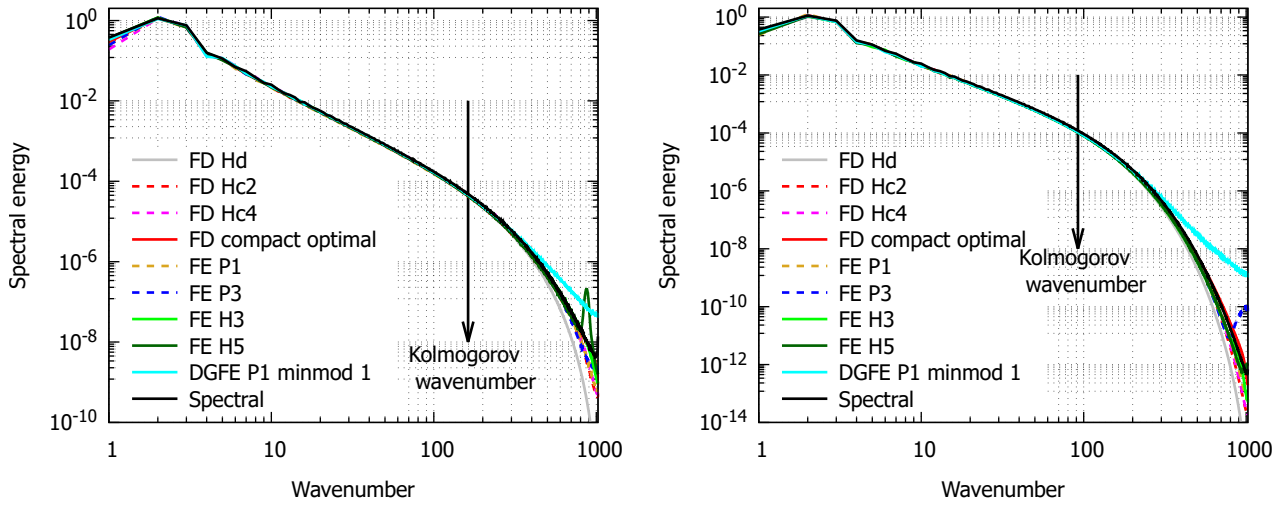
to ${}^{(2,1)}B$. The Padé filter with $\alpha = 0.1$ shows an energy spectrum located between ${}^{(3,1)}B$ and ${}^{(4,1)}B$. The Padé filters with high values of α show a high energy pile-up at high frequencies. The best compromise between over-damping and energy pile-up is reached by ${}^{(2,1)}B$ and $\alpha = -0.5$. The Fourier cut-off filter is located at the wavenumber $k = N/4$. The transfer functions plotted on Fig. 1.1 explain why the filters ${}^{(2,1)}B$ and $\alpha = -0.5$ are best suited to attenuate the energy pile-up in the range $N/4 \leq k \leq N/2$. Indeed their cut-off wavenumbers are located before the energy pile-up. The binomial filters with larger stencils and the Padé filters with positive α are not suited because their cut-off wavenumbers are located inside the energy pile-up.

2.2.2 Resolved DNS

Figure 2.11 shows the results of the different numerical schemes for resolved DNS computations. The problem size is equal to 2048. The relative error is computed with the spectral method used as the reference. The inertial range is very well computed by all the schemes. The main differences appear at higher wavenumbers than the Kolmogorov wavenumber:

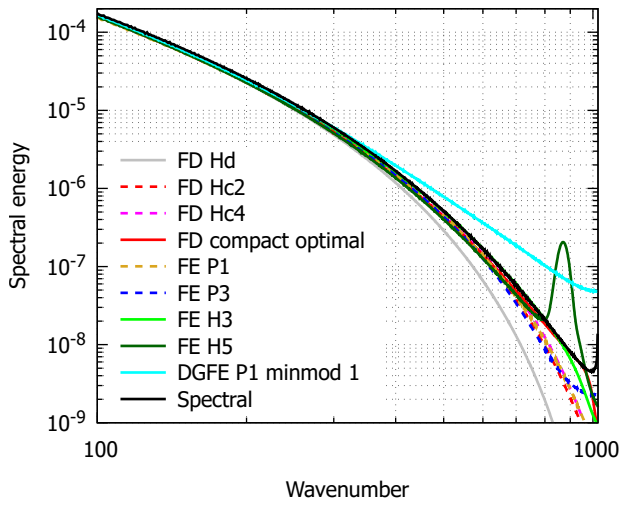
- The second order energy dissipative finite difference scheme Hd2 has the worst behaviour at large wavenumbers: the diffusive range is not satisfactorily computed with a too high damping.
- The second and fourth order energy conservative scheme Hc2/Hc4, the linear Lagrange finite element P1 and the cubic Lagrange finite element P3 present similar results.
- The compact optimal FD scheme offers the best results for both values of the viscosity.
- The Hermite H5 element shows a parasite excitation at very high wavenumbers.
- The Hc4, P3, H3 and H5 still have an parasite excitation of the highest wavenumbers, mainly for $\nu = 0.0075m^2/s$.

