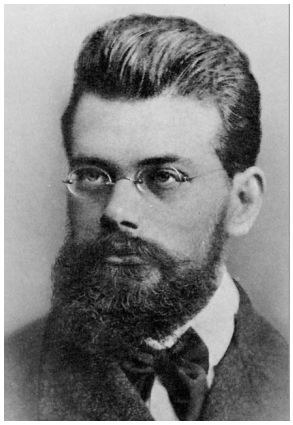
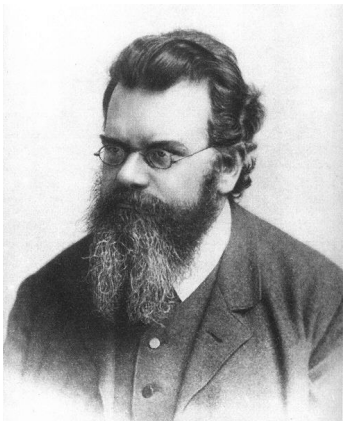
université
PARIS-SACLAY

Université Paris-Saclay

the Lattice Boltzmann Method

M2 – Analyse Modélisation Simulation

B. Graille
March 12, 2025

Pictures of Ludwig Eduard Boltzmann

Sei irgend ein Raum mit sehr vielen Gasmolekülen erfüllt, deren jedes ein einfacher materieller Punkt ist. Jede Molekül fliege während des grössten Theiles der Zeit geradlinig mit gleichförmiger Geschwindigkeit fort. Nur wenn sich zwei Moleküle zufällig sehr nahe kommen, beginnen sie auf einender einzuwirken. Ich nenne diesen Vorgang, während dessen zwei Moleküle auf einander einwirken, einen Zusammenstoss der beiden Moleküle, ohne dass jedoch dabei an einen Stoss elastischer Körper zu denken ist ; die während des Zusammenstosses wirksamen Kräfte können vielmehr ganz beliebig sein. Selbst wenn zu Anfang der Zeit alle Moleküle dieselbe Geschwindigkeit besessen hätten, würden sie dieselbe im Verlaufe der Zeit nicht immer beibehalten. In Folge der Zusammenstösse werden vielmehr einige Moleküle grössere, andere kleinere Geschwindigkeiten annehmen, bis sich endlich eine solche Verteilung der Geschwindigkeiten unter den Molekülen hergestellt hat, dass dieselbe durch die Zusammenstösse nicht weiter verändert wird. Bei dieser schliesslich sich herstellenden Geschwindigkeitsvertheilung werden im Allgemeinen alle möglichen Geschwindigkeit von Null bis zu einer sehr grossen Geschwindigkeit vorkommen. Die Zahl der Molekülen, deren Geschwindigkeit zwischen v und $v + dv$ liegt, wollen wir mit $F(v)dv$ bezeichnen. Dann bestimmt uns also die Function F die Geschwindigkeitsvertheilung vollständig. Für den Fall einatomiger Moleküle, den wir jetzt betrachten, fand bereits Maxwell für $F(v)$ den Werth $Av^2e^{-Bv^2}$, wobei A und B Constanten sind, so dass also die Wahrscheinlichkeit der verschiedenen Geschwindigkeit durch eine ähnliche Formel gegeben wird, wie die Wahrscheinlichkeit der verschiedenen Beobachtungsfehler in der Theorie der Methode der kleinsten Quadrate. Der erste Beweis jedoch, den Maxwell für diese Formel gab, wird von ihm selbst als unrichtig bezeichnet. Später gab er zwar einen sehr eleganten Beweis dafür, dass, wenn man die obige Geschwindigkeitsvertheilung einmal unter den Glasmolekülen hergestellt hat, dieselbe in der That durch die Zusammenstösse nicht weiter verändert wird. Er sucht auch zu beweisen, dass es die einzige Geschwindigkeitsvertheilung von der betrachteten Eigenschaft ist. Allein der letztere Beweis scheint mir wieder Fehlschlüsse zu enthalten. Es ist somit noch nicht bewiesen, dass, wie immer der Zustand des Gases zu Anfang gewesen sein mag, er sich immer dieser von Maxwell gefundenen Grenze nähern muss. Es könnte sein, dass es ausser dieser noch verschiedene andere mögliche Grenzen gibt. Dieser Beweis gelingt aber leicht mittelst der Auffassungsweise des Problems, zu deren Auseinandersetzung ich jetzt schreiten will, und welche zudem den Vorteil bitter, dass sie sich direct auf mehratomige Moleküle, also auf den in der Natur wahrscheinlich allein vorkommenden Fall übertragen lässt.

Ludwig Eduard Boltzmann, extract from [2]

Contents

1 A numerical method inspired by the kinetic theory **7**
1 Thermodynamics of gases . . . 7
2 Cellular automata . . . 9
3 The lattice Boltzmann method . . . 11

2 Example of the D1Q2 **15**
1 Description of the scheme . . . 15
2 Asymptotic Analysis : the Taylor expansion method . . . 16
3 Numerical illustrations . . . 19
3.1 The transport equation . . . 20
3.2 Burgers equation . . . 21
4 Notions of stability . . . 22
4.1 The von Neumann analysis . . . 22
4.2 The maximum principle . . . 24

3 A notion of consistency **29**
1 Algebraic form of LB schemes . . . 29
1.1 Spatial and temporal discretization . . . 29
1.2 Discrete velocities and shift operators . . . 30
1.3 Monolithic scheme . . . 32
2 Corresponding multi-step scheme . . . 33
2.1 Characteristic polynomial and Cayley-Hamilton theorem . . . 33
2.2 Main result . . . 34
3 Example of the D_1Q_2 . . . 34

A Fourier Transform **37**

B Schur and von Neumann polynomials **39**
1 Definitions . . . 39
2 Theorems . . . 40

References **41**

Chapter 1

A numerical method inspired by the kinetic theory

We first propose an elementary introduction to the lattice Boltzmann method. This numerical method has its origins in gas automata or cellular automata.

1 Thermodynamics of gases

The fluid flows can be described at several spatial and time scalings according to the investigated phenomena. When the number of particles in the volume of reference is low (and in particular when the number of collisions involving these particles is low), a microscopic description is required and yields to predict the evolution of each individual particule. Many models are relevant and the most precise use quantum mechanics. One of the simplest models considers the particles as hard spheres moving in a straight line and interacting by bouncing. These collisions are infinitely brief phenomena during which certain physical quantities are exchanged like momentum, energy, *etc*. Unfortunately, as the number of particles increases, these models become unusable due to too many variables: remember that Avogadro's number equals to the number of carbon atoms in 12 grams is worth approximately 6×10^{23}.

