SIAM J. Numer. Anal.
Vol. 50, No. 3, pp. 1058–1085

STABILITY CONDITIONS FOR THE NUMERICAL SOLUTION OF CONVECTION-DOMINATED PROBLEMS WITH SKEW-SYMMETRIC DISCRETIZATIONS*

ERWAN DERIAZ†

Abstract. This paper presents original and close to optimal stability conditions linking the time step and the space step, stronger than the CFL criterion, $\delta t \leq C\delta x^{\alpha}$ with $\alpha = \frac{2r}{2r-1}$, r an integer for some numerical schemes we produce when solving convection-dominated problems. We test this condition numerically and prove that it applies to nonlinear equations under smoothness assumptions.

Key words. CFL condition, von Neumann stability, transport equation, Euler equation, Runge–Kutta schemes, Adams–Bashforth schemes

AMS subject classifications. 65M02, 34D02, 41A02

DOI. 10.1137/100808472

1. Introduction.

In numerical fluid mechanics, many simulations for transport-dominated problems employ explicit second order time discretization schemes, of either Runge–Kutta [15, 8] or Adams–Bashforth [18, 19] type. Although widely in use and proved efficient, the stability domains of these order two numerical schemes (see Figure 1) exclude the (Oy) axis corresponding to transport problems. Nonetheless, actual experiments [24, 8] show that even in this case, a convergent solution can be obtained. If the problem admits a sufficiently smooth, classical solution, the second order time-stepping is stable at worst under a condition of type $\delta t \leq C(\delta x/u_{\max})^{4/3}$, where δt is the time step, δx the space step, and $u_{\max}$ the maximum velocity of the transport problem.

A close look at the stability condition provided by an analysis of von Neumann type applied to transport equation provides an explanation. To the best of our knowledge, this result is new (for instance, it is not presented in [23] which has collected the state of the art in numerical stability) despite the fact that it applies to a wide variety of numerical problems. Under some smoothness conditions, it readily extends to the Burgers equation, incompressible Euler equations, Navier–Stokes equations with a high Reynolds number on domains possibly bounded by walls, and conservation laws.

For the single step numerical method (i.e., the explicit Euler scheme), a stability result relying on a similar approach and providing a stability constraint of the type $\delta t \leq C(\delta x/u_{\max})^2$ has been presented by several authors [12, 17, 21]. The square originates from a completely different kind of numerical instability than the usual stability condition for the heat equation with explicit schemes. As we will see in this article, it comes from the order of tangency of the stability domain to the (Oy) axis and applies only to some first order schemes while for the heat equation it comes from the second derivative notwithstanding the order of the scheme. We present the generalization of this stability constraint to other schemes. Incidentally, we show that

*Received by the editors September 13, 2010; accepted for publication (in revised form) December 27, 2011; published electronically May 10, 2012.
http://www.siam.org/journals/sinum/50-3/80847.html
†Laboratoire de Mécanique, Modélisation et Procédés Propres, 38 rue Frédéric Joliot-Curie, 13451, Marseille Cedex 20, France (erwan.deriaz@l3m.univ-mrs.fr).

for transport-dominated problems there exists a direct connection between the order and the stability of numerical schemes.

As the numerical viscosity may stabilize the time scheme, this $\frac{4}{3}$-CFL[1] criterion applies essentially to pseudospectral methods and conservative numerical methods [22, 24]. A basic numerical experiment allows us to validate our approach.

The paper is organized as follows. First we recall the definition and the computation of the von Neumann stability. Then we focus on the linear transport problem, predicting a stability condition of the type $\delta t \leq C(\delta x/u)^{2r/(2r-1)}$ with r an integer, for several schemes. Then we construct numerical schemes for which such a stability condition appears for $r = 1, 2, 3, 4$ and corresponds to exponents equal to 2, $\frac{4}{3}$, $\frac{6}{5}$ and $\frac{8}{7}$. Finally we show how this stability criterion extends to nonlinear equations and to multicomponent transport equations (including wave equations).

2. The von Neumann stability condition. Let us consider the equation

$$\partial_t u = F\, u, \quad u(0, \cdot) = u_0, \tag{2.1}$$

where $u : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$, $(t, x) \mapsto u(t, x)$, and F is a linear operator. We denote by $\sigma(\xi)$ the symbol associated to F, i.e., $\widehat{F\, u}(t, \xi) = \sigma(\xi)\widehat{u}(t, \xi)$, where $u \mapsto \widehat{u}$ stands for the Fourier transform.[2]

In the following, we explain how to apply the von Neumann stability analysis as presented in [3, 13, 14, 16, 18, 21, 23]. We note $u_k \sim u(k\delta t, \cdot)$ the approximation at time $k\delta t$ for k an integer, δt denoting the time step. We consider a spectral discretization or that all the terms are orthogonally reprojected in our discretization space. The scheme can be of Runge–Kutta type, relying on the computation of intermediate time steps $u_{(\ell)}$

$$u_{(0)} = u_n, \quad u_{(\ell)} = \sum_{i=0}^{\ell-1} a_{\ell i}\, u_{(i)} + \delta t \sum_{i=0}^{\ell-1} b_{\ell i}\, F\, u_{(i)} \ \text{ for } 1 \leq \ell \leq s', \quad u_{n+1} = u_{(s')} \tag{2.2}$$

with $(a_{\ell i})_{\ell,i}$ and $(b_{\ell i})_{\ell,i}$ well chosen to ensure the accuracy of the integration.

Or it can be an explicit multistep (Adams–Bashforth) scheme involving the previous time steps:

$$u_{n+1} = \sum_{i=0}^{s} c_i\, u_{n-i} + \delta t \sum_{i=0}^{s} d_i\, F\, u_{n-i}. \tag{2.3}$$

We can also mix these two types of integration schemes:

$$\begin{aligned} u_{(\ell)} &= \sum_{i=1}^{\ell-1} a_{\ell i}\, u_{(i)} + \sum_{i=0}^{s} c_{\ell i}\, u_{n-i} + \delta t \sum_{i=0}^{\ell-1} b_{\ell i}\, F\, u_{(i)} \\ &\quad + \delta t \sum_{i=0}^{s} d_{\ell i}\, F\, u_{n-i} \text{ for } 1 \leq \ell \leq s', \\ u_{n+1} &= u_{(s')}. \end{aligned} \tag{2.4}$$

[1]CFL stands for the names of the three authors of the founding paper [5]: R. Courant, K. Friedrichs, and H. Lewy

[2]The Fourier transform of a function $f \in L^1(\mathbb{R})$ is noted $\hat{f}(\xi) = \int_{-\infty}^{+\infty} f(x)\, e^{-ix\xi} dx$; we recall that $f \mapsto \frac{1}{\sqrt{2\pi}} \hat{f}$ defines an isometry on $L^2(\mathbb{R})$.

The von Neumann stability analysis consists in isolating a Fourier mode ξ by taking $u_n(x) = \phi_n\, e^{i\xi x}$. Actually, if δx is the space step, then $-\frac{\pi}{\delta x} \leq \xi \leq \frac{\pi}{\delta x}$.

In the case when several previous time samples are necessary, like in the case of an Adams–Bashforth scheme, we set

$$X_n = \begin{pmatrix} u_n \\ u_{n-1} \\ \vdots \\ u_{n-s} \end{pmatrix}. \tag{2.5}$$

Remarking that each time we apply F to a term in (2.4), we also multiply this term by δt, it turns out that

$$X_{n+1} = M(\sigma(\xi)\delta t)X_n, \tag{2.6}$$

where setting $\zeta = \sigma(\xi)\delta t$, $M(\zeta)$ is a $(s+1)\times(s+1)$ square matrix whose elements are polynomials in ζ. Note that if F is a differential operator with derivatives of maximal order γ, then $|\zeta| \leq K\frac{\delta t}{\delta x^\gamma}$. In the case of hyperbolic equations, γ is equal to one.

Let $\lambda_0(\zeta), \dots, \lambda_s(\zeta)$ denote the eigenvalues of $M(\zeta)$. The spectral radius is defined by

$$\rho(M(\zeta)) = \max_{0\leq i\leq s}|\lambda_i(\zeta)|. \tag{2.7}$$

Then

$$\rho(M(\zeta))^n \leq \|M(\zeta)^n\| \leq \|M(\zeta)\|^n. \tag{2.8}$$

For almost every ζ, $\exists K_\zeta > 0$ such that $\forall n \geq 0$, $\|M(\zeta)^n\| \leq K_\zeta\, \rho(M(\zeta))^n$, where the constant K_ζ becomes large near the singularities of $M(\zeta)$. Actually, in numerical experiments, this does not play any crucial role. (See [21] for a complete discussion on this topic.) Hence, overlooking this latest point, the von Neumann stability of the scheme (2.4) is assured by

$$\forall i, \zeta, \quad |\lambda_i(\zeta)| \leq 1 + C\delta t \tag{2.9}$$

with C a positive constant independent of δx and δt. Sometimes, C is taken equal to zero to enforce an absolute stability. The assumption (2.9) allows any error ε_0 to stay bounded after an elapsed time T, since

$$\|\varepsilon_T\| = \|M(\zeta)^{T/\delta t}\varepsilon_0\| \leq K_\zeta\,(1 + C\delta t)^{T/\delta t}\|\varepsilon_0\| \leq K_\zeta\, e^{CT}\|\varepsilon_0\|. \tag{2.10}$$

The von Neumann stability domain of the scheme (2.4) is given by $\mathcal{S} = \{\zeta \in \mathbb{C}, \rho(M(\zeta)) \leq 1\}$. In Figures 1, 2, and 6, the x-axis represents the real part of ζ and the y-axis its imaginary part.

We represent the domain $\{\zeta \in \mathbb{C}, |\lambda_\ell(\zeta)| \leq 1\, \forall \ell\}$ delimited by the curves $\{\zeta \in \mathbb{C}, \lambda_\ell(\zeta) = e^{i\theta}, \theta \in [0, 2\pi[\}$ for $0 \leq \ell \leq s$.

In Figure 1 we plotted such domains for the first four Runge–Kutta schemes and the first five Adams–Bashforth schemes. The curves correspond to all the values of ζ, symbol of the operator $\delta t\, F$, for which there exists an eigenvalue with modulus equal to one. For order four and five Adams–Bashforth schemes, the stability domains only correspond to the semi-disks located on the left of the (Oy) axis. The loops on the right of the (Oy) axis do not correspond to any stable domain.

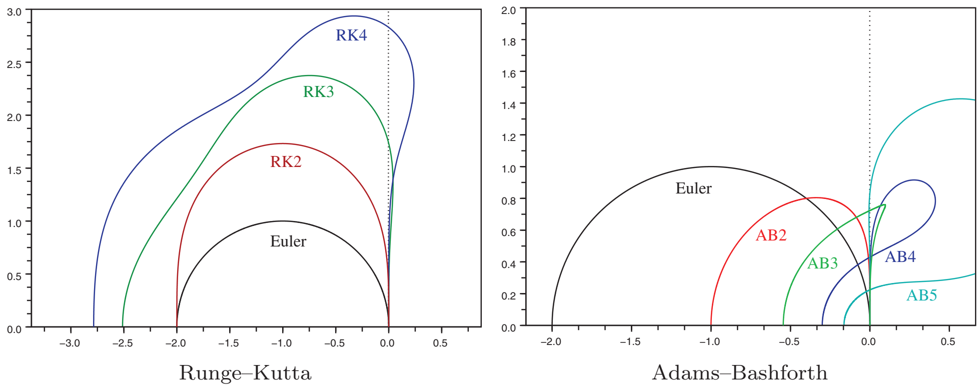

FIG. 1. *Von Neumann stability domains for the first four Runge–Kutta and five Adams–Bashforth schemes.*

We discretize the differential equation (2.1) with respect to the space variables. We assume that $u(t,x) = \sum_k u_k(t)\varphi_{\delta x,k}(x) \in V_{\delta x}$, where δx is a parameter corresponding to the space step. If $\overline{u} = (u_k)_k$, we obtain a discretized version of (2.1),

$$\partial_t \overline{u} = M_{\delta x}\,\overline{u}, \quad \overline{u}(0,\cdot) = \overline{u_0}, \tag{2.11}$$

with, for instance, $M_{\delta x} = \mathbb{P}_{\delta x} F$, where $\mathbb{P}_{\delta x}$ denotes the orthogonal projector on $V_{\delta x}$. To ensure the stability of the simulation, the stability domain with *thick line* must include the spectrum $Sp = \{\lambda(M_{\delta x})\}$ of the matrix $M_{\delta x}$. The thick line is due to the term $C\delta t$ in $\sup_{\xi \in Sp} \rho(M(\delta t \xi)) \le 1 + C\delta t$, where M from (2.6) depends on the temporal scheme.

The behavior of the stability domain along the (Oy) axis indicates how the scheme will be stable under the condition $\rho(M(\zeta)) \le 1 + C\delta t$—which gives more relevant stability conditions than the more classical $\rho(M(\zeta)) \le 1$—for convection-dominated problems. The next parts of our study will be dedicated to finding precise stability conditions on δt and δx in the frame of von Neumann stability.