- The discontinuous Lagrange P1 element displays a parasite excitation at high wavenumbers even with the minmod 1 slope limiter that has an aggressive damping characteristics.

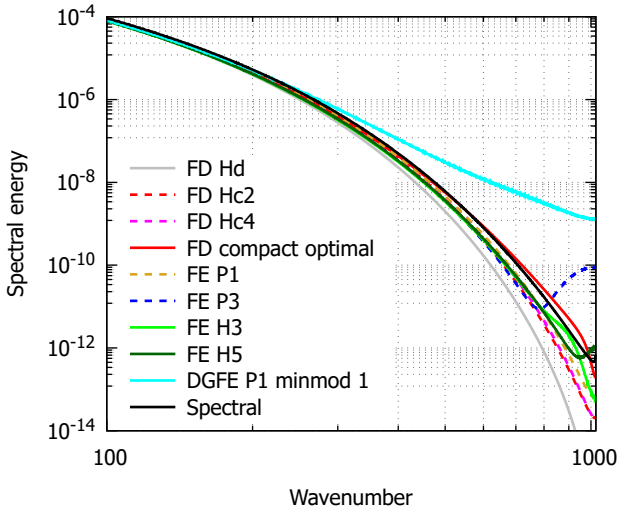


(a) Energy spectra

(b) Energy spectra



(c) Enlargement at high wavenumbers



(d) Enlargement at high wavenumbers

Figure 2.11: Resolved DNS - Matrix size=2048 - Energy spectra for (left) $\nu = 0.0035m^2/s$ and (right) $\nu = 0.0075m^2/s$. (top) Full spectrum and (bottom) enlargement at high wavenumbers.

Bibliography

[1] Romit Maulik and Omer San. Evaluation of explicit and implicit LES closures for Burgers turbulence. *Journal of Computational and Applied Mathematics*, 327:12–40, 2018.

[2] Sanjiva K. Lele. Compact finite difference schemes with spectral-like resolution. *Journal of Computational Physics*, 103:16–42, 1992.

[3] D. Fauconnier and E. Dick. On the spectral and conservation properties of nonlinear discretization operators. *Journal of Computational Physics*, 230:4488–4518, 2011.

[4] Rong Wang and Raymond J. Spiteri. Linear instability of the fifth-order WENO method. *SIAM Journal on Numerical Analysis*, 45:1871–1901, 2007.

[5] Foluso Ladeinde, Xiaodan Cai, Miguel R. Visbal, and Datta V. Gaitonde. Turbulence spectra characteristics of high order schemes for direct and large eddy simulation. *Applied Numerical Mathematics*, 36:447–474, 2001.

[6] Chi-Wang Shu. Different formulations for the discontinuous Galerkin method for the viscous term. 2000.

[7] Carlos Erik Baumann and J. Tinsley Oden. A discontinuous hp finite element method for convection-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 175:311–341, 1999.

[8] Bernardo Cockburn and Chi-Wang Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II: General framework. *Mathematics of Computation*, 52(186):411–435, 1989.

[9] Jan S. Hesthaven and Tim Warburton. *Nodal discontinuous Galerkin methods. Algorithms analysis and applications.* Texts in Applied Mathematics. Springer Science+Business Media, LLC, 2008.

[10] J. M. C. Pereira and J. C. F. Pereira. Fourier analysis of several finite difference schemes for the one-dimensional unsteady convection-diffusion equation. *International Journal for Numerial Methods in Fluids*, 36:417–439, 2001.

[11] S. Gottlieb and C.-W. Shu. Total variation diminishing Runge-Kutta schemes. *Mathematics of Computation*, 67(221):73–85, 1998.

[12] Raymond J. Spiteri and Steven J. Ruuth. A new class of optimal high-order strong-stability-preserving time discretization methods. Technical Report CS-2001-01, Faculty of Computer Science, Halifax, Canada, 2001.