On the mesoscopic scale, particles are described by a continuous variable called the velocity distribution function which represents the number of particles at a given time, at a given point in space, with a given velocity. This scale is for example used to describe the statistical velocity distribution of a gas when it is far from thermodynamic equilibrium. This intermediate spatial scale is relevant in the case where the unit volume contains a large number of particles. The particle density equation proposed by Boltzmann [2] describes the evolution of a population of particles which relaxes towards an equilibrium given by a

Maxwell-Boltzmann distribution. This equilibrium distribution is characterized by only a few state variables (mass, momentum, energy, *etc.*) and depends very little on the underlying microscopic system: many different microscopic systems have the same equilibrium. Boltzmann further shows that a function (later called the Boltzmann's H function) plays the role of an entropy: it is necessarily decreasing over time for a closed system. He introduces a statistical interpretation of entropy in 1877, which marked a major breakthrough in the understanding of the transition from reversible microscopic dynamics to irreversible macroscopic evolution—the formula in engraved on his gravestone in Vienna $S = k \log W$.

Various attempts have been proposed to obtain approximate solutions of the Boltzmann equation: Chapman [4, 5, 6] and Enskog [13, 14] independently gave two methods with identical results, those of Enskog reusing Hilbert's asymptotic expansion method. Enskog's method consists of jointly expanding the Boltzmann equation and its solution: the approximation of order zero is a Maxwell-Boltzmann distribution and its moments satisfy the compressible Euler equations; the first order approximation is the solution of a linear equation and its moments satisfy the compressible Navier-Stokes equations. When the second order approximation is considered, the Burnett equations are obtained. However, these equations of order 2 have never made it possible to obtain notable improvements [15, 22], the essential reason being the real difficulty of having to impose boundary conditions for higher order derivatives. Full details of the Enskog method—generally referred to as the Chapman-Enskog method in the literature—are given in classic references such as Hirschfleder et al. [22], Ferziger and Kaper [15], Vincenti and Kruger [30], Chapman and Cowling [7] and Liboff [23]. These approximate solutions and the resulting models are relevant at a macroscopic scale for which the mean free path—average distance traveled by a particle between two collisions—is very small.

We consider a simple gas with a single type of particles and no external force lying in $\mathbb{R}^d$ for $d \in \{1, 2, 3\}$. Let's define the particle distribution function f, function of the time t, the position x, and the velocity c, such that the mass of the particles $\mathrm{d}\rho$ lying in a box located at x with a small volume $\mathrm{d}x$ and having a velocity c defined with a precision $\mathrm{d}c$, at the time t, is given by

$$\mathrm{d}\rho = f(t, x, c)\,\mathrm{d}x\,\mathrm{d}c.$$

The particle distribution function can then be used to define the macroscopic quantities as, for example, the mass ρ, the momentum q, and the energy E by

$$\rho(t, x) = \int_{\mathbb{R}^d} f(t, x, c)\,\mathrm{d}c, \tag{1.1}$$

$$q(t, x) = \int_{\mathbb{R}^d} cf(t, x, c)\,\mathrm{d}c, \tag{1.2}$$

$$E(t, x) = \int_{\mathbb{R}^d} \frac{1}{2} c{\cdot}cf(t, x, c)\,\mathrm{d}c. \tag{1.3}$$

The fluid velocity is defined as $u = q/\rho$.

The particle distribution function f satisfies the Boltzmann equation

$$\partial_t f(t,x,c) + c\cdot\partial_x f(t,x,c) = \mathscr{Q}(f)(t,x,c), \qquad t>0,\ x\in\mathbb{R}^d,\ c\in\mathbb{R}^d. \tag{1.4}$$

In this equation, the term $\partial_t f + c\cdot\partial_x f$ corresponds to a free transport at the velocity c, whereas the right member $\mathscr{Q}(f)$ describes the collision between the particles of the gas.In most cases, for dilute gases, this operator $\mathscr{Q}$ is quadratic taking into account "two-points" collisions.

A microscopic analysis of the molecular collisions shows that the mass, the momentum, and the energy are conserved during the collisions. The effect on the macroscopic scale that interest us here is that, in particular, the collision kernel $\mathscr{Q}(f)$ has zero integral when tested against 1, c, and $\frac{1}{2}c\cdot c$:

$$\int_{\mathbb{R}^d}\mathscr{Q}(f)\,\mathrm{d}c = 0, \qquad \int_{\mathbb{R}^d} c\mathscr{Q}(f)\,\mathrm{d}c = 0, \qquad \int_{\mathbb{R}^d}\frac{1}{2}c\cdot c\mathscr{Q}(f)\,\mathrm{d}c = 0. \tag{1.5}$$

By injecting these hypothesis into the Boltzmann equation (1.4), the conserved quantities $m = (\rho, q, E)$ are functions of time and space that satisfy the Euler equations of gas dynamics:

$$\partial_t m + \partial_x\cdot\varphi = 0. \tag{1.6}$$

In this equation, the flux φ reads

$$\varphi = \left(\int_{\mathbb{R}^d} cf\,\mathrm{d}c, \int_{\mathbb{R}^d} c\otimes cf\,\mathrm{d}c, \int_{\mathbb{R}^d} c\frac{1}{2}c\cdot cf\,\mathrm{d}c\right).$$

The equation (1.6) is a system of partial differential equations if the flux φ reads as a function of the conserved quatities m. We speak about the *closure* of the system. The asymptotic expansion of Hilbert [21] or the Chapman-Enskog development [7] can be used to obtain an asymptotic expression of the flux in the limit of small *Knudsen* number. The Knudsen number ϵ is a dimensionless number defined as the ratio of the mean free path (average length covered by a particle between two collisions) over a characteristic macroscopic length.

2 Cellular automata

In 1986, Frisch, Hasslacher, and Pomeau were responsible for a revolution: a simple cellular automaton obeying a few microscopic conservation laws was capable of reproducing the complexity of a real fluid flow [16]. The principle of this cellular automaton is to consider a hexagonal network in which each site hosts up to six particles, each of which has a speed enabling it to move to a neighbouring point during one time step. All the particles have the same mass, and therefore the same energy. Moreover, only one particle can be at a given site with a given speed (exclusion principle). Without going into detail about the algorithm, it is made up of two phases, a transport phase during which each particle changes

site according to its own velocity, and a collision phase during which the particles reorganise their velocities according to simple principles that satisfy the conservation of mass and momentum. The hexagonal lattice allows this automaton to respect a fundamental property of symmetry in the Navier-Stokes equation: invariance by rotation–which was not the case for the automaton of Hardy, Pomeau, and De Pazzis [18, 19] ten years earlier, in particular for the impulse flow.