3. Stability conditions for the transport equation. The von Neumann stability analysis for the transport equation presents some subtleties which explain why Runge–Kutta order two and Adams–Bashforth order two schemes are still used in numerical fluid dynamics although the transport operator $ia\xi$ for fixed $a \in \mathbb{R}^*$ and $\xi \in [-\frac{\pi}{\delta x}, \frac{\pi}{\delta x}]$ is located outside the stability domains of these schemes. In this section, we show that what matters is the behavior of the stability domain along the (Oy) axis.

3.1. Accurate theoretical stability condition. Let us consider the most basic transport equation:

$$\partial_t u + a\,\partial_x u = 0 \qquad \text{with} \;\; u : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R},\; (t,x) \mapsto u(t,x). \tag{3.1}$$

Since $\widehat{f'}(\xi) = i\xi \widehat{f}(\xi)$, the symbol of the operator $F\,u = -a\,\partial_x u$ is equal to $\sigma(\xi) = -i\,a\,\xi$.

As explained in the previous section, considering an explicit scheme (2.4), taking $u_n(x) = \phi_n\, e^{i\,\xi\,x}$, and setting X_n as in (2.5), we can write

$$X_{n+1} = A(\xi) X_n \tag{3.2}$$

with A a matrix whose coefficients are polynomials in $-i\,a\,\xi\,\delta t$.

In the case the numerical scheme is of Runge–Kutta type, $A(\xi)$ is a polynomial:

$$A(\xi) = \sum_{\ell=0}^{s} \beta_\ell (-i\, a\, \xi)^\ell \delta t^\ell. \tag{3.3}$$

The coefficients (β_ℓ) of this polynomial play an important role in our stability analysis. In [18], the polynomial $g(\zeta) = \sum_{\ell=0}^{s} \beta_\ell \zeta^\ell$ is called the *amplification factor*. We are able to compute the norm of $A(\xi)$ explicitly,

$$|A(\xi)|^2 = \sum_{\ell=0}^{s} S_\ell \delta t^{2\ell} a^{2\ell} \xi^{2\ell}, \tag{3.4}$$

with (assuming $\beta_j = 0$ for $j > s$)

$$S_\ell = \sum_{j=0}^{2\ell} (-1)^{\ell+j} \beta_j \beta_{2\ell-j}. \tag{3.5}$$

The von Neumann stability condition $|A(\xi)| \le 1 + C\delta t$ for all frequencies ξ remaining to the computational domain and for a given C, implies that for $\xi \in [-\frac{1}{\delta x}, \frac{1}{\delta x}]$ (usually the computational domain is rather $[-\frac{\pi}{\delta x}, \frac{\pi}{\delta x}]$, but we discard π for simplicity)

$$\sum_{\ell=0}^{s} S_\ell \delta t^{2\ell} a^{2\ell} \xi^{2\ell} \le 1 + 2C\delta t. \tag{3.6}$$

For the sake of consistency of the numerical scheme, $\beta_0 = \beta_1 = 1$ so $S_0 = \beta_0^2 = 1$. Then if $S_1 = \cdots = S_{r-1} = 0$ and $S_r > 0$ for a given integer r, we can write for small $\delta t \xi$,

$$|A(\xi)|^2 = 1 + S_r \delta t^{2r} a^{2r} \xi^{2r} + o(\delta t^{2r} \xi^{2r}). \tag{3.7}$$

With (3.6) it implies $S_r \delta t^{2r} a^{2r} \delta x^{-2r} \le 2C\delta t$, so $\delta t^{2r} a^{2r} \xi^{2r} \to 0$ for $\delta t \to 0$ implies $\delta t \xi = o(1)$; i.e., as $\xi \sim 1/\delta x$, we must have $\delta t = o(\delta x)$. Hence (3.7) is valid for all the computational domain $[-\frac{1}{\delta x}, \frac{1}{\delta x}]$. And the stability condition (3.6) is reduced to

$$S_r \delta t^{2r} a^{2r} \delta x^{-2r} \le 2C\delta t, \tag{3.8}$$

i.e.,

$$\delta t \le \left(\frac{2C}{S_r}\right)^{\frac{1}{2r-1}} \left(\frac{\delta x}{a}\right)^{\frac{2r}{2r-1}}. \tag{3.9}$$

This surprising stability condition is directly linked to the tangency of the stability domain $\mathcal{S} = \{\zeta \in \mathbb{C}, \max_i |\lambda_i(\zeta)| \le 1\}$ to the vertical axis (Oy). We have the following theorem.

THEOREM 3.1 (thick line stability theorem). *Consider a numerical time integration of type* (2.4) *with stability domain $\mathcal{S}$ bounded near zero by the parameterized curve $\{\zeta(\theta), \theta \in \mathcal{V}(0)\}$ with $\mathcal{V}(0)$ a real neighborhood of zero. If for some integer r, the Taylor expansion of ζ yields*

$$\zeta = i(\theta + o(\theta)) + T_{2r}\theta^{2r} + o(\theta^{2r}) \tag{3.10}$$

with $T_{2r} < 0$. Then the corresponding stability condition for the transport equation reads

$$\delta t \leq \left(\frac{C}{-T_{2r}} \right)^{\frac{1}{2r-1}} \left(\frac{\delta x}{a} \right)^{\frac{2r}{2r-1}}. \tag{3.11}$$

Note that $T_{2r} = -\frac{S_r}{2}$, where the quantity S_r defined by (3.5) provides (3.9).

Proof. The amplification factor $g(\zeta)$ is given by $g(\zeta) = \lambda_{\max}(A(\xi))$ with $|\lambda_{\max}(A(\xi))| = \max |\lambda_i(\zeta)|$ and $\{\lambda_i(\zeta)\}_{0\leq i\leq s}$ the eigenvalues of the matrix $A(\xi)$ from (3.2). There is a finite number of eigenvalues. Due to the polynomial form of the elements of the matrix $A(\xi)$, the eigenvalues can be written as holomorphic functions: $\lambda_i(\zeta) = \sum_{\ell\geq 0} \beta_\ell^{(i)} \zeta^\ell$. Among these eigenvalues $\lambda_i(\zeta)$, we consider the one such that $|\lambda_{i_0}(\zeta)| \geq |\lambda_i(\zeta)|$ for all indices i and complex number ζ in a neighborhood of 0. This corresponds to the largest sequence $(S_0, S_1, \ldots, S_r, \ldots)$ with S_ℓ defined by (3.5) for the usual order relation on sequences. Then $g(\zeta) = \lambda_{i_0}(\zeta)$ is a holomorphic function in a neighborhood of 0,

$$g(\zeta) = \sum_{\ell\geq 0} \beta_\ell \zeta^\ell = 1 + \zeta + \beta_2\zeta^2 + \cdots + \beta_s\zeta^s + \cdots, \tag{3.12}$$

where, for consistency we consider $\beta_0 = \beta_1 = 1$.

We already know that for (S_ℓ) given by (3.5) satisfying $S_1 = \cdots = S_{r-1} = 0$ and $S_r > 0$, the CFL stability condition (3.9) applies. We show that this same condition provides the tangency of the stability domain $\mathcal{S}$ to the (Oy) axis at zero.

Near O, let $\zeta = p + i\,q$,

$$g(\zeta) = 1 + p + iq + \beta_2(p + iq)^2 + \cdots + \beta_s(p + iq)^s + \cdots \tag{3.13}$$

with p and q independent variables close to zero. Then,

$$|g(\zeta)|^2 = (1 + p + \beta_2(p^2 - q^2) + \cdots)^2 + (q + 2\beta_2\, p\, q + \cdots)^2. \tag{3.14}$$

Looking for the first significant terms of this sum makes 1 and $2p$ appear. But we do not know which is the lowest power of q existing in this sum. Nevertheless, all the terms $p^\ell q^j$ with $\ell, j \geq 1$ and p^ℓ with $\ell \geq 2$ are negligible with respect to p, so we have from (3.13)

$$\begin{aligned} |g(\zeta)|^2 &= |1 + p + iq + \beta_2(iq)^2 + \cdots + \beta_s(iq)^s + \cdots|^2 + o(p) \\ &= 1 + 2p + S_r\, q^{2r} + o(p) + o(q^{2r}) \end{aligned} \tag{3.15}$$

with q^{2r} the lowest power of q with nonzero coefficient S_r (given by (3.5)).

As a result, the curve $\{|g(\zeta)| = 1\}$ is approximated by $p = -\frac{S_r}{2} q^{2r}$ near the origin. This curve can also be parameterized by $\theta \in [-\pi, \pi]$ in $\{g(\zeta) = e^{i\theta}, \theta \in [-\pi, \pi]\}$. For multistep schemes (see section 6), it is convenient to express ζ as a function of θ and to write it as a Taylor series:

$$\zeta = \sum_{\ell\geq 1} iT_{2\ell-1}\theta^{2\ell-1} + T_{2\ell}\theta^{2\ell}. \tag{3.16}$$

Then, for $\theta \in \mathbb{R}$ close to 0

$$\zeta = i(\theta + o(\theta)) - \frac{S_r}{2}\theta^{2r} + o(\theta^{2r}), \tag{3.17}$$

so $T_{2r} = -\frac{S_r}{2}$. Hence this tangency implies the CFL (3.9). $\square$

Theorem 3.2. *An order $2p$ numerical time integration applied to the transport equation is, at worst, stable under the CFL-like condition:*

$$\delta t \leq C \left(\frac{\delta x}{a}\right)^{\frac{2p+2}{2p+1}}. \tag{3.18}$$

Proof. For an order $2p$ scheme, we have

$$u_{n+1} = u_n + \delta t\, \partial_t u_n + \frac{\delta t^2}{2}\partial_t^2 u_n + \cdots + \frac{\delta t^{2p}}{(2p)!}\partial_t^{2p} u_n + o(\delta t^{2p}). \tag{3.19}$$

The transport operator F commutes with ∂_t. So iterating $\partial_t u = F\, u$ we obtain $\partial_t^\ell u_n = F^\ell(u_n)$. Hence (3.19) yields the amplification factor

$$g(\zeta) = 1 + \zeta + \frac{\zeta^2}{2} + \cdots + \frac{\zeta^{2p}}{(2p)!} + o(\zeta^{2p}) \tag{3.20}$$

with $o()$ gathering the negligible terms under the condition $\delta t = o(\delta x)$. In this case, the (β_ℓ) of (3.12) are given by $\beta_\ell = \frac{1}{\ell!}$. Then for $q \in [1, p]$, the coefficients S_q of the sum (3.4) are given by

$$S_q = \sum_{\ell=0}^{2q} (-1)^{(q-\ell)} \frac{1}{\ell!}\, \frac{1}{(2q-\ell)!} = \frac{(-1)^q}{(2q)!} \sum_{\ell=0}^{2q} \mathrm{C}_{2q}^\ell (-1)^\ell = 0. \tag{3.21}$$

Hence, in the worst case regarding the stability, the first nonzero significant term in the sum (3.4) is $S_{p+1}\delta t^{2p+2} a^{2p+2} \xi^{2p+2}$ with $S_{p+1} > 0$ implying the stability condition (3.18). If $S_{p+1} < 0$, then a linear CFL condition is sufficient. ∎

3.2. Examples with usual schemes. We apply our analysis to some popular schemes in fluid dynamics. This provides the following stability conditions for some of the most used schemes for transport problem (3.1):

- The simplest example is the Euler explicit scheme, order one in time:

$$u_{n+1} = u_n - \delta t\; a \partial_x u_n. \tag{3.22}$$

  For this scheme, $g(\zeta) = 1 + \zeta$ so $r = 1$, $S_1 = 1$ and we find the stability condition

$$\delta t \leq 2C \left(\frac{\delta x}{a}\right)^2. \tag{3.23}$$

- An improved version of this scheme allows us to construct an order two centered scheme:

$$\begin{cases} u_{n+1/2} = u_n - \frac{\delta t}{2} a \partial_x u_n, \\ \quad u_{n+1} = u_n - \delta t\; a \partial_x u_{n+1/2}. \end{cases} \tag{3.24}$$

  For this scheme, $g(\zeta) = 1 + \zeta + \frac{1}{2}\zeta^2$ so $r = 2$ because $S_1 = 0$ and $S_2 = \frac{1}{4}$. Compared to the previous case, the stability is improved:

$$\delta t \leq 2C^{1/3} \left(\frac{\delta x}{a}\right)^{4/3}. \tag{3.25}$$

- For a Runge–Kutta scheme of order four, we have

$$\begin{cases} u_{n(1)} = u_n - \frac{\delta t}{2} a\partial_x u_n, \\ u_{n(2)} = u_n - \frac{\delta t}{2} a\partial_x u_{n(1)}, \\ u_{n(3)} = u_n - \delta t\ a\partial_x u_{n(2)}, \\ u_{n+1} = u_n - \frac{\delta t}{6} a\partial_x u_n - \frac{\delta t}{3} a\partial_x u_{n(1)} - \frac{\delta t}{3} a\partial_x u_{n(2)} - \frac{\delta t}{6} a\partial_x u_{n(3)}. \end{cases} \tag{3.26}$$

From the amplification factor $g(\zeta) = 1 + \zeta + \frac{\zeta^2}{2} + \frac{\zeta^3}{6} + \frac{\zeta^4}{24}$ we infer

$$S_1 = S_2 = 0 \quad \text{and} \quad S_3 = -\frac{1}{72},\ S_4 = \frac{1}{576}. \tag{3.27}$$

As $S_3 < 0$, our study doesn't apply to this case, and the stability domain, Figure 1, indicates that a classical linear CFL condition has to be satisfied.