Instead of seeking mathematical models, the development of computer modelling tools led to the idea of discrete simulators easy to program. In such an approach, the space, time, velocity, number of molecules present at a given time at a given point are discrete variables. The development of these cellular automata have been three highlights.

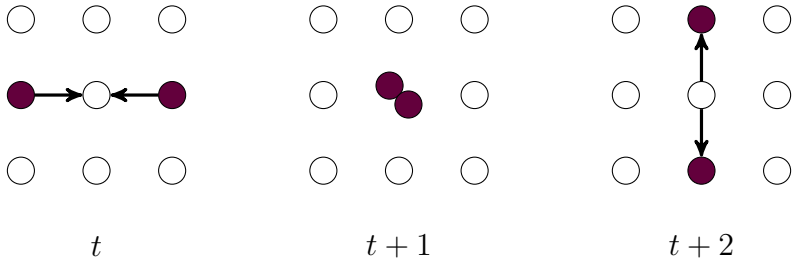


Figure 1.1 – *Frontal collision dynamics in the HPP [18] model.*

The first idea is to use a square two-dimensional lattice. The lattice set of Gaussian integers ($\mathbb{Z}[i] \cong \mathbb{Z}^2$) has a state defined by a binary variable field being 0 or 1. A value of 0 indicates that the site (i, j) is empty and the value 1 it is occupied. The discret evolution of the particles on the lattice is described by the velocities linking a vertex (i, j) to its four neighbors $(i\pm1, j\pm 1)$. With a unity space step and unity time step, the speed range therefore take values in the set $\{e_1, -e_1, e_2, -e_2\}$, with $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Each particle (or occupied site) has one of the four previous proposed velocities. It remains to define the collision rules when there is a conflict to occpy a site at a new discret time. Without describing in detail here the model of Hardy, de Pazzis, and Pomeau [18], we must build collision rules that respect the conservation of mass, momentum, and energy while taking into account a discret time and space. The figure 1.1 describes the dynamics in the evet of a frontal collision. We remark that during the intermediate time $t + 1$, two discret particles are present at the same time on the same vertex of the lattice.

A remarkable point in the study of the cellular automata is that it is possible (at least formally) to pass to the limit. Taking blocks of larger and larger size allows to define a macroscopic density ρ^∞ (ratio of the number of occupied sites towards the number of sample sites) and a macroscopic momentum q^∞. We introduce also a "large" time scale compared to the time scale of the simulation (equal to 1) and a "large" spatial scale compared to the lattice step (still equal to 1!). Using these continuous variables, the limit equations take the form

$$\partial_t \rho^\infty + \partial_x \cdot q^\infty = 0, \qquad \partial_t q^\infty + \partial_x \cdot P(\rho^\infty, q^\infty) = 0,$$

where P is a computable function. The conservations of the mass and of the momentum are then satisfied by the cellular automata at the macroscopic limit. But the pressure tensor P is not isotropic.

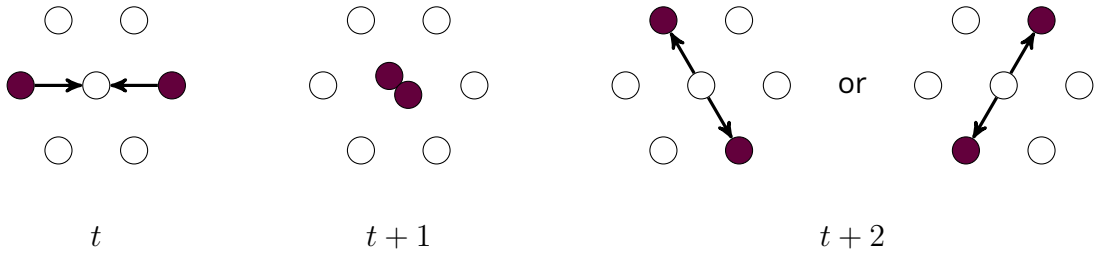


Figure 1.2 – *Frontal collision dynamics in the FPP [16] model.*

To remedy this defect of isotropy, Frisch, Hasslacher, and Pommeau [16] proposed to use a hexagonal lattice, *i.e.* vertices of the form $a + bj$, $a, b \in \mathbb{Z}$ and $1 + j + j^2 = 0$. The discrete velocity space contains more directions, the collision dynamics is therefore more complex (see figure 1.2). A random draw is needed to describe the post-collision state after a frontal collision. With this new model, the hydrodynamic limit is isotropic, therefore physically admissible. The extension to three space dimensions was realized soon after by d'Humières, Lallemand and Frisch [10] using a 4-dimensional model with 24 velocities and a face-centered cubic lattice.

However, a statistical noise—inherent to particle methods and highlighted by Orszag and Yakhot [26]—as well as the exponential complexity of the collision operator—this complexity is due to the fact that collisions between N bodies are taken into account, which is not the case for the Boltzmann equation at the continuous level in the case of sparse gases—making it less competitive than spectral methods in particular—and this despite numerous efforts, both theoretical and operational.

3 The lattice Boltzmann method

The first lattice Boltzmann scheme was proposed by Mac Namara and Zanetti in 1988 [24] with the explicit intention of avoiding the problem of statistical noise in cellular automata. The approach essentially consists of replacing Boolean numbers—a particle either occupies or does not occupy a site in the lattice—with an averaged population represented by a floating number. The change in perspective is exactly the same as that involved in moving from the particle view of a gas—where each particle is tracked in its movement—to continuous kinetic theory—for which the quantity tracked is an average particle density.

Mac Namara and Zanetti keep a discret lattice but seek real unknowns f which describes the average population on a given site, with a given velocity imposed by the geometry.

In order to eliminate the exponential complexity of the collision operator, Higuera, Succi and Benzi [20] introduce equilibrium distributions and a collision matrix, which makes it possible to write the collision operator collision in linear form. In [27], Quian, d'Humières and Lallemand propose a network Boltzmann scheme—later called BGK by the community—where the equilibria are polynomials in mass and velocity and where the collision matrix is propor-

tional to the identity. This scheme has since been widely used because it has satisfactory isotropy properties and it is of order 2. This approach was then enriched by d'Humières [9] who proposed a matrix of diagonal collisions in a space other than that of density functions, called moment space. It is within this framework that we systematically place ourselves in this dissertation. The reader interested in the historical aspect of Boltzmann network diagrams can refer to the works [8, 29].

We consider a physical space with d dimensions, $d \in \{1, 2, 3\}$. In this physical space there is an infinite regular lattice denoted by $\mathscr{L}$. Usually, the lattice is made up of squares for $d = 2$ or cubes for $d = 3$, but it could be made up of triangles, hexagons or tetrahedrons *etc*. In this lecture, we consider lattices like $\mathscr{L} = \Delta x \mathbb{Z}^d$, where $\Delta x > 0$ is the spatial step of the lattice. Moreover, we introduce the time step of the scheme $\Delta t > 0$ and the lattice velocity λ linked by

$$\lambda = \frac{\Delta x}{\Delta t}. \tag{1.7}$$

On each point of the lattice $x \in \mathscr{L}$, some particle distribution functions are put, denoted by f_j, $0 \leqslant j < q$. For each j, the particle distribution function $f_j(x, t)$ represents the number of particles on the vertex $x \in \mathscr{L}$ at time t that have the velocity c_j, $0 \leqslant j < q$. It could be seen as a discret representation of the particle distribution function of the kinetic theory. A lattice Boltzmann scheme is then a numerical evolution of these discret values that mimics the Boltzmann equation (1.4) with an operator splitting. We then describe one time step of the scheme, that is the agebraic operations used to transform the particle distribution functions at time t to $t+\Delta t$.

The first phase is the *relaxation phase* that mimics the collision: the particles on the same vertex of the lattice interact and recombine according to the laws prescribed by the scheme. We note with a superscript $\star$ the quantities after the relaxation phase. The framework proposed by d'Humières [9] consists of changing the basis for writing the relaxation phase. A invertible matrix $\boldsymbol{M}$ is chosen to transform the vector of the particle distribution functions $\boldsymbol{f} = (f_0, \dots, f_{q-1})^T$ into the vector of the moments $\boldsymbol{m} = (m_0, \dots, m_{q-1})^T$ according to the relations:

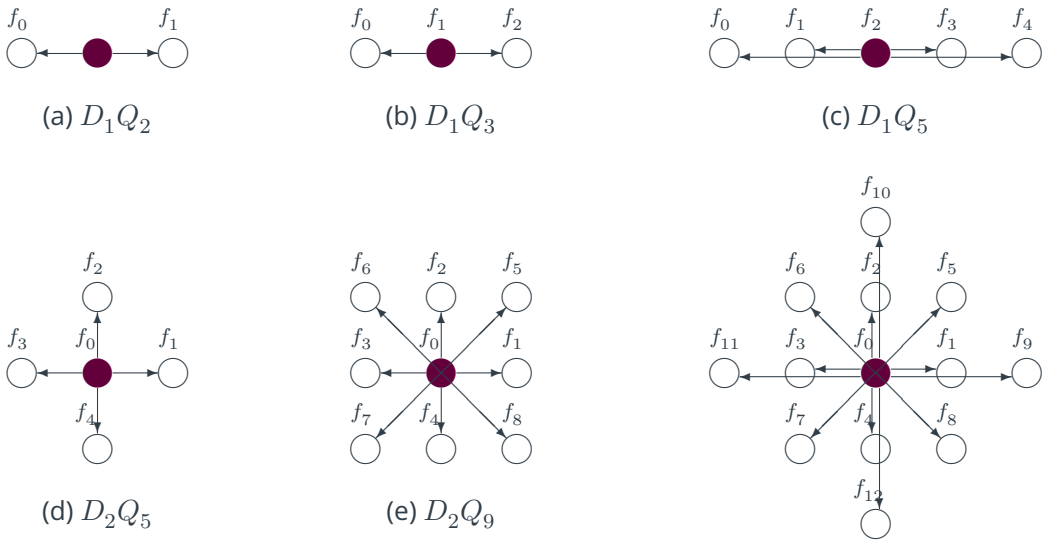
$$\boldsymbol{m} = \boldsymbol{M}\boldsymbol{f}, \qquad \boldsymbol{f} = \boldsymbol{M}^{-1}\boldsymbol{m}. \tag{1.8}$$

The relaxation phase preserves some of these moments: we consider that the N first moments $m_0, \dots, m_{N-1}$ are not modified, that reads:

$$m_k^\star = m_k, \quad 0 \leqslant k < N. \tag{1.9}$$

Usually, the N conserved moments are chosen among the conserved physical quantities like the mass, the momentum or the energy. This choice is obviously decisive, as we will see below. The $q-N$ non-conserved moments m_k, $N \leqslant k < q$ are modified through a linear relaxation toward an equilibrium state denoted m_k^{eq} thanks to a scalar relaxation parameter s_k:

$$m_k^\star = m_k + s_k(m_k^{\text{eq}} - m_k), \quad N \leqslant k < q. \tag{1.10}$$



(a) D_1Q_2 (b) D_1Q_3 (c) D_1Q_5

(d) D_2Q_5 (e) D_2Q_9 (f) D_2Q_{13}

Figure 1.3 – *Examples of the more classical free transport phases: the filled circle represents the initial cell and the arrows indicate the adapted velocities used to change the site. The terminology for the scheme D_dQ_q is q velocities on the lattice $\Delta x\mathbb{Z}^d$ in a d-dimensional space.*

The second phase is the *free transport*: each particle move freely on the lattice according to its personnal velocity. Thus, the particles with the velocity c_j on the vertex $x \in \mathscr{L}$ find themselves on the vertex $x+c_j\Delta t$ after the time step Δt. A key point is that, the spatial step Δx and the time step Δt being fixed, the velocities c_j, $0 \leqslant j < q$, are chosen to satisfy $c_j \in \lambda\mathbb{Z}^d$, so that the point $x+c_j\Delta t$ is also on the lattice. This free transport phase reads then

$$f_j(t+\Delta t, x+c_j\Delta t) = f_j^\star(t,x), \quad 0 \leqslant j < q. \tag{1.11}$$

An important property of the lattice Boltzmann method is that the transport step is treated exactly using the method of characteristics since the discrete velocities and mesh characteristics allow it. No projection or reconstruction step is necessary.

Now that we have clarified the two steps of the Lattice Boltzmann method, it remains to choose the free parameters in order to simulate the target system of partial differential equations. These parameters are of two types: the relaxation parameters s_k, $N \leqslant k < q$ are scalar numbers and the equilibrium values of the non-conserved moments m_k^{eq}, $N \leqslant k < q$, are functions of the conserved moments m_k, $0 \leqslant k < N$. The goal of this lesson is to give mathematical tools that can be used to fix these parameters.

Chapter 2

Example of the D_1Q_2

This chapter is devoted to the first simplest scheme with only two velocities in a mono-dimensional space. This scheme is however very robust and makes it possible to simulate any scalar conservation law, even non-linear. The interested reader could find more details and more results in [17].