- The order 5 Runge–Kutta scheme from [6, p. 115] provides the amplification factor $g(\zeta) = 1 + \zeta + \frac{\zeta^2}{2} + \frac{\zeta^3}{6} + \frac{\zeta^4}{24} + \frac{\zeta^5}{120} + \frac{\zeta^6}{1280}$. Therefore it is stable under the condition

$$\delta t \le \left(\frac{11520}{7}\right)^{1/5} C^{1/5} \left(\frac{\delta x}{a}\right)^{6/5}, \qquad \left(\frac{11520}{7}\right)^{1/5} \sim 4.398. \tag{3.28}$$

- The order two Adams–Bashforth scheme goes as follows:

$$u_{n+1} = u_n - \frac{3}{2}\delta t\ a\partial_x u_n + \frac{1}{2}\delta t\ a\partial_x u_{n-1}. \tag{3.29}$$

So, according to section 2, we consider $X_n = \left[\begin{smallmatrix} u_n \\ u_{n-1} \end{smallmatrix}\right]$ and we apply the numerical scheme to a pure Fourier mode $\phi_n\, e^{i\xi\, x}$. We obtain

$$X_{n+1} = \begin{bmatrix} 1 + \frac{3}{2}\zeta & -\frac{\zeta}{2} \\ 1 & 0 \end{bmatrix} X_n \tag{3.30}$$

with $\zeta = -ia\,\delta t\,\xi$.

We compute the eigenvalue of this 2×2 matrix; the characteristic polynomial is given by $\chi(Y) = Y^2 - (1 + \frac{3}{2}\zeta)Y + \frac{\zeta}{2}$. Owing to the fact that $\delta t = o(\delta x)$, we have $\zeta \to 0$. An expansion of the larger eigenvalue Y_0 in terms of powers of ζ provides

$$Y_0 = 1 + \zeta + \frac{\zeta^2}{2} - \frac{\zeta^3}{4} - \frac{\zeta^4}{8} + o(\zeta^4). \tag{3.31}$$

With $\zeta = -ia\frac{\delta t}{\delta x}$, we obtain

$$|Y_0| = 1 + \frac{1}{4} a^4 \frac{\delta t^4}{\delta x^4} + o\left(\frac{\delta t^4}{\delta x^4}\right). \tag{3.32}$$

As we want $|Y_0| \le 1 + C\delta t$, this drives to the following stability condition:

$$\delta t \le 2^{2/3} C^{1/3} \left(\frac{\delta x}{a}\right)^{4/3}. \tag{3.33}$$

Therefore, two popular second order schemes, Runge–Kutta two (RK2) and Adams–Bashforth two (AB2), require a CFL-like condition, $\delta t \le C\delta x^{4/3}$. The $\delta t_{\max}$ is $2^{1/3}$ larger for RK2 than for AB2, but RK2 necessitates twice as many computations as

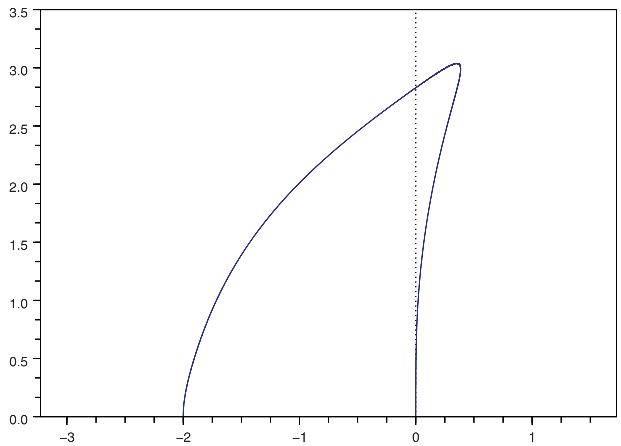

FIG. 2. *Von Neumann stability domain for the pseudo-leap-frog scheme* (3.34).

AB2 for each time step. So, regarding only the stability, AB2 is $2^{2/3}$ cheaper than RK2.

Not all the second order numerical schemes need to satisfy a 4/3-CFL condition. For instance, the leap-frog scheme calls a usual linear CFL stability condition. The following second order scheme is also stable under a linear CFL condition:

$$u_{n+1} = u_n + \delta t\, a\partial_x \left(\frac{u_n + u_{n-1}}{2} + \delta t\, a\partial_x u_n \right). \tag{3.34}$$

Its stability domain is drawn in Figure 2. The fact that $r = 2$ with $S_2 < 0$ in (3.17) is reflected by a tangent to (Oy) oriented to the right.

3.3. Effect of the space discretization. The space discretization impacts the stability condition (3.18) if it dissipates or creates energy, as do the upwind and downwind schemes. Graphically this means that the spectra of these discretizations for the transport operator $F : u \mapsto -a\partial_x u$ are not contained in the (Oy) axis; see Figure 3.

In the frame of the von Neumann stability analysis we consider the function $u(x) = e^{i\xi x}$ for $\xi \in \mathbb{R}$. Then, having a closer look at the three academic cases for finite differences, we obtain the following:

1. The downwind schemes are always unstable. For the first order downwind scheme (see Figure 3 for its spectrum)

$$a\frac{\partial u}{\partial x} + O(\delta x) = a\frac{u(x) - u(x - \delta x)}{\delta x} = a\, e^{i\xi x}\frac{-e^{-i\xi\delta x} + 1}{\delta x} \tag{3.35}$$

for $a < 0$ provides the symbol $\sigma = -a\frac{-e^{-i\,\xi\,\delta x}+1}{\delta x}$ instead of $-i\,a\,\xi$ in formula (3.3). Combined with the Euler scheme for time integration, the amplification factor $G(\sigma) = 1 + \delta t\,\sigma$ becomes

$$|G|^2 = 1 - 2a\frac{\delta t}{\delta x}\left(1 + \frac{\delta t}{\delta x}\right)(1 - \cos(\xi\,\delta x)) \tag{3.36}$$

and the error

$$\varepsilon_T \sim |G|^{\frac{T}{\delta t}}\varepsilon_0 \sim e^{\frac{-2a}{\delta x}}\varepsilon_0 \tag{3.37}$$

goes unconditionally to $+\infty$.

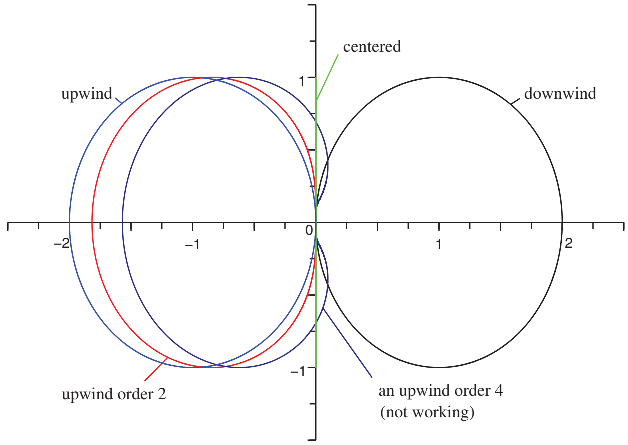

FIG. 3. *Spectra of various finite difference schemes for space differentiation. The numerical method is stable if the spectrum of the discretized differentiation fits into the domain of stability of the temporal scheme (as those of Figure 1). Here, the spectra have been normalized in order to have the same vertical size.*

2. The centered space discretizations satisfy $Sp \subset i\mathbb{R}$, where $Sp = \{\sigma(\xi), \xi \in [-\frac{\pi}{\delta x}, \frac{\pi}{\delta x}]\}$. They include most of the compact finite difference schemes. For instance, the usual centered scheme

$$\frac{\partial u}{\partial x} + O(\delta x^2) = \frac{u(x+\delta x) - u(x-\delta x)}{2\delta x} = e^{i\xi x}\frac{e^{i\xi\delta x} - e^{-i\xi\delta x}}{2\delta x} \tag{3.38}$$

has a spectrum given by $\sigma = i\frac{\sin(\xi\,\delta x)}{\delta x}$ which goes along the (Oy) axis, so its distance to the domain of stability of the time scheme goes almost the same as in the spectral case. The stability results (3.11) presented in this section apply fully to this case with a constant C which depends on the space discretization.
3. The upwind schemes can be unconditionally unstable if part of their spectra is located on the right side of the (Oy) axis. (See the spectrum of the order four upwind scheme plotted in Figure 3.) If their spectra remain in the left part of the complex plane, then the exponent in (3.11) is modified in the following way: assume that the domain of stability of the time discretization satisfies

$$g(\theta) = i\theta + T_{2q}\theta^{2q} + o(i\theta) + o(\theta^{2q}) \tag{3.39}$$

in a neighborhood of 0, with $T_{2q} < 0$ and $q \in \mathbb{N}$; assume that the spectrum of the discretized derivative satisfies

$$\sigma(\theta) = i\theta + V_{2p}\theta^{2p} + o(i\theta) + o(\theta^{2p}) \tag{3.40}$$

with $V_{2p} < 0$ (i.e., upwind scheme) and $p \in \mathbb{N}$. Then the thick line stability condition (3.11) becomes
   - the CFL condition $\delta t \leq C\delta x$, with C a constant independent of δt and δx if $p \leq q$,
   - a mitigation of the nonlinear condition (3.9)

$$\delta t \leq C\delta x^{\frac{q(2p-1)}{p(2q-1)}} \tag{3.41}$$

   with C a constant independent of δt and δx if $p \geq q$.

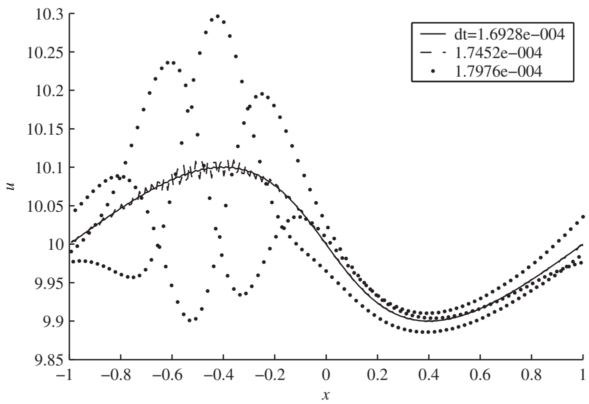

FIG. 4. *Numerical solution obtained at time $T = 1$ for three different time steps:* $0.97\,\delta t_{\max}$*,* $\delta t_{\max}$*, and* $1.03\,\delta t_{\max}$ *for* $N = 256$ *with RK2 (order two Runge–Kutta) with a* $\delta t_{\max}$ *corresponding to* $K = 5$*; i.e.,* $||u_T||_{TV} = 5||u_0(x)||_{TV}$.

The details of the proofs and the numerical tests for these assertions will be presented in a further article. Note that the case $p = +\infty$ (i.e., switching to a centered finite difference scheme) makes the condition (3.11) appear.

4. Numerical experiment with the Burgers equation. In order to test our assertions, we proceed to a numerical experiment with the inviscid Burgers equation. Although this is a nonlinear equation, we choose initial conditions such that it assimilates to a transport equation: the sinusoidal part represents only 1% of the transport amplitude. So technically, regarding the stability, it behaves like a transport equation. Then the various Fourier modes are *naturally* activated during the experiment.

$$\partial_t u + u\partial_x u = 0 \quad \text{for} \quad (t, x) \in [0, T] \times \mathbb{T}, \quad \text{and} \quad u(0, \cdot) = u_0. \tag{4.1}$$

In sections 5 and 6 we show numerical evidence that stability conditions (3.23), (3.25), (3.33), and (3.11) hold for this problem (replacing a by $\|u\|_{L^\infty}$).

We solve (4.1) numerically using a Fourier pseudospectral method [2]. The scheme is de-aliased by truncation. Most of the time integration methods presented in this paper are tested on this classical basic problem.

The initial condition for the numerical experiment is $u_0(x) = 10 - 0.1 \sin(\pi x)$ in a periodic domain $x \in \Omega = [-1, 1]$. For $t < t_{\max} = 10/\pi$ the equation admits a smooth exact solution $u(x, t) = u_0(a)$, where $a = a(x, t)$ is solution of the equation $a - x + u_0(a)t = 0$. However, the numerical solution is only sought for $t \in [0, 1]$ in order to satisfy some regularity requirements on the solution (see Proposition 7.1). To determine the admissibility of the numerical solution, we apply a criterion based on the total variation norm (which is expected to be constant): the numerical solution u_n has to satisfy $||u_n||_{TV} \le K||u_0(x)||_{TV}$ with $K = 1.1$ $\forall n$ such that $n\delta t \le T = 1$.

The $\delta t_{\max}$ we compute has very little dependence on the divergence criterion K. Actually, below $\delta t_{\max}$ (97%), the numerical solution shows no spurious oscillations, while above it (103%), these oscillations create some kind of explosion destroying the profile of the solution completely; see Figure 4.