We consider the following mono-dimensional hyperbolic equation

$$\partial_t \mathsf{u}(t,x) + \partial_x \varphi(\mathsf{u})(t,x) = 0, \qquad t > 0,\ x \in \mathbb{R}, \tag{2.1}$$

where the flux φ is a smooth function on $\mathbb{R}$. A two-velocities lattice Boltzmann scheme is used to approximate the solution of this equation.

1 Description of the scheme

We use the notation proposed by d'Humières in [9] by considering $\mathscr{L} = \Delta x\mathbb{Z}$, a regular lattice in one dimension of space with typical mesh size Δx. The time step Δt is determined after the specification of the velocity scale λ by the relation $\Delta t = \frac{\Delta x}{\lambda}$.

Definition 2.1 – *lattice velocity*

The lattice velocity is defined by

$$\lambda = \frac{\Delta x}{\Delta t}. \tag{2.2}$$

For the scheme denoted by D_1Q_2, we introduce $\{c_0 = -\lambda, c_1 = \lambda\}$ the set of the two velocities. We have therefore that for each node x of $\mathscr{L}$, and each c_j, $j \in \{0,1\}$, the point $x + c_j\Delta t$ is also a node of the lattice $\mathscr{L}$. The aim of the D_1Q_2 scheme is to compute a particles distribution vector $\boldsymbol{f} = (f_0, f_1)^T$ on the lattice $\mathscr{L}$ at discrete values of time. The

scheme splits into two phases for each time iteration: first, the relaxation phase that is local in space, and second, the transport phase for which an exact characteristic method is used.

The framework proposed by d'Humières reduced here to the two moments denoted by $\boldsymbol{m} = (m_0, m_1)^T$ and defined for each space point $x \in \mathscr{L}$ and for each time t by

$$m_0 = f_0 + f_1, \qquad m_1 = \lambda\,(-f_0 + f_1)\,. \tag{2.3}$$

The matrix of the moments $\boldsymbol{M}$ such that $\boldsymbol{m} = \boldsymbol{M}\boldsymbol{f}$ satisfies

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 \\ -\lambda & \lambda \end{pmatrix}, \qquad \boldsymbol{M}^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2\lambda} \\ \frac{1}{2} & \frac{1}{2\lambda} \end{pmatrix}. \tag{2.4}$$

Let us now describe one time step of the scheme. The start point is the density vector $\boldsymbol{f}(t,x)$ in $x \in \mathscr{L}$ at time t, the moments are computed by

$$\boldsymbol{m}(t,x) = \boldsymbol{M}\boldsymbol{f}(t,x). \tag{2.5}$$

The relaxation phase then reads

$$m_0^\star(t,x) = m_0(t,x), \qquad m_1^\star(t,x) = m_1(t,x) + s_1(m_1^{\mathrm{eq}}(t,x) - m_1(t,x)), \tag{2.6}$$

where s_1 is the relaxation parameter and m_1^{eq} the second moment at equilibrium that is a function of m_0. As a consequence, the first moment m_0 is conserved during the relaxation phase. The densities are then computed after the relaxation phase by

$$\boldsymbol{f}^\star(t,x) = \boldsymbol{M}^{-1}\boldsymbol{m}^\star(t,x). \tag{2.7}$$

The transport phase finally reads

$$f_0(t+\Delta t, x) = f_0^\star(t, x+\Delta x), \quad f_1(t+\Delta t, x) = f_1^\star(t, x-\Delta x). \tag{2.8}$$

2 Asymptotic Analysis : the Taylor expansion method

The aim of this section is to find the equivalent equations of the scheme and in particular to fix the equilibrium value m_1^{eq} as a function of m_0 in order to ensure that the scheme is consistent with (2.1). This reasoning consists in a formal development of the distribution functions $\boldsymbol{f}(t,x)$ at small Δt and Δx, assuming that these functions are regular enough to use the Taylor formula. The results of this section are particular cases of the general expansion of Dubois [11, 12].

Remark 2.2

The Taylor expansion method is a formal method to obtain the equivalent equations. A rigorous method will be given in the sequel based on a rewriting of the scheme.

The Taylor expansion method consists in expanding the distribution functions with respect to the small parameter Δt. Considering Eq. (2.8), we have

$$f_j + \Delta t \partial_t f_j + \tfrac{1}{2}\Delta t^2 \partial_{tt} f_j = f_j^\star - c_j \Delta t \partial_x f_j^\star + \tfrac{1}{2} c_j \cdot c_j \Delta t^2 \partial_{xx} f_j^\star + \mathcal{O}(\Delta t^3), \quad 0 \leqslant j \leqslant 1, \tag{2.9}$$

where the variables $x \in \mathscr{L}$ and t have been removed for readability. As the relaxation phase is written in the space of moments, we immediately take the moments of order 0 and 1 of Eq. (2.9) by summing over j after multiplication by 1 or by c_j:

$$m_0 + \Delta t \partial_t m_0 + \tfrac{1}{2}\Delta t^2 \partial_{tt} m_0 = m_0^\star - \Delta t \partial_x m_1^\star + \tfrac{\lambda^2}{2} \Delta t^2 \partial_{xx} m_0^\star + \mathcal{O}(\Delta t^3), \tag{2.10}$$

$$m_1 + \Delta t \partial_t m_1 + \tfrac{1}{2}\Delta t^2 \partial_{tt} m_1 = m_1^\star - \lambda^2 \Delta t \partial_x m_0^\star + \tfrac{\lambda^2}{2} \Delta t^2 \partial_{xx} m_1^\star + \mathcal{O}(\Delta t^3). \tag{2.11}$$

We then consider Eqs. (2.10) and (2.11) at order k for $0 \leqslant k \leqslant 2$.

At the zeroth order, the particle distribution functions are at equilibrium.

Proposition 2.3 – *zeroth order*

We assume that the relaxation parameter is non zero: $s_1 \neq 0$. Then, the particle distribution functions are close to their equilibrium values:

$$f_j = f_j^{\mathrm{eq}} + \mathcal{O}(\Delta t), \qquad f_j^\star = f_j^{\mathrm{eq}} + \mathcal{O}(\Delta t), \qquad 0 \leqslant j \leqslant 1, \tag{2.12}$$

and the non-conserved moment is close to its equilibrium value:

$$m_1 = m_1^{\mathrm{eq}} + \mathcal{O}(\Delta t), \qquad m_1^\star = m_1^{\mathrm{eq}} + \mathcal{O}(\Delta t). \tag{2.13}$$

Proof. Considering Eq. (2.10) at zeroth order does not give information as m_0 is conserved during the relaxation phase ($m_0 = m_0^\star$). Considering now Eq. (2.11) at zeroth order yields to $m_1 = m_1^\star + \mathcal{O}(\Delta t)$. Injecting this relation in Eq. (2.6) reads

$$m_1 = m_1 + s_1 (m_1^{\mathrm{eq}} - m_1) + \mathcal{O}(\Delta t).$$

As $s_1 \neq 0$, we obtain relations (2.13). Relations (2.12) on f_j and $f_j^\star$, $0 \leqslant j < 2$, are given by

$$f_0 = \tfrac{1}{2} m_0 - \tfrac{1}{2\lambda} m_1, \qquad f_0^\star = \tfrac{1}{2} m_0 - \tfrac{1}{2\lambda} m_1^\star, \qquad f_0^{\mathrm{eq}} = \tfrac{1}{2} m_0 - \tfrac{1}{2\lambda} m_1^{\mathrm{eq}},$$
$$f_1 = \tfrac{1}{2} m_0 + \tfrac{1}{2\lambda} m_1, \qquad f_1^\star = \tfrac{1}{2} m_0 + \tfrac{1}{2\lambda} m_1^\star, \qquad f_1^{\mathrm{eq}} = \tfrac{1}{2} m_0 + \tfrac{1}{2\lambda} m_1^{\mathrm{eq}}.$$

That ends the proof. □

Proposition 2.4 – *First order macroscopic equation*

We assume that the relaxation parameter is non zero: $s_1 \neq 0$. Then, the first moment m_0 satisfies the partial differential equation

$$\partial_t m_0 + \partial_x m_1^{\mathrm{eq}} = \mathcal{O}(\Delta t). \tag{2.14}$$

As a consequence, if $m_1^{\mathrm{eq}} = \varphi(m_0)$ then the first moment m_0 satisfies (2.1) at order 1.

Proof. We consider Eq. (2.10) at order 1:

$$m_0 + \Delta t \partial_t m_0 = m_0 - \Delta t \partial_x m_1^\star + \mathcal{O}(\Delta t^2).$$

We then formally derive relation (2.13) $m_1^\star = m_1^{\mathrm{eq}} + \mathcal{O}(\Delta t)$ to obtain the target partial differential equation. We concede that this proof is only formal for now... □

Definition 2.5 – *particular derivative*

The j-th particular derivative associated to the velocity c_j is defined by

$$\mathrm{d}_t^j = \partial_t + c_j \cdot \partial_x, \quad 0 \leqslant j < q. \tag{2.15}$$

We then define the equilibrium default θ by using the particular derivatives.

Definition 2.6 – *equilibrium default*

The equilibrium default for the D_1Q_2 is defined by

$$\theta = \sum_{j=0}^{1} c_j \cdot \mathrm{d}_t^j \, f_j^{\mathrm{eq}}. \tag{2.16}$$

Proposition 2.7

The equilibrium default θ can then be rewritten into the form

$$\theta = \partial_t m_1^{\mathrm{eq}} + \lambda^2 \partial_x m_0. \tag{2.17}$$

Proof. We expand the equilibrium default for the D_1Q_2:

$$\theta = -\lambda\big(\partial_t - \lambda\partial_x\big) f_0^{\mathrm{eq}} + \lambda\big(\partial_t + \lambda\partial_x\big) f_1^{\mathrm{eq}} = \partial_t\big(-\lambda f_0^{\mathrm{eq}} + \lambda f_1^{\mathrm{eq}}\big) + \lambda^2 \partial_x\big(f_0^{\mathrm{eq}} + f_1^{\mathrm{eq}}\big).$$

Using Eq. (2.3) ends the proof. □

Lemma 2.8 – *Transition lemma*

The second moment m_1 satisfies

$$m_1 = m_1^{\mathrm{eq}} - \frac{\Delta t}{s_1}\theta + \mathcal{O}(\Delta t^2), \qquad m_1^\star = m_1^{\mathrm{eq}} + \Delta t\Big(1 - \frac{1}{s_1}\Big)\theta + \mathcal{O}(\Delta t^2). \tag{2.18}$$

Moreover, we have

$$f_j^\star - f_j = \Delta t \, \mathrm{d}_t^j \, f_j^{\mathrm{eq}} + \mathcal{O}(\Delta t^2), \quad 0 \leqslant j \leqslant 1.$$

Proof. We consider Eq. (2.11) at order 1:

$$m_1 + \Delta t \partial_t m_1 = m_1^\star - \lambda^2 \Delta t \partial_x m_0 + \mathcal{O}(\Delta t^2).$$

Using Eq. (2.6) of the relaxation phase, we obtain

$$s_1(m_1^{\mathrm{eq}} - m_1) = \Delta t\big(\partial_t m_1 + \lambda^2 \partial_x m_0\big) + \mathcal{O}(\Delta t^2).$$

Once again, we derive a Taylor expansion without any rigorous argument. We replace $\partial_t m_1$ by $\partial_t m_1^{\mathrm{eq}}$ assuming that the error is of order 1 to obtain

$$s_1(m_1^{\mathrm{eq}} - m_1) = \Delta t\big(\partial_t m_1^{\mathrm{eq}} + \lambda^2 \partial_x m_0\big) + \mathcal{O}(\Delta t^2) = \Delta t\theta + \mathcal{O}(\Delta t^2).$$

Recombining this first relation with Eq. (2.6) ends the proof. □

Proposition 2.9 – *Second order macroscopic equation*

We assume that the relaxation parameter is non zero: $s_1 \neq 0$ and that the equilibrium value is given by $m_1^{\mathrm{eq}} = \varphi(m_0)$. Then, the first moment m_0 satisfies the second-order partial differential equation

$$\partial_t m_0 + \partial_x \varphi(m_0) = \Delta t\ \sigma_1\, \partial_x \left[\Big(\lambda^2 - \big(\varphi'(m_0)\big)^2\Big)\partial_x m_0\right] + \mathcal{O}(\Delta t^2), \tag{2.19}$$

with Henon's parameter $\sigma_1 = 1/s_1 - 1/2$.

Let us remark that this second-order macroscopic equation (2.19) then contains a diffusion term with a regularization effect if $\sigma_1 > 0$ (that is $0 < s_1 < 2$) and $|\varphi'(m_0)| < \lambda$. These conditions are indeed compatible with the stability conditions we derive in the sequel. In order to simulate the hyperbolic equation (2.1), the relaxation parameter s_1 could be taken equal to 2. But this term has a stabilization effect and it could be sometime useful to choose s_1 smaller to minimize the oscillations around the discontinuities.