The computations are performed for different numbers of grid points, $16 \le N \le 7758$. For each N, we find $\delta t_{\max}$ by dichotomy with a 0.5% accuracy. The results are represented as $\delta t_{\max}(N)$ curves in Figure 5. They evidence the theoretically

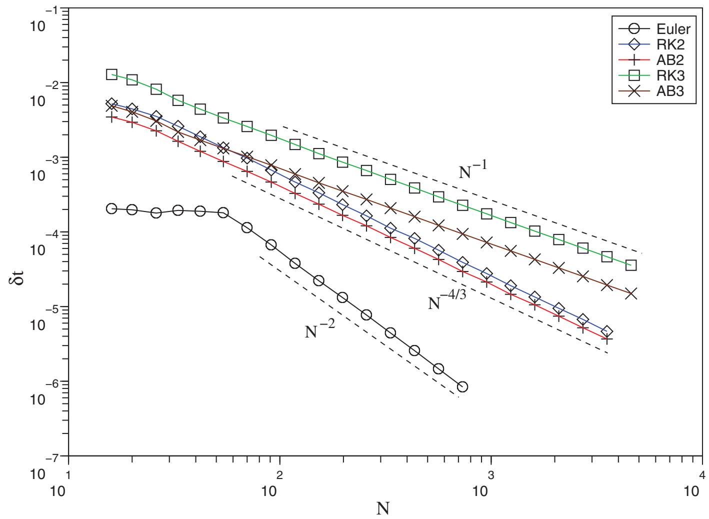

FIG. 5. *Maximal time step $\delta t_{\max}$ depending on the number of points N for Runge–Kutta schemes and Adams–Bashforth schemes, obtained experimentally.*

predicted power law $\delta t_{\max} = CN^{\alpha}$ when the number of grid points is sufficiently large. The explicit Euler scheme displays $\alpha = -2$ slope in log-log scale. The two curves corresponding to the second order schemes asymptotically both show an asymptotic slope equal to $CN^{-\frac{4}{3}}$, but the constant C is $2^{1/3}$ times larger for the Runge–Kutta scheme.

When the order is increasing to 3 and 4 for Runge–Kutta schemes and Adams–Bashforth schemes, the slope equals -1. But, while the constant C increases with the order for Runge–Kutta (yielding a larger stability domain), it diminishes for Adams–Bashforth schemes with the increasing order (see Figure 1 or, e.g., [2]).

5. Simple 2N-storage numerical schemes with "shrinking CFL" stability conditions. In order to illustrate the phenomenon presented in section 3, we construct numerical schemes having stability conditions of the type $\delta t \le C\left(\frac{\delta x}{a}\right)^{\frac{2r}{2r-1}}$ and which only necessitate two time levels to be stored in the computer memory. Four of the five schemes presented here need to satisfy this stability condition with exponents $2r/(2r-1)$ different from 1: 2, $\frac{4}{3}$, $\frac{6}{5}$, and $\frac{8}{7}$. All of these numerical schemes are of order two, so they show relatively poor consistency given the number of intermediate steps. Other efficient low storage schemes can be found in [18] and [11].

To solve the equation

$$\partial_t u = F(u), \tag{5.1}$$

let us consider the following family of schemes:

$$\begin{aligned} u_{(0)} &= u_n \\ u_{(1)} &= u_n + \alpha_p \delta t F(u_{(0)}) \\ &\dots \\ u_{(\ell)} &= u_n + \alpha_{p-\ell} \delta t F(u_{(\ell-1)}) \\ &\dots \\ u_{n+1} &= u_n + \alpha_1 \delta t F(u_{(p-1)}). \end{aligned} \tag{5.2}$$

These can also be written

(5.3)
$$u_{n+1} = u_n+\alpha_1\delta tF(u_n+\alpha_2\delta tF(u_n+\alpha_3\delta tF(u_n+\cdots+\alpha_{p-1}\delta tF(u_n+\alpha_p\delta tF(u_n))\dots))).$$

If F is *linear*, this corresponds to

$$u_{n+1} = u_n + \beta_1\delta tFu_n + \beta_2\delta t^2F^2u_n + \beta_3\delta t^3F^3u_n + \cdots + \beta_p\delta t^pF^pu_n \tag{5.4}$$

with $\beta_m = \prod_{\ell=1}^m \alpha_\ell$. Owing to $F^\ell u = \partial_t^\ell u$,

$$u_{n+1} = u_n + \beta_1\delta t\partial_t u_n + \beta_2\delta t^2\partial_t^2 u_n + \beta_3\delta t^3\partial_t^3 u_n + \cdots + \beta_p\delta t^p\partial_t^p u_n. \tag{5.5}$$

Here we recognize an expansion similar to the Taylor expansion of the function u_n, and we are able to tell exactly the order of the scheme for linear equations by comparing the coefficients β_ℓ with those of the Taylor expansion which is provided by

$$u_{n+1} = \sum_{\ell=0}^{+\infty}\frac{\delta t^\ell}{\ell!}\partial_t^\ell u_n = u_n + \delta t\partial_t u_n + \frac{1}{2}\delta t^2\partial_t^2 u_n + \frac{1}{6}\delta t^3\partial_t^3 u_n + \cdots \tag{5.6}$$

and the smallest ℓ such that $\beta_{\ell+1} \neq 1/(\ell+1)!$ indicates the order of the scheme. Note that this holds only if F is linear or if the order of the scheme is less than or equal to two. The interest of such schemes is that the coefficients α_ℓ are easily deduced from the β_ℓ.

We assume that F is a convection operator. Using the stability analysis section 3, we know that the values of

$$S_\ell = \beta_\ell^2 - 2\beta_{\ell-1}\beta_{\ell+1} + 2\beta_{\ell-2}\beta_{\ell+2} - \dots \tag{5.7}$$

provide the stability condition. We verify the validity of this stability condition using the numerical test from section 4.

For a given p in (5.3), maximizing the number of S_ℓ equal to zero leads to the following schemes of order two (except the first one) and stability conditions:

- With $\beta_1 = 1$ and $\beta_\ell = 0$ for $\ell \geq 2$, this is the Euler explicit scheme,

  $$u_{n+1} = u_n + \delta tFu_n, \tag{5.8}$$

  $\beta_2 \neq \frac{1}{2}$, so it is of order one, and $S_1 = 1$ implies $\delta t \leq 2C(\frac{\delta x}{a})^2$.
- With $\beta_1 = 1$, $\beta_2 = 1/2$, and $\beta_\ell = 0$ for $\ell \geq 3$, this is a second order Runge–Kutta scheme,

  $$u_{n+1} = u_n + \delta tF\left(u_n + \frac{1}{2}\delta tFu_n\right), \tag{5.9}$$

  $\beta_3 \neq \frac{1}{6}$, so it is of order two, and $S_2 = 1/4$ implies (3.25) $\delta t \leq 2\,C^{1/3}(\frac{\delta x}{a})^{4/3}$.
- With $\beta_1 = 1$, $\beta_2 = 1/2$, $\beta_3 = 1/8$ and $\beta_\ell = 0$ for $\ell \geq 4$, it is an order two numerical scheme ($\beta_3 \neq 1/6$),

  $$\text{(scheme 3)}\quad u_{n+1} = u_n + \delta tF\left(u_n + \frac{1}{2}\delta tF\left(u_n + \frac{1}{4}\delta tFu_n\right)\right), \tag{5.10}$$

  and as $S_1 = S_2 = 0$ and $S_3 = 1/64$, we have the stability condition

  $$\delta t \leq 2^{7/5}\,C^{1/5}\left(\frac{\delta x}{a}\right)^{6/5}. \tag{5.11}$$

Table 1

Coefficients β_m for different m. In the expression for $m = 7$, α is a real solution of the equation $\alpha^3 - 9\alpha^2 - \alpha + 1 = 0$.

m	1	2	3	4	5	6	7
β_m	1	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{3-2\sqrt{2}}{8}$	$\frac{5\sqrt{5}-11}{64}$	$\frac{26-15\sqrt{3}}{16}$	$-\frac{1}{64} + \frac{7\alpha}{128}$
$\left(\frac{1}{\beta_m^2}\right)^{\frac{1}{2m-1}}$	1	1.587...	2.297...	2.997...	3.687...	3.395...	5.045...

- The schemes verifying $\beta_\ell = 0$ for $\ell \geq 5$ and $S_1 = S_2 = S_3 = 0$ are given by $\beta_1 = 1$, $\beta_2 = 1/2$, $\beta_3 = \frac{2\pm\sqrt{2}}{4}$, and $\beta_4 = \frac{3\pm2\sqrt{2}}{8}$. If we choose the minus sign for β_3 and β_4, this means

$$\text{(scheme 4)} \quad u_{n+1} = u_n + \delta t F \left(u_n + \frac{1}{2}\delta t F\left(u_n + \frac{2-\sqrt{2}}{2}\delta t F\left(u_n + \frac{2-\sqrt{2}}{4}\delta t F u_n\right)\right)\right). \tag{5.12}$$

It is a second order scheme and has to satisfy the CFL-like stability condition

$$\delta t \leq \left(\frac{2C}{\beta_4^2}\right)^{1/7}\left(\frac{\delta x}{a}\right)^{8/7}. \tag{5.13}$$

- In the general case, we consider $\beta_0 = 1, \beta_1 = 1, \beta_2, \dots, \beta_m$, and for $1 \leq \ell \leq m-1$, $S_\ell = 0$. As $S_m = \beta_m^2 > 0$, the schemes resulting from this system of equations has to satisfy

$$\delta t \leq \left(\frac{2C}{\beta_m^2}\right)^{\frac{1}{2m-1}}\left(\frac{\delta x}{a}\right)^{\frac{2m}{2m-1}}. \tag{5.14}$$

For β_m positive and minimum, it results in the constants indicated in Table 1. On the other hand, if we impose the order to be three with five nonzero β_ℓ, maximizing the number of S_ℓ equal to zero provides $\beta_1 = 1$, $\beta_2 = 1/2$, $\beta_3 = 1/6$, $\beta_4 = 1/24$, and $\beta_5 = 1/144$. Hence it is written

$$\text{(scheme 5)} \quad u_{n+1} = u_n + \delta t F\left(u_n + \frac{\delta t}{2}F\left(u_n + \frac{\delta t}{3}F\left(u_n + \frac{\delta t}{4}F\left(u_n + \frac{\delta t}{6}F u_n\right)\right)\right)\right). \tag{5.15}$$

As $S_4 = \beta_4^2 - 2\beta_3\beta_5 < 0$, a classical linear CFL condition $\delta t \leq C\frac{\delta x}{a}$ applies. Even as $\beta_\ell = 1/\ell!$ until $\ell = 4$, this scheme is of order four.

Figure 6 shows the stability domains of these schemes, and the numerical experiments presented in Figure 7 confirm our predictions regarding the stability condition on δt.

Remark 5.1. Maximizing the tangency of the stability domain to the (Oy) axis is equivalent to optimizing the energy conservation scale by scale. This explains why people simulating convection-dominated problems tend to prefer the Crank–Nicholson scheme $u_{n+1} = u_n + \frac{\delta t}{2}(Fu_n + Fu_{n+1})$ (see [7], for instance), whose stability domain boundary coincides with the (Oy) axis.

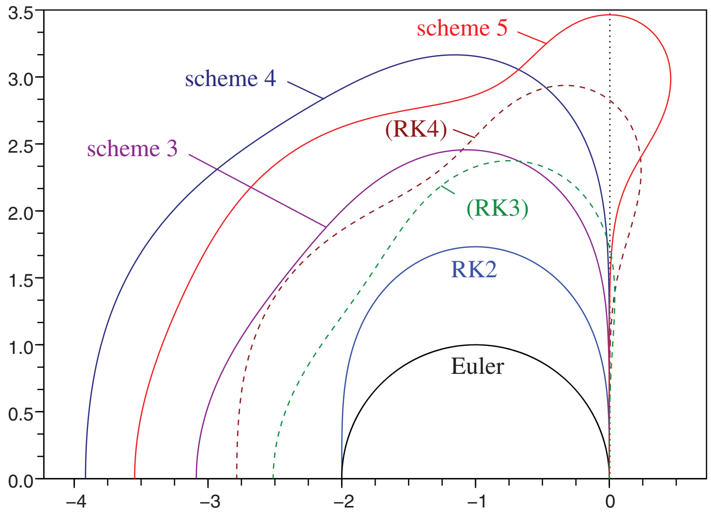

FIG. 6. *Von Neumann stability domains for schemes of Runge–Kutta type.*

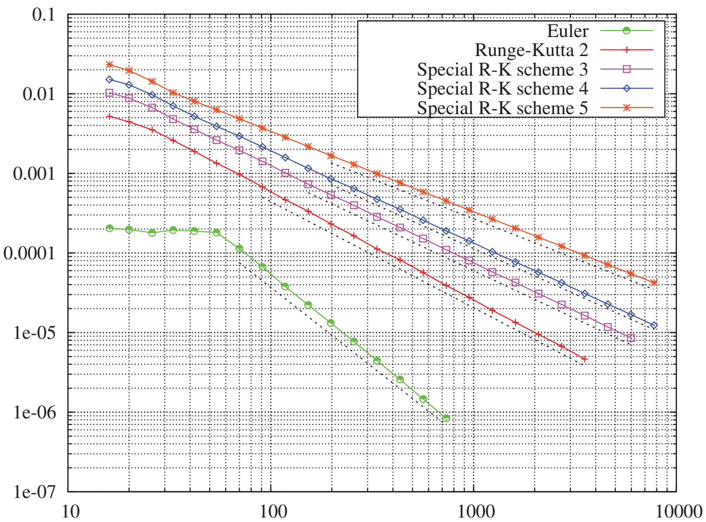

FIG. 7. *Maximal time step ensuring stability for the test case of section 4 with the Runge–Kutta schemes stable under the condition* $\delta t \leq C\delta x^{\frac{2r}{2r-1}}$ *for* $r = 1, 2, 3,$ *and* 4 *whose slops we plot in parallel beneath the experimental results.* (Ox) *axis represents the number of points* $N = \frac{2}{\delta x}$ *and* (Oy) *the maximal time step* $\delta t_{\max}$ *above which the numerical solution becomes unstable.*

6. Adams–Bashforth schemes with "shrinking CFL" stability conditions. Let us consider an Adams–Bashforth scheme with coefficients (α_k):

$$u_{n+1} = u_n + \sum_{k=0}^{K} \alpha_k \, \delta t \, F(u_{n-k}). \tag{6.1}$$

The order of scheme (6.1) depends on the sums

$$\Upsilon_\ell = \sum_{k=0}^{K} k^\ell \alpha_k \qquad \text{for } 0 \leq \ell \leq K\ . \tag{6.2}$$

The scheme has order m iff for $0 \le \ell \le m-1$, $\Upsilon_\ell = \frac{(-1)^\ell}{\ell+1}$ (see [18]). Solving the system for $m = K+1$ provides the Adams–Bashforth scheme of order $K+1$, properly speaking.