Proof. We consider Eq. (2.10) at order 2:

$$m_0 + \Delta t\partial_t m_0 + \tfrac{1}{2}\Delta t^2 \partial_{tt} m_0 = m_0 - \Delta t\partial_x\big(m_1^{\mathrm{eq}} + \Delta t(1 - \tfrac{1}{s_1})\theta\big) + \tfrac{1}{2}\lambda^2 \Delta t^2 \partial_{xx} m_0,$$

by combining with the previous result of the transition lemma 2. The zeroth-order term vanishes, the first-order term reads $\partial_t m_0 + \partial_x \varphi(m_0)$ and the second-order one on the right-hand side

$$-(\tfrac{1}{2} - \tfrac{1}{s_1})\partial_x \theta = \sigma_1 \partial_x\big(\partial_t \varphi(m_0) + \lambda^2 \partial_x m_0\big) = \sigma_1 \partial_x\big[\big(\lambda^2 - \varphi'(m_0)^2\big)\partial_x m_0\big].$$

as $\partial_{tt} m_0 = -\partial_{tx} m_1^{\mathrm{eq}} + \mathcal{O}(\Delta t)$ and $\partial_t \varphi(m_0) = \varphi'(m_0)\partial_t m_0 = -\varphi'(m_0)\partial_x m_1^{\mathrm{eq}} + \mathcal{O}(\Delta t)$. □

3 Numerical illustrations

In this section, we perform two numerical simulations, one for the transport equation with a constant velocity, and one for Burger's equation. The lattice $\mathscr{L}$ is reduced to $[0, 1]$ and a homogeneous Neumann condition is added to treat the boundaries. In order to visualize the properties of the D_1Q_2 scheme, the initial condition is chosen as a regular bump function.

3.1 The transport equation

Let c be a real constant, we consider in this section $\varphi(\mathsf{u}) = c\mathsf{u}$. The equation reads

$$\partial_t \mathsf{u} + \partial_x(c\mathsf{u}) = 0.$$

We consider a smooth initial function. The solution has the same regularity for $t \geq 0$, the asymptotic analysis is then supposed to give the good estimation of the error. The results are given in Fig. 2.1.

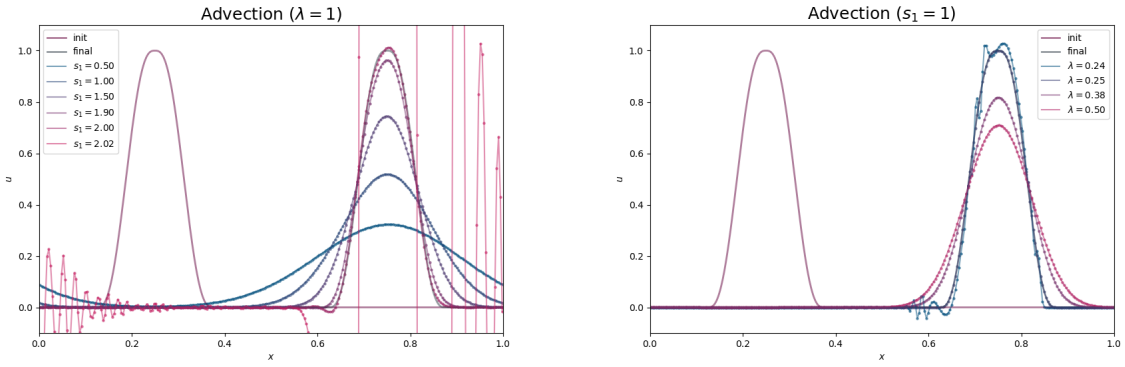


Figure 2.1 – *Transport equation with constant velocity $c = 0.25$ simulated with a D_1Q_2. Left: $\lambda = 1$ and several values for the relaxation parameter; Right: $s_1 = 1$ and several values for the lattice velocity.*

We observe that

- For $\lambda = 1 > c = 0.25$, the scheme proposes a reasonable numerical solution. For small values of s_1, the solution is however deteriorated with a large reduction of high and small values. For high values of s_1 (close to 2), we also observe that the solution is no more positive!
- For $s_1 = 1$, the scheme proposes a better solution when the lattice velocity λ is close to c by higher value.

In that case, the equivalent equation (2.19) reads

$$\partial_t m_0 + c\partial_x m_0 = \Delta t \sigma_1(\lambda^2 - c^2)\partial_{xx} m_0 + \mathcal{O}(\Delta t^2) = \Delta x \sigma_1 \frac{\lambda^2 - c^2}{\lambda}\partial_{xx} m_0 + \mathcal{O}(\Delta x^2).$$

If the space step Δx is fixed (in the simulations, we fixed $\Delta x = 1/256$), the numerical diffusion operator is proportional to σ_1 and to $(\lambda^2 - c^2)/\lambda$. Even if this equivalent condition only gives asymptotic informations, the diffusion operator acts with a regularization effect if the scalar coefficient is positive, that is if $0 < s_1 < 2$ and $|c| < \lambda$. The other case corresponding to $\lambda < |c|$ and $s_1 \notin [0, 2]$ is not taken into account as the scheme has to be stable for vanishing advection velocity $c = 0$. We then recover the behavior of the curves according to the variations of s_1 with fixed λ and *vice versa*.

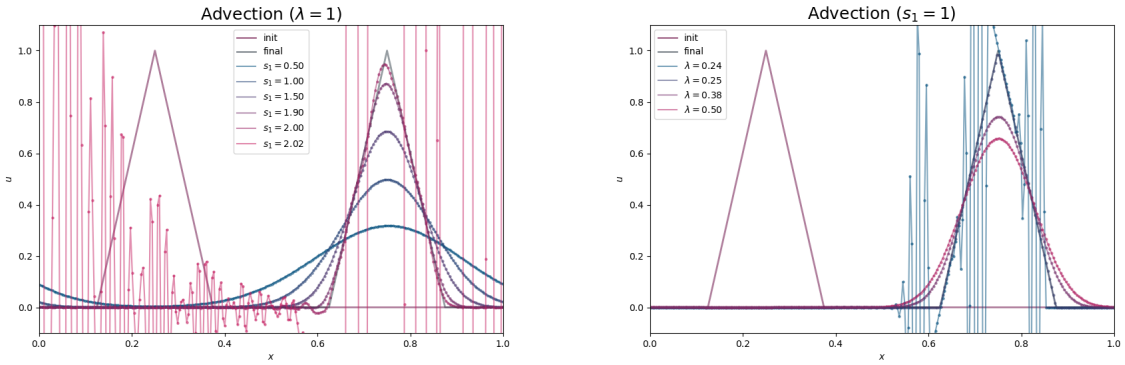


Figure 2.2 – *Transport equation with constant velocity $c = 0.25$ simulated with a D_1Q_2. Left: $\lambda = 1$ and several values for the relaxation parameter; Right: $s_1 = 1$ and several values for the lattice velocity.*

The same simulations are provided for a continuous initial condition but with discontinuous derivatives. The asymptotic analysis is then not directly possible but the conclusions remain the same.