The von Neumann stability domain is computed as indicated in section 2. Let

$$X_n = \begin{pmatrix} u_n \\ u_{n-1} \\ \vdots \\ u_{n-K} \end{pmatrix} \qquad \text{and} \qquad \widehat{\delta t\, F(u_\ell)} = \zeta \widehat{u_\ell} \;\; \text{with} \;\; \zeta \in \mathbb{C}. \tag{6.3}$$

Then $\widehat{X_{n+1}} = M(\zeta)\widehat{X_n}$ with the matrix $M(\zeta) \in \mathcal{M}_{K+1}(\mathbb{C})$ given by

$$M(\zeta) = \begin{bmatrix} 1+\alpha_0\zeta & \alpha_1\zeta & \dots & \dots & \alpha_K\zeta \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}. \tag{6.4}$$

The characteristic polynomial is given by

$$P(X) = \det(X\,Id - M(\zeta)) = X^{K+1} - X^K - \sum_{k=0}^{K} \alpha_k \zeta X^{K-k}. \tag{6.5}$$

As the eigenvalues of this polynomial provide the multiplication factor of the scheme (6.1), the boundaries of the stability domain are therefore obtained by considering the curve $\{\zeta \in \mathbb{C}$ such that (s.t.) $\exists \theta \in [-\pi,\pi], P(e^{i\theta}) = 0\}$, i.e.,

$$\zeta = \frac{e^{i\theta} - 1}{\sum_{k=0}^{K} \alpha_k\, e^{-ik\theta}}, \qquad \theta \in [-\pi,\pi]. \tag{6.6}$$

According to Theorem 3.1, the stability condition depends on the tangency to the imaginary axis (Oy) obtained for θ close to 0. Assuming order one at least (i.e., $\Upsilon_0 = 1$), a Taylor expansion of expression (6.6) provides

$$\zeta = \sum_{r\ge 1} i^r \theta^r \sum_{\substack{p+q=r \\ p\ge 1,\, q\ge 0}} \frac{(-1)^q}{p!} \sum_{\sum_{n\ge 1} n\kappa_n = q} (-1)^{\sum_{n\ge 1}\kappa_n} \frac{\left(\sum_{n\ge 1}\kappa_n\right)!}{\prod_{n\ge 1}\kappa_n!} \prod_{n\ge 1} \left(\frac{\Upsilon_n}{n!}\right)^{\kappa_n}. \tag{6.7}$$

The first two elements of this sum are given by

$$T_2 = -\Upsilon_1 - \frac{1}{2}, \quad T_4 = \frac{1}{6}\Upsilon_1 - \frac{1}{4}\Upsilon_2 - \Upsilon_1\Upsilon_2 + \frac{1}{6}\Upsilon_3 + \frac{1}{2}\Upsilon_1^2 + \Upsilon_1^3 \tag{6.8}$$

with Υ_ℓ from (6.2).

For a given K, maximizing the tangency to (Oy) (i.e., the number m s.t. $T_{2\ell}$ is equal to 0 for $l < m$) provides the following numerical schemes:

- For $K = 1$, $T_2 = 0$ implies $\alpha_0 = \frac{3}{2}$ and $\alpha_1 = -\frac{1}{2}$, i.e., the Adams–Bashforth scheme of order two. As $T_4 = -\frac{1}{4}$, it is stable under the condition (3.33) $\delta t \le 2^{2/3} C^{1/3} \left(\frac{\delta x}{a}\right)^{4/3}$.

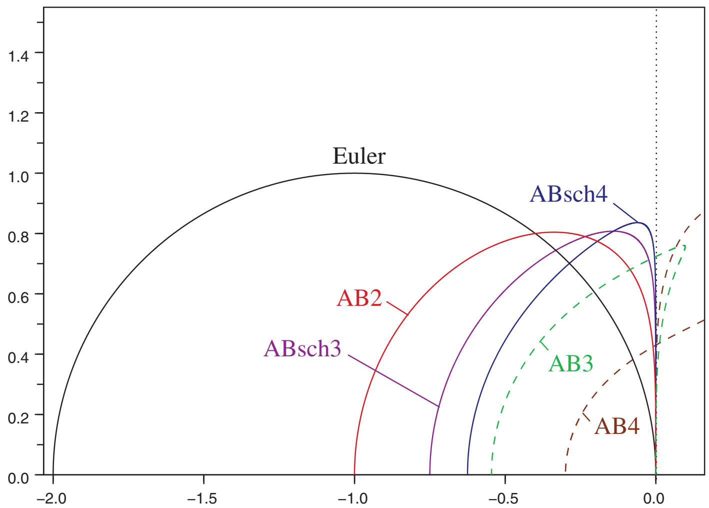

FIG. 8. *Von Neumann stability domains for modified Adams–Bashforth schemes maximizing the tangency to the* (Oy) *axis.*

- With three time steps, $T_2 = T_4 = 0$ leads to the scheme we call (ABsch3) with $\alpha_0 = \frac{5}{3}$, $\alpha_1 = -\frac{5}{6}$, and $\alpha_2 = \frac{1}{6}$. Given that $\Upsilon_1 = -1/2$ and $\Upsilon_2 = -1/6 \neq 1/3$, it is of order two. And $T_6 = -1/12$ induces the CFL condition

$$\delta t \leq 12^{1/5} C^{1/5} \left(\frac{\delta x}{a}\right)^{6/5}. \tag{6.9}$$

- With four time steps, enforcing $T_2 = T_4 = T_6 = 0$ yields the scheme (ABsch4) with

$$(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = \left(\frac{7}{4}, -\frac{21}{20}, \frac{7}{20}, -\frac{1}{20}\right). \tag{6.10}$$

  As $\Upsilon_1 = -1/2$ and $\Upsilon_2 = 1/10 \neq 1/3$, this is also a second order scheme. On the other hand, we have $T_8 = -\frac{1}{40}$, so this scheme is stable under the condition

$$\delta t \leq 40^{1/7} C^{1/7} \left(\frac{\delta x}{a}\right)^{8/7}. \tag{6.11}$$

We plot the stability domains corresponding to these schemes in Figure 8 and verify our stability predictions with the Burgers equation test from section 4. The results of these experiments in Figure 9 confirm the predicted stability conditions (3.33), (6.9), and (6.11) but less accurately than for Runge–Kutta schemes (3.25), (5.11), and (5.13).

7. Extension to some nonlinear equations. We show that these results extend to regular solutions to nonlinear problems such as the incompressible Euler equations on a domain Ω bounded with walls and scalar conservation laws. We proceed in three steps with gradually increasing complexity:

- First we consider the transport equation with nonconstant velocity on bounded domains. Hence we step outside the strict frame of von Neumann stability analysis.

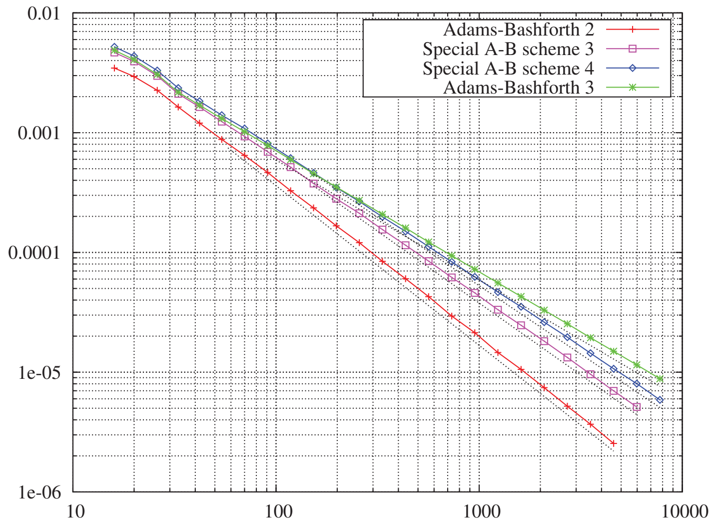

FIG. 9. *Maximal time step ensuring the stability, obtained experimentally with the one-dimensional Burgers equation* (7.15) *for Adams–Bashforth schemes. It evidences three slopes,* $\delta t_{\max} = C\delta x^{2r/(2r-1)}$ *with* $r = 2, 3,$ *and* 4, *plus a CFL condition slope, plotted in parallel with dotted lines.* (Ox) *axis represents the number of points* $N = \frac{2}{\delta x}$ *and* (Oy) *the maximal time step* $\delta t_{\max}$ *above which the numerical solution becomes unstable.*

- Then we study the simplest nonlinear equation involving transport, the one-dimensional Burgers equation, and we show that the previous results still hold true under a smoothness condition.
- Then we transpose our results to the scalar conservation laws and the incompressible Euler equations on a domain Ω possibly bounded by walls.

7.1. Transport by a variable velocity. The transport of a scalar θ by a divergence-free velocity $\mathbf{u}$ on an open set $\Omega \subset \mathbb{R}^d$ with regular boundaries satisfies the equation

$$\partial_t \theta + \mathbf{u}(\mathbf{x}) \cdot \nabla\theta = 0 \text{ for } \mathbf{x} \in \Omega,\ t \in [0, T] \tag{7.1}$$
$$\text{with}\quad \operatorname{div}(\mathbf{u}) = 0 \text{ for } \mathbf{x} \in \Omega, \quad \mathbf{u}(\mathbf{x}) \cdot \mathbf{n} = 0 \quad \text{for}\ \ \mathbf{x} \in \partial\Omega.$$

In order to generalize the stability analysis to this case, we need the following lemma, which corresponds to $\mathbf{v} = (\theta, \dots, \theta)$ and $\mathbf{w} = (\varphi, \dots, \varphi)$ in Lemma 7.2.

LEMMA 7.1. *Let* $\theta, \varphi : \Omega \to \mathbb{R}$, $\mathbf{u} : \Omega \to \mathbb{R}^d$ *such that* $\operatorname{div}(\mathbf{u}) = 0$ *on* Ω *and* $\mathbf{u} \cdot \mathbf{n} = 0$ *on* $\partial\Omega$*; then*

$$\langle \theta, \mathbf{u} \cdot \nabla\varphi \rangle_{L^2(\Omega)} = -\langle \mathbf{u} \cdot \nabla\theta, \varphi \rangle_{L^2(\Omega)}. \tag{7.2}$$

Equivalently, we have

$$\langle \theta, \mathbf{u} \cdot \nabla\theta \rangle_{L^2(\Omega)} = 0. \tag{7.3}$$

The computations using the skew-symmetry relationship leads to the same stability conditions as those relying on complex numbers in section 3 under the following assumptions regarding the space discretization.

Assumption 7.1. The space discretization conserves the skew-symmetry of the equation; i.e., with the notation of (2.11)

$$\langle M_{\delta x}\overline{\theta}, \overline{\theta} \rangle = 0, \qquad \text{which is equivalent to} \quad \langle M_{\delta x}\overline{\theta}, \overline{\varphi} \rangle = -\langle \overline{\theta}, M_{\delta x}\overline{\varphi} \rangle \quad \forall \overline{\theta}, \overline{\varphi} \in V_{\delta x}. \tag{7.4}$$

Such *conservative* discretizations are presented in some computational fluid mechanics publications such as [22], for instance.

Assumption 7.2. The discretization is sufficiently regular to enforce

$$\|M_{\delta x}\overline{\theta}\| \leq C\frac{\|\overline{\theta}\|}{\delta x} \quad \forall \overline{\theta} \in V_{\delta x}. \tag{7.5}$$

This assumption is satisfied with $C \sim \|\mathbf{u}\|_{L^\infty}$ for almost all the discretizations. One would need special properties to avoid this happening.

Let $F(\theta) = \mathbb{P}_{\delta x}(\mathbf{u} \cdot \nabla\theta)$ with $\mathbb{P}_{\delta x}$ the orthogonal projector onto the space of discretization $V_{\delta x}$. In sections 7.2, 7.3, and 7.4, even if the operator F is not linear, we omit the projector $\mathbb{P}_{\delta x}$ since it does not change the computations we present because it can be set or removed when needed:

$$\langle \mathbb{P}_{\delta x}\theta, \mathbb{P}_{\delta x}\varphi\rangle = \langle \mathbb{P}_{\delta x}\theta, \varphi\rangle = \langle \theta, \mathbb{P}_{\delta x}\varphi\rangle \quad \text{for } \theta, \varphi \in L^2(\Omega). \tag{7.6}$$

For the Runge–Kutta scheme (2.2), we find the following expression for θ_{n+1}:

$$\theta_{n+1} = \sum_{i=0}^{k} \beta_i \delta t^i F^i(\theta_n). \tag{7.7}$$

Starting from this expression and according to Lemma 7.1 along with Assumption 7.1,

$$\langle F^i(\theta_n), F^j(\theta_n)\rangle_{L^2(\Omega)} = \begin{cases} 0 & \text{if} \quad i+j = 2\ell+1 \quad \text{for} \quad \ell \in \mathbb{N}, \\ (-1)^{\ell-i}\|F^\ell(\theta_n)\|_{L^2}^2 & \text{if} \quad i+j = 2\ell \quad \text{for} \quad \ell \in \mathbb{N}. \end{cases} \tag{7.8}$$

We compute the L^2 norm of θ_{n+1} as a function of the L^2 norm of θ_n. From (7.7), (7.8) and under the Assumption 7.1, we have

$$\|\theta_{n+1}\|_{L^2}^2 = \sum_{\ell=0}^{k} S_\ell\, \delta t^{2\ell}\|F^\ell(\theta_n)\|_{L^2}^2 \tag{7.9}$$

with (S_ℓ) given by (3.5), i.e.,

$$S_\ell = \sum_{j=-\min(\ell,k-\ell)}^{\min(\ell,k-\ell)} (-1)^j \beta_{\ell-j}\beta_{\ell+j}. \tag{7.10}$$

For consistency in the numerical scheme, we must have $S_0 = 1$. On the other hand, let us suppose that $S_1 = S_2 = \cdots = S_{r-1} = 0$ and $S_r > 0$. Under Assumption 7.2, for $\theta_n \in V_{\delta x}$,

$$\|F^r(\theta_n)\|_{L^2} \leq \|\mathbf{u}\|_{L^\infty}^r \frac{\|\theta_n\|_{L^2}}{\delta x^r}, \tag{7.11}$$

and knowing that for $x \geq -1$,

$$\sqrt{1+x} \leq 1 + \frac{x}{2}, \tag{7.12}$$

we derive from (7.9)

$$\begin{aligned} \|\theta_{n+1}\|_{L^2} &\leq \left(1 + \frac{\delta t^{2r}}{\delta x^{2r}} S_r \|\mathbf{u}\|_{L^\infty}^{2r} + o(\delta t)\right)^{1/2} \|\theta_n\|_{L^2}, \\ \|\theta_n\|_{L^2} &\leq \left(1 + \left(\frac{\delta t^{2r-1} S_r}{2\delta x^{2r}}\|\mathbf{u}\|_{L^\infty}^{2r} + o(1)\right)\delta t\right) \end{aligned} \tag{7.13}$$

where $o()$ gathers all the negligible terms.

Let $a = \|\mathbf{u}\|_{L^\infty}$; then the numerical scheme (2.2) is stable for small perturbations under the condition

$$\delta t \leq C \left(\frac{\delta x}{a} \right)^{\frac{2r}{2r-1}}. \tag{7.14}$$

Hence the results obtained in the von Neumann stability framework remain valid in the case of the convection by a variable velocity on a bounded domain. This is still a linear equation but outside the von Neumann stability analysis framework which assumes a periodic or unbounded domain.

7.2. The Burgers equation. In order to clarify the role of the nonlinearity and validate our analysis under smoothness conditions on the solution, we have a look at the simplest nonlinear case, the one-dimensional inviscid Burgers equation:

$$\partial_t u + u \partial_x u = 0 \quad \text{for} \quad (t, x) \in [0, T] \times \mathbb{R}, \quad u(0, \cdot) = u_0. \tag{7.15}$$

In order to infer the numerical stability for this problem, we linearize it. Assume u_n is a discretized version of the solution u in time and in space. As proposed in [10], we consider a perturbed solution $u_n + \varepsilon_n$. Under regularity assumptions on u, each time discretization will involve a specific evolution equation on ε_n.

Actually, the small error ε_n that we introduce corresponds to oscillations at the smallest scales in space $V_{\delta x}$. This instability propagates and may increase at each time step. In the following, we demonstrate that under CFL-like conditions similar to those of section 3, the L^2 norm of the small error ε_n is amplified in a limited way,

$$\|\varepsilon_{n+1}\|_{L^2} \leq (1 + C\delta t)\|\varepsilon_n\|_{L^2}, \tag{7.16}$$

where C is a constant that neither depends on δx nor on δt. Thus, after an elapsed time T, the error increases at most exponentially as a function of the time:

$$\|\varepsilon_{t_0+T}\|_{L^2} \leq (1 + C\delta t)^{T/\delta t} \|\varepsilon_{t_0}\|_{L^2} \leq e^{CT} \|\varepsilon_{t_0}\|_{L^2}. \tag{7.17}$$

As $\partial_t u = -u \partial_x u$, $\partial_t^\ell u = \sum_\alpha \lambda_\alpha u^{\alpha_1} (\partial_x u)^{\alpha_2} \dots (\partial_x^{\ell-1} u)^{\alpha_{\ell-1}} + (-1)^\ell u^\ell \partial_x^\ell u$, we remark a kind of equivalence between the space regularity and the time regularity. If $\partial_x^\ell u \in L^\infty$, then $\partial_t^\ell u \in L^\infty$. In the general case, for Runge–Kutta schemes (2.2), we have for $0 \leq \ell \leq s$,

$$u_{(\ell)} + \varepsilon_{(\ell)} = \sum_{i=0}^{\ell-1} a_{\ell i} \, (u_{(i)} + \varepsilon_{(i)}) + \delta t \sum_{i=0}^{\ell-1} b_{\ell i} \, F(u_{(i)} + \varepsilon_{(i)}) \tag{7.18}$$

and $u_{n+1} + \varepsilon_{n+1} = u_{(s)} + \varepsilon_{(s)}$, so

$$\varepsilon_{(\ell)} = \sum_{i=0}^{\ell-1} a_{\ell i} \, \varepsilon_{(i)} + \delta t \sum_{i=0}^{\ell-1} b_{\ell i} \, (F(u_{(i)} + \varepsilon_{(i)}) - F(u_{(i)})) \tag{7.19}$$

and $\varepsilon_{n+1} = \varepsilon_{(s)}$.

PROPOSITION 7.1. *Consider a solution u of the Burgers equation* (7.15) *s-times differentiable such that $\|\partial_x^s u\|_{L^\infty(\mathbb{R} \times [0,T])} < +\infty$. Under the condition $\delta t = o(\delta x)$, a stability error ε_{n+1} in the explicit scheme* (7.19), *small enough at the initial time $\|\varepsilon_0\|_{L^2} = o(\delta x^{3/2})$, can be expressed as*

$$\varepsilon_{n+1} = \varepsilon_n + \sum_{i=1}^{s} \beta_i \, \delta t^i \, u_n^i \, \partial_x^i \varepsilon_n + \delta t \, \varepsilon_n \, \partial_x u_n + R_n \tag{7.20}$$

with $\|R_n\|_{L^2} = o(\delta t \|\varepsilon_n\|_{L^2})$. *The coefficients* (β_i) *derive from scheme* (2.2) *similarly to those in* (3.3).

Proof. All the terms we have to deal with are projections in the space discretization $V_{\delta x}$. In order to simplify the notation, we omit this projection that we assume orthogonal, as for the Galerkin methods [24, 1].

We prove that $\varepsilon_{(\ell)}$ can be put under the form (7.20) by recurrence on $\ell = 0 \dots s$.

As $\varepsilon_{(0)} = \varepsilon_n$, the assertion is true for $\ell = 0$.

Let us assume the assertion true for i from 0 to $\ell - 1$,

$$\varepsilon_{(i)} = \varepsilon_n + \sum_{j=1}^{i} \beta_{(i)j}\, \delta t^j\, u_n^j\, \partial_x^j \varepsilon_n + \alpha_{(i)}\, \delta t\, \varepsilon_n\, \partial_x u_n + R_{(i)}, \tag{7.21}$$

with $\|R_{(i)}\|_{L^2} = o(\delta t \|\varepsilon_n\|_{L^2})$. The coefficients $(\beta_{(i)j})$ correspond to the partial step i of the Runge–Kutta scheme distant by $\alpha_{(i)}\delta t$ from the time $n\delta t$. Note that $\alpha_{(s)} = 1$. Then, given that $\sum_{i=0}^{\ell-1} a_{\ell i} = 1$,

$$\begin{aligned}
\varepsilon_{(\ell)} &= \sum_{i=0}^{\ell-1} a_{\ell i}\, \varepsilon_{(i)} + \delta t \sum_{i=0}^{\ell-1} b_{\ell i}\, (F(u_{(i)} + \varepsilon_{(i)}) - F(u_{(i)})) \\
&= \varepsilon_n + \sum_{i=0}^{\ell-1} a_{\ell i} \left(\sum_{j=1}^{i} \beta_{(i)j} \delta t^j u_n^j \partial_x^j \varepsilon_n + \alpha_{(i)} \delta t \varepsilon_n \partial_x u_n + R_{(i)} \right) \\
&\quad + \delta t \sum_{i=0}^{\ell-1} b_{\ell i}\, \left(u_{(i)} \partial_x \varepsilon_{(i)} + \varepsilon_{(i)} \partial_x u_{(i)} + \varepsilon_{(i)} \partial_x \varepsilon_{(i)} \right).
\end{aligned} \tag{7.22}$$

Knowing that

$$\begin{aligned}
\partial_x \varepsilon_{(i)} &= \partial_x \varepsilon_n + \sum_{j=1}^{i} \beta_{(i)j} \delta t^j \left(\partial_x (u_n^j) \partial_x^j \varepsilon_n + u_n^j \partial_x^{j+1} \varepsilon_n \right) \\
&\quad + \alpha_{(i)} \delta t \left(\partial_x \varepsilon_n \partial_x u_n + \varepsilon_n \partial_x^2 u_n \right) + \partial_x R_{(i)},
\end{aligned} \tag{7.23}$$

it follows that

$$\begin{aligned}
R_{(\ell)} = \sum_{i=0}^{\ell-1} a_{\ell i}\, R_{(i)} + \delta t \sum_{i=0}^{\ell-1} b_{\ell i} \Bigg(u_{(i)} \Bigg(& \sum_{j=1}^{i} \beta_{(i)j} \delta t^j \partial_x (u_n^j) \partial_x^j \varepsilon_n \\
& + \delta t \left(\partial_x \varepsilon_n \partial_x u_n + \varepsilon_n \partial_x^2 u_n \right) + \partial_x R_{(\ell)} \Bigg) \\
& + \varepsilon_{(i)} \partial_x \varepsilon_{(i)} + (\varepsilon_{(i)} \partial_x u_{(i)} - \varepsilon_n \partial_x u_n) \\
& + (u_{(i)} - u_n) \sum_{j=1}^{i} \beta_{(i)j} \delta t^j u_n^j \partial_x^{j+1} \varepsilon_n \Bigg).
\end{aligned} \tag{7.24}$$

Now, we need to show that these terms are $o(\delta t \|\varepsilon_n\|_{L^2})$. According to assumption (7.21) and due to $\delta t = o(\delta x)$, Assumption 7.2 provides

$$\|\delta t^j \partial_x^j \varepsilon_n\| \le \left(\frac{\delta t}{\delta x} \right)^j \|\varepsilon_n\|. \tag{7.25}$$

Hence $\varepsilon_{(i)} = (1 + o(1))\varepsilon_n$ in the sense that $\varepsilon_{(i)} = \varepsilon_n + \eta_{(i)}$ with $\|\eta_{(i)}\|_{L^2} = o(\|\varepsilon_n\|_{L^2})$ (the stability condition being $\|\varepsilon_{(i)}\|_{L^2} = (1+O(\delta t))\|\varepsilon_n\|_{L^2}$). As we assumed $\|\varepsilon_n\|_{L^2} = o(\delta x^{3/2})$, then

$$\|\varepsilon_n\|_{L^\infty} \leq \frac{\|\varepsilon_n\|_{L^2}}{\delta x^{1/2}} = o(\delta x). \tag{7.26}$$

As a result, the cross term $\delta t \varepsilon_{(i)} \partial_x \varepsilon_{(i)}$ satisfies

$$\|\delta t \varepsilon_{(i)} \partial_x \varepsilon_{(i)}\|_{L^2} \leq \frac{\delta t}{\delta x} \|\varepsilon_{(i)}\|_{L^\infty} \|\varepsilon_{(i)}\|_{L^2} = o(\delta t \|\varepsilon_n\|_{L^2}). \tag{7.27}$$

As for $i \leq s-1$,

$$u_{(i)} = u_n + \delta t B_{(i)}(u_n, \partial_x u_n, \ldots, \partial_x^i u_n, \delta t) \tag{7.28}$$

with B a polynomial, $\|B\|_{L^\infty}$ and $\|\partial_x B\|_{L^\infty}$ are bounded, so $\|u_{(i)} - u_n\|_{L^\infty}$=o(1) and $\|\partial_x u_{(i)} - \partial_x u_n\|_{L^\infty}$=o(1). It allows us to replace $u_{(i)}$ by u_n in the expansion (7.22), the difference going into $R_{(\ell)}$; see (7.24).

Hence, using the fact that $\varepsilon_{(i)}, R_{(i)} \in V_{\delta x}$ the discretization space, $\|\partial_x^j \varepsilon_{(i)}\|_{L^2} \leq \frac{\|\varepsilon_{(i)}\|_{L^2}}{\delta x^j}$ and the same for $R_{(i)}$. Let r be an element of the sum $R_{(\ell)}$; then it satisfies

$$\|r\|_{L^2} \leq \frac{\delta t^p}{\delta x^q} \tau(\|u_n\|_{L^\infty}, \|\partial_x u_n\|_{L^\infty}, \ldots, \|\partial_x^\ell u_n\|_{L^\infty}) \|\varepsilon_n\|_{L^2} \tag{7.29}$$

with τ a polynomial and $p \geq q+1$.

Given the fact that $\delta t = o(\delta x)$, we obtain that $\|R_{(\ell)}\|_{L^2} = o(\delta t \|\varepsilon_n\|_{L^2})$. Using the recurrence, we obtain the result for $\ell = s$, i.e., for ε_{n+1}.

Actually, taking into account the orthogonality of $\varepsilon_n \partial_x \varepsilon_n$ with ε_n, we can relax one of the assumptions; i.e., it is sufficient to have $\|\varepsilon_0\|_{L^2} = o(\delta x)$, and with the cancellations, it is even only necessary that $\|\varepsilon_0\|_{L^2} = o(\delta x^{1/2})$. ∎

THEOREM 7.1. *If we solve the Burgers equation* (7.15) *with the numerical scheme* (2.2), *then for a sufficiently regular solution u, the stability condition is provided by*

$$\delta t \leq \left(\frac{2(C - \|\partial_x u\|_{L^\infty})}{S_r} \right)^{\frac{1}{2r-1}} \left(\frac{\delta x}{\|u\|_{L^\infty}} \right)^{\frac{2r}{2r-1}}, \tag{7.30}$$

where δt is the time step, δx the space step, r an integer, and S_r a quantity defined by (3.3), (3.4), *and* (3.5) *and C the constant in the exponential growth of the error $\varepsilon(t) \sim \varepsilon_0 e^{Ct}$.*

Proof. Thanks to Proposition 7.1, we are able to write

$$\|\varepsilon_{n+1}\|_{L^2} \leq \|\varepsilon_n + \sum_{i=1}^{s} \beta_i \delta t^i u_n^i \partial_x^i \varepsilon_n\|_{L^2} + \delta t \|\varepsilon_n \partial_x u_n\|_{L^2} + o(\delta t \|\varepsilon_n\|_{L^2}). \tag{7.31}$$

On the other hand we have $\|\varepsilon_n \partial_x u_n\|_{L^2} \leq \|\partial_x u\|_{L^\infty} \|\varepsilon_n\|_{L^2}$, and since for $i, j \leq s$,

$$\begin{aligned} &\langle \delta t^i u_n^i \partial_x^i \varepsilon_n, \delta t^j u_n^j \partial_x^j \varepsilon_n \rangle_{L^2} \\ &\quad = \begin{cases} o(\delta t \|\varepsilon_n\|_{L^2}^2) & \text{if} \quad i+j = 2\ell+1, \\ (-1)^{\ell-i} \delta t^{2\ell} \|u_n^\ell \partial_x^\ell \varepsilon_n\|_{L^2}^2 + o(\delta t \|\varepsilon_n\|_{L^2}^2) & \text{if} \quad i+j = 2\ell, \end{cases} \end{aligned} \tag{7.32}$$

we derive

$$\left\| \varepsilon_n + \sum_{i=1}^{s} \beta_i \delta t^i u_n^i \partial_x^i \varepsilon_n \right\|_{L^2}^2 = \sum_{\ell=0}^{2s} S_\ell \delta t^\ell \| u_n^\ell \partial_x^\ell \varepsilon_n \|_{L^2}^2 + o(\delta t \|\varepsilon_n\|_{L^2}^2) \tag{7.33}$$

with S_ℓ given by (3.5).

Then, as $S_0 = 1$ and $\|u_n^\ell \partial_x^\ell \varepsilon_n\|_{L^2} \le \|u_n\|_{L^\infty}^\ell \frac{\|\varepsilon_n\|_{L^2}}{\delta x^\ell}$,

$$\left\| \varepsilon_n + \sum_{i=1}^{s} \beta_i \delta t^i u_n^i \partial_x^i \varepsilon_n \right\|_{L^2}^2 \le \sum_{\ell=0}^{s} S_\ell \left(\frac{\delta t}{\delta x}\right)^{2\ell} \|u\|_{L^\infty}^{2\ell} \|\varepsilon_n\|_{L^2}^2 + o(\delta t \|\varepsilon_n\|_{L^2}^2) \tag{7.34}$$

so

$$\left\| \varepsilon_n + \sum_{i=1}^{s} \beta_i \delta t^i u_n^i \partial_x^i \varepsilon_n \right\|_{L^2} \le \left(1 + \frac{1}{2} \sum_{\ell=1}^{s} S_\ell \left(\frac{\delta t}{\delta x}\right)^{2\ell} \|u\|_{L^\infty}^{2\ell} + o(\delta t)\right) \|\varepsilon_n\|_{L^2} \tag{7.35}$$

and finally

$$\|\varepsilon_{n+1}\|_{L^2} \le \left(1 + \frac{1}{2} \sum_{\ell=1}^{s} S_\ell \left(\frac{\delta t}{\delta x}\right)^{2\ell} \|u\|_{L^\infty}^{2\ell} + \delta t \|\partial_x u\|_{L^\infty} + o(\delta t)\right) \|\varepsilon_n\|_{L^2}. \tag{7.36}$$

Let r be the first power in the sum where $S_r \neq 0$, and let us assume that $S_r > 0$. Then, the stability condition $\|\varepsilon_{n+1}\|_{L^2} \le (1 + C\delta t) \|\varepsilon_n\|_{L^2}$ is reduced to

$$\frac{1}{2} S_r \frac{\delta t^{2r-1}}{\delta x^{2r}} \|u\|_{L^\infty}^{2r} \le (C - \|\partial_x u\|_{L^\infty}), \tag{7.37}$$

i.e., the condition (7.30). We recognize the same power law as the one obtained in the linear case (3.9). The term $-\|\partial_x u\|_{L^\infty}$ should usually be discarded since its contribution is external to the instability phenomenon and random. ∎

7.3. Scalar conservation laws. Scalar conservation laws group equations of the type

$$\partial_t u + \sum_{i=1}^{d} \partial_{x_i} f_i(u) = 0 \quad \text{for} \quad (\mathbf{x}, t) \in \mathbb{R}^d \times [0, T], \tag{7.38}$$

$$u(0, \mathbf{x}) = u_0(\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^d \tag{7.39}$$

with $f_i : \mathbb{R} \to \mathbb{R}$ differentiable functions and $u : \mathbb{R}^d \to \mathbb{R}$, the scalar unknown function. A stability analysis of the solution of these equations in the frame of a Runge–Kutta discontinuous Galerkin formulation was presented in [24] for space accuracy of orders two and three, with the $\delta t \le C \delta x^{4/3}$ CFL-like condition, but as the byproduct of a long and rigorous computational process. This work was the continuation of [4], where the authors observed that first and second order Runge–Kutta methods are unstable under any linear CFL conditions when the space discretization is sufficiently accurate and so does not dissipate too much. In this section, we link their results to our analysis and refine the stability criteria. We have the following result.

THEOREM 7.2. *Let us apply the numerical scheme* (2.2) *to solve* (7.38). *If* $f \in C^{p+1}$ *and* $u \in C^p$, *i.e.*, $f^{(p+1)}, u^{(p)} \in L^\infty$, *and if the* (S_ℓ) *defined by* (7.10) *satisfy* $S_1 = \cdots = S_{r-1} = 0$ *and* $S_r > 0$, *then, given a constant* C *limiting the exponential* *growth of the stability error $\varepsilon_T \leq e^{CT}\varepsilon_0$, the numerical scheme is conditionally stable under the CFL-like condition*

$$\delta t \leq \left(\frac{2C}{S_r}\right)^{1/(2r-1)} \left(\frac{\delta x}{\sum_{i=1}^d \|f_i'(u)\|_{L^\infty}}\right)^{\frac{2r}{2r-1}}. \tag{7.40}$$

Proof. The proof is more or less the same as for the Burgers case (cf. part 7.2) using the following facts:

- $f_i(u_n + \varepsilon_n) = f_i(u_n) + f_i'(u_n)\varepsilon_n + o(\varepsilon_n)$,
- $u_{(\ell)} - u_n = o(1)$,
- $\partial_{x_i}(f_i'(u_n)\varepsilon) \sim f_i'(u_n)\partial_{x_i}\varepsilon$ for stability analysis, and
- for all functions η and ε,

$$\left\langle \eta, \sum_{i=1}^d \sum_{j=1}^d \partial_{x_i}\left(f_i'(u_n) f_j'(u_n)\partial_{x_j}\varepsilon\right)\right\rangle = -\left\langle \sum_{i=1}^d f_i'(u_n)\partial_{x_i}\eta, \sum_{i=1}^d f_i'(u_n)\partial_{x_i}\varepsilon\right\rangle \tag{7.41}$$

  allowing equalities of the type (7.50).

Finally, we obtain

$$\|\varepsilon_{n+1}\|_{L^2}^2 = (1+2C_1\delta t + o(\delta t))\|\varepsilon_n\|_{L^2}^2 + S_r\,\delta t^{2r} \left\| \sum_{\mathbf{i}\in[1,d]^r} \left(\prod_{s=1}^r f_{i_s}'(u_n)\right)\left(\prod_{s=1}^r \partial_{x_{i_s}}\right)\varepsilon_n \right\|_{L^2}^2. \tag{7.42}$$

Then, knowing that for $\varepsilon_n \in V_{\delta x}$,

$$\begin{aligned} \left\| \sum_{\mathbf{i}\in[1,d]^r} \left(\prod_{s=1}^r f_{i_s}'(u_n)\right)\left(\prod_{s=1}^r \partial_{x_{i_s}}\right)\varepsilon_n \right\|_{L^2}^2 &\leq \left(\sum_{\mathbf{i}\in[1,d]^r} \left(\prod_{s=1}^r \|f_{i_s}'(u_n)\|_{L^\infty}\right) \frac{\|\varepsilon_n\|_{L^2}}{\delta x^r}\right)^2 \\ &\leq \left(\left(\sum_{i\in[1,d]} \|f_i'(u_n)\|_{L^\infty}\right)^r \frac{\|\varepsilon_n\|_{L^2}}{\delta x^r}\right)^2, \end{aligned} \tag{7.43}$$

and neglecting the constant C_1, the von Neumann stability criteria

$$\|\varepsilon_{n+1}\|_{L^2}^2 \leq (1 + 2C\delta t + o(\delta t))\|\varepsilon_n\|_{L^2}^2 \tag{7.44}$$

is satisfied if

$$\left(\sum_{i\in[1,d]} \|f_i'(u)\|_{L^\infty}\right)^{2r} \frac{S_r\,\delta t^{2r}}{\delta x^{2r}} \leq 2C\delta t, \tag{7.45}$$

i.e., condition (7.40). ◻

7.4. Incompressible Euler equation. The Euler equations model incompressible fluid flows with no viscous term:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} - \nabla p = 0, \quad \operatorname{div}\mathbf{u} = 0, \quad \text{for } (t,\mathbf{x}) \in \mathbb{R}_+ \times \Omega. \tag{7.46}$$

The use of the Leray projector $\mathbb{P}$, which is the L^2-orthogonal projector on the divergence-free space, allows us to remove the pressure term:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbb{P}\left[(\mathbf{u}\cdot\nabla)\mathbf{u}\right] = 0. \tag{7.47}$$

The stability analysis of this case proceeds somehow as a synthesis of sections 7.1 and 7.2. An important property is then the skewness property of the transport term. (See [9, chapter IV, Lemma 2.1] or [8] for the proof, also used for the stability of the incompressible Navier–Stokes equations in [20, 17].)

Lemma 7.2. *Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in H^1(\Omega)^d$, $H^1(\Omega)$ denoting the Sobolev space on the open set $\Omega \subset \mathbb{R}^d$ be such that $(\mathbf{u}\cdot\nabla)\mathbf{v}, (\mathbf{u}\cdot\nabla)\mathbf{w} \in L^2$. If $\mathbf{u} \in \boldsymbol{H}_{\mathrm{div},0}(\Omega) = \{\mathbf{f} \in (L^2(\Omega))^d,\ \mathrm{div}\,\mathbf{f} = 0 \text{ on } \Omega,\ \mathbf{f}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega\}$, then*

$$\langle \mathbf{v}, (\mathbf{u}\cdot\nabla)\mathbf{w}\rangle_{L^2(\Omega)} = -\langle (\mathbf{u}\cdot\nabla)\mathbf{v}, \mathbf{w}\rangle_{L^2(\Omega)}. \tag{7.48}$$

Corollary 7.1. *With the same assumptions as in Lemma* 7.2,

$$\langle \mathbf{v}, (\mathbf{u}\cdot\nabla)\mathbf{v}\rangle_{L^2(\Omega)} = \int_{\mathbf{x}\in\Omega} \mathbf{v}\cdot(\mathbf{u}\cdot\nabla)\mathbf{v}\, d\mathbf{x} = 0. \tag{7.49}$$

Considering the scheme (2.2), we introduce a stability error $\varepsilon_{(\ell)}$ at level ℓ. Then, under the condition $\delta t = o(\delta x)$ and for ε_n small enough, most of the terms appearing in the expression of $\varepsilon_{(\ell)}$ are negligible with respect to

- the terms $\delta t^i\, F_n^i(\varepsilon_n)$, where $F_n(\varepsilon_n) = \mathbb{P}[(\mathbf{u}_n\cdot\nabla)\varepsilon_n]$ and $F_n^i = \underbrace{F_n \circ F_n \circ \cdots \circ F_n}_{i \text{ times}}$,
- the term $\delta t\, \mathbb{P}[(\varepsilon_n\cdot\nabla)\mathbf{u}_n]$.

Then most of the arguments used in section 7.2 apply with even more accuracy since we have the orthogonality relation

$$\langle F_n^i(\varepsilon_n), F_n^j(\varepsilon_n)\rangle_{L^2(\Omega)} = \begin{cases} 0 & \text{if} \quad i+j = 2\ell+1 \quad \text{for} \quad \ell \in \mathbb{N}, \\ (-1)^{\ell-i}\|F_n^\ell(\varepsilon_n)\|^2_{L^2(\Omega)} & \text{if} \quad i+j = 2\ell \quad \text{for} \quad \ell \in \mathbb{N} \end{cases} \tag{7.50}$$

instead of (7.32). This leads to the following result.

Proposition 7.2. *Assume that the incompressible Euler equations* (7.46) *have an s-times space-differentiable solution $\mathbf{u}$ such that $\|\nabla^s\mathbf{u}\|_{L^\infty([0,T]\times\Omega)} < +\infty$, that the discretization conserves the skew-symmetry relation* (7.48)*, and that $\forall\varepsilon \in V_{\delta x}$, $\|\mathbb{P}_{\delta x}\, F_n^k(\varepsilon)\|_{L^2} \le C\|\mathbf{u}\|^k_{L^\infty}\frac{\|\varepsilon\|_{L^2}}{\delta x^k}$. Then a stability error ε small enough at the initial time $\|\varepsilon_0\|_{L^2} = o(\delta x^{d/2})$ remains bounded for $t \in [0,T]$ under the condition*

$$\delta t \le \left(\frac{2C}{S_r}\right)^{1/(2r-1)} \left(\frac{\delta x}{\|\mathbf{u}\|_{L^\infty}}\right)^{\frac{2r}{2r-1}} \tag{7.51}$$

with δt the time step, δx the space step, and r and S_r obtained as in (7.10).

This proposition extends to Navier–Stokes equations for a high Reynolds number. The incompressible Navier–Stokes equations are written

$$\begin{cases} \partial_t\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} - \nu\Delta\mathbf{u} + \nabla p = 0, \\ \mathrm{div}\mathbf{u} = 0, \\ \mathbf{u}(0,x) = \mathbf{u}_0(x). \end{cases} \quad x \in \mathbb{R}^d,\ t \in [0,T], \tag{7.52}$$

Using the Leray projector $\mathbb{P}$ (the orthogonal projector on divergence-free vector fields) we reduce the equation to

$$\partial_t\mathbf{u} + \mathbb{P}\left[\mathbf{u}\cdot\nabla\mathbf{u}\right] - \nu\Delta\mathbf{u} = 0. \tag{7.53}$$

Two second order schemes are widely in use for the solution of this equation: the order two Runge–Kutta scheme [15, 8] and the second order Adams–Bashforth scheme [18, 19].

When the Reynolds number $\mathrm{Re} = \frac{\|u\|_{L^\infty} L}{\nu}$ is sufficiently large, the contribution of the heat kernel to the stability vanishes [8], and the same instability effects as for the incompressible Euler equation appear as observed in two-dimensional experiments [8]. New tests with boundaries comply with the stability condition $\delta t \leq \delta t_{\max} = C \delta x^{4/3}$ for dipole/wall numerical experiments. These results will be presented in a forthcoming paper.

8. Multicomponent transport. We extend the scope of application of the stability conditions (3.9) to other cases with multiple derivatives in time, like wave equations or multiple components, like in some MHD models [7]. Let us consider the one-dimensional equation

$$\partial_t X = M \partial_x X \quad \text{with} \quad X = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad u_\ell : \mathbb{R} \to \mathbb{R}, \quad \text{and} \quad M \in \mathcal{M}_n(\mathbb{R}). \tag{8.1}$$

For example, for $X = (u, v)^t$ and $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ we obtain the wave equation $\partial_t^2 u = \partial_x^2 u$.

Regarding the general case, we diagonalize the matrix M in $\mathbb{C}$:

$$M = P^{-1} D P, \quad \text{with} \quad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}. \tag{8.2}$$

Considering $Y = PX$, the equation $\partial_t Y = D \partial_x Y$ has physical meaning only if $\lambda_\ell \in \mathbb{R} \forall \ell$. Under this form all the components are independent. Therefore all our results on the transport equation apply to this case, taking $a = \max_\ell |\lambda_\ell|$.

When the matrix M cannot be diagonalized, like in the case $M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, we remark that the second component is independent from the first component,

$$\begin{cases} \partial_t u_1 = \partial_x u_1 + \partial_x u_2, \\ \partial_t u_2 = \partial_x u_2, \end{cases} \tag{8.3}$$

then the term $\partial_x u_2$ in the first equation plays the role of a source term.

In the case when there are several space variables,

$$\partial_t X = M_1 \partial_{x_1} X + M_2 \partial_{x_2} X + \cdots + M_n \partial_{x_n} X \tag{8.4}$$

applying a von Neumann stability analysis, we obtain

$$\partial_t \hat{X} = (M_1 \, i\xi_1 + M_2 \, i\xi_2 + \cdots + M_n \, i\xi_n) \, \hat{X}. \tag{8.5}$$

We consider $M(\xi) = M_1 \, \xi_1 + M_2 \, \xi_2 + \cdots + M_n \, \xi_n$ and diagonalize $M(\xi) = P(\xi)^{-1} D(\xi) P(\xi)$. As previously, taking $\hat{Y}_\xi = P(\xi) \hat{X}$, we obtain the stability constraint (3.11) as in the scalar case.

9. Conclusion. The stability CFL-like conditions presented in this paper may be encountered in many simulations of convection-dominated problems using explicit numerical schemes. Although based on a classical von Neumann stability analysis, this kind of stability analysis is not performed elsewhere.

Two arguments support our approach. First, we explained some "CFL shrinking" effects for second order schemes already in use: people remarked that they had to take $C \to 0$ in the usual linear CFL condition $\delta t \leq C\delta x$. Second, we predicted some exotic CFL conditions $\delta t \leq C\delta x^{\frac{2r}{2r-1}}$ for certain Runge–Kutta schemes and Adams–Bashforth schemes which optimize the energy conservation. Numerical tests validate these predictions.

We showed why increasing the temporal order of a scheme increases the stability. We even linked the order of a scheme, its stability, and the tangency of its stability domain to the (Oy) axis in the von Neumann stability analysis. Nevertheless, the numerical viscosity may erase these instability effects especially when using an upwind scheme [4].

We extended the domain of application of these results to different equations, including equations on bounded domains, nonlinear equations, and equations with multiple derivatives in time. These extensions assume smoothness properties for the solution. This smoothness assumption restrains the frame of application to a rather limited area. Nevertheless, this clear exposing of actual numerical artifacts plus the correlate accurate stability conditions should be useful to a wide community, in particular to those who perform numerical simulations of turbulent flows with spectral codes.

Acknowledgments. The author gratefully acknowledges the CEMRACS 2007 organizers for his stay in the CIRM in Marseilles and for his access to its rich bibliographical resources, as well as Institute of Fundamental Technological Research Polish Academy of Sciences (IPPT PAN) and Commissariat à l'Énergie Atomique (CEA) for his stays there in 2007 and 2009, respectively. He wishes to express his gratitude to Yvon Maday and Frédéric Coquel for fruitful discussions, as well as to Dmitry Kolomenskiy for his help in the redaction of this paper and in the realization of the numerical experiments. He also acknowledges the anonymous referees, whose comments substantially improved the quality of the paper.

REFERENCES

[1] E. Burman, A. Ern, and M.A. Fernandez, *Explicit Runge–Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems*, SIAM J. Numer. Anal., 48 (2010), pp. 2019–2042.

[2] C. Canuto, M.T. Hussaini, A. Quarteroni, and T.A. Zang, *Spectral Methods in Fluid Dynamics*, Springer-Verlag, New York, 1988.

[3] J.G. Charney, R. Fjörtoft, and J. von Neumann, *Numerical Integration of the Barotropic Vorticity Equation*, Tellus 2, John Wiley, New York, 1950, pp. 237–254.

[4] B. Cockburn and C.-W. Shu, *Runge-Kutta discontinuous Galerkin methods for convection-dominated problems*, J. Sci. Comput., 16 (2001), pp. 173–261.

[5] R. Courant, K. Friedrichs, and H. Lewy, *On the partial difference equations of mathematical physics*, IBM Journal, 1967, pp. 215–234.

[6] M. Crouzeix and A.L. Mignot, *Analyse Numérique des Équations Différentielles*, Masson, Paris, 1992.

[7] O. Czarny and G. Huysmans, *Bézier surfaces and finite elements for MHD simulations*, J. Comput. Phys., 227 (2008), pp. 7423–7445.

[8] E. Deriaz and V. Perrier, *Direct numerical simulation of turbulence using divergence-free wavelets*, SIAM Multiscale Model. Simul., 7 (2008), pp. 1101–1129.

[9] V. Girault and P.A. Raviart, *Finite Element Methods for Navier-Stokes Equations*, Springer-Verlag, Berlin, 1986.

[10] E. Godlewski and P.A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Springer-Verlag, New York, 1996.

[11] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112.

[12] D. Gottlieb and E. Tadmor, *The CFL condition for spectral approximations to hyperbolic initial-boundary value problems*, Math. Comp., 56 (1991), pp. 565–588.

[13] E. Hairer, S.P. Nørsett, and G. Wanner, *Solving ordinary differential equations* I: *Nonstiff problems*, Comput. Math. 8, Springer-Verlag, New York, 1987.

[14] E. Hairer and G. Wanner, *Solving ordinary differential equations* II: *Stiff and differential-algebraic problems*, Comput. Math. 14, Springer-Verlag, New York, 1991.

[15] R. Kupferman and E. Tadmor, *A fast, high resolution, second-order central scheme for incompressible flows*, Proc. Natl. Acad. Sci. USA, 94 (1997), pp. 4848–4852.

[16] D. Levy and E. Tadmor, *From semidiscrete to fully discrete: Stability of Runge–Kutta schemes by the energy method*, SIAM Rev., 40 (1998), pp. 40–73.

[17] M. Marion and R. Temam, *Numerical Methods for Fluids*, Handb. Numer. Anal. 6, Elsevier Science, New York, 1998.

[18] R. Peyret, *Handbook of Computational Fluid Mechanics*, Academic Press, New York, 2000.

[19] K. Schneider, *Numerical simulation of the transient flow behaviour in chemical reactors using a penalization method*. Comput. & Fluids, 34 (2005), pp. 1223–1238.

[20] R. Temam, *The Navier-Stokes Equations*, North-Holland, Amsterdam, 1984.

[21] L.N. Trefethen, *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*, http://www.comlab.ox.ac.uk/nick.trefethen/pdetext.html (1996).

[22] R.W.C.P. Verstappen and A.E.P. Veldman, *Symmetry-preserving discretization of turbulent flow*, J. Comput. Phys., 187 (2003), pp. 343–368.

[23] P. Wesseling, *Principles of Computational Fluid Dynamics*, Springer-Verlag, New York, 2001.

[24] Q. Zhang and C.-W. Shu, *Error estimates to smooth solutions of Runge–Kutta discontinuous Galerkin methods for scalar conservation laws*, SIAM J. Numer. Anal., 42 (2004), pp. 641–666.