3.2 Burgers equation

The Burgers equation is the simplest non-linear scalar equation that reads

$$\partial_t \mathrm{u} + \partial_x(\mathrm{u}^2/2) = 0.$$

We consider a piecewise affine function as initial condition. The exact solution can then be computed analyticaly also as a piecewise affine function. This solution is continuous until $t = 0.125$. Then a discontinuity appears and goes to the right vith a decreasing velocity. The results are given in Fig. 2.3.

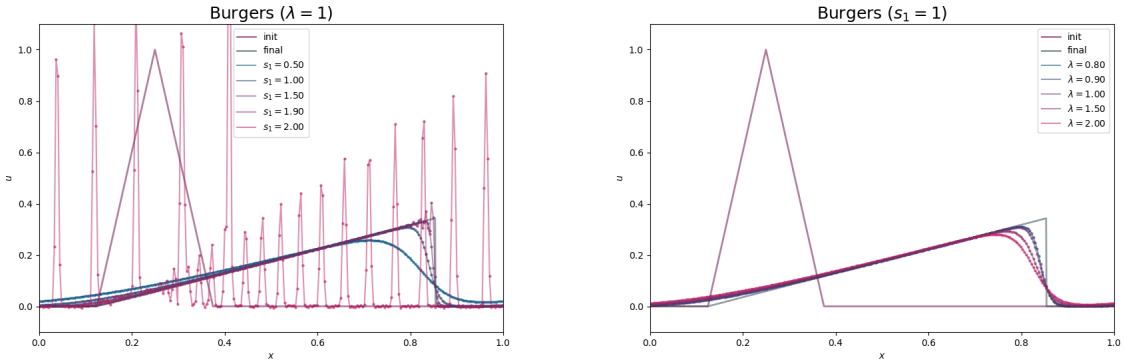


Figure 2.3 – *Burgers equation simulated with a D_1Q_2. Left: $\lambda = 1$ and several values for the relaxation parameter; Right: $s_1 = 1$ and several values for the lattice velocity.*

We observe the same behaviour

- For $\lambda = 1$, the scheme proposes a reasonable numerical solution if the relaxation parameter $s_1 \leqslant 2$. For small values of s_1, the solution is however deteriorated with a large reduction of high and small values. For high values of s_1 (close to 2), we also observe that the solution is no more positive!
- For $s_1 = 1$, the scheme proposes a better solution when the lattice velocity λ is close to 0.8 by higher value. Note that the value 0.8 is not an exact value for the change of behaviour and has no explaination for the moment...

4 Notions of stability

Many notions of stability are usefull for the Lattice Boltzmann method. Essentially, each notion is associated to the choice of a norm and the good one depends on the application. In this section, we present the von Neumann stability and the L^∞-stability for the D_1Q_2.

4.1 The von Neumann analysis

Attention, est-ce qu'il y a une inversion des vitesses ?

The von Neumann analysis concerns the linear schemes. We consider in this section $\varphi(\mathrm{u}) = c\mathrm{u}$ for a constant real value $c \in \mathbb{R}$. For $\xi \in [-\pi, \pi]$, we inject in the scheme a wave with ξ as a frequency:

$$f_0(t,x) = \alpha_0 e^{ix\xi/\Delta x}, \qquad f_1(t,x) = \alpha_1 e^{ix\xi/\Delta x}.$$

After one time step, we obtain

$$\begin{aligned}
m_0^\star(t,x) &= (\alpha_0 + \alpha_1) e^{ix\xi/\Delta x},\\
m_1^\star(t,x) &= (1-s_1)\lambda(\alpha_0 - \alpha_1) e^{ix\xi/\Delta x} + s_1 c(\alpha_0+\alpha_1) e^{ix\xi/\Delta x},\\
f_0^\star(t,x) &= (1 - \tfrac{s_1}{2}(1+\tfrac{c}{\lambda}))\alpha_0 e^{ix\xi/\Delta x} + \tfrac{s_1}{2}(1+\tfrac{c}{\lambda})\alpha_1 e^{ix\xi/\Delta x},\\
f_1^\star(t,x) &= \tfrac{s_1}{2}(1-\tfrac{c}{\lambda})\alpha_0 e^{ix\xi/\Delta x} + (1 - \tfrac{s_1}{2}(1-\tfrac{c}{\lambda}))\alpha_1 e^{ix\xi/\Delta x},\\
f_0(t+\Delta t,x) &= (1 - \tfrac{s_1}{2}(1+\tfrac{c}{\lambda}))e^{-i\xi} f_0(t,x) + \tfrac{s_1}{2}(1+\tfrac{c}{\lambda}) e^{-i\xi} f_1(t,x),\\
f_1(t+\Delta t,x) &= \tfrac{s_1}{2}(1-\tfrac{c}{\lambda})e^{i\xi} f_0(t,x) + (1 - \tfrac{s_1}{2}(1-\tfrac{c}{\lambda}))e^{i\xi} f_1(t,x).
\end{aligned}$$

The amplification matrix then reads

$$\boldsymbol{G}(\xi) = \begin{pmatrix} (1 - \frac{s_1}{2}(1+\frac{c}{\lambda}))e^{-i\xi} & \frac{s_1}{2}(1+\frac{c}{\lambda})e^{-i\xi} \\ \frac{s_1}{2}(1-\frac{c}{\lambda})e^{i\xi} & (1 - \frac{s_1}{2}(1-\frac{c}{\lambda}))e^{i\xi} \end{pmatrix}.$$

Definition 2.10 – *von Neumann stability*

The lattice Boltzmann scheme characterized by the amplification matrix $\boldsymbol{G}(\xi)$ is said stable if the sequel $(\boldsymbol{f}^n)_{n\geq 0}$ defined by $\boldsymbol{f}^{n+1} = \boldsymbol{G}(\xi)\boldsymbol{f}^n$ is bounded for any $\boldsymbol{f}^0 \in \mathbb{C}^3$ and for any $\xi \in [-\pi, \pi]$.

Proposition 2.11 – *von Neumann stability*

The scheme is stable in the sense of von Neumann if, and only if, the minimal polynomial of the amplification matrix $\boldsymbol{G}(\xi)$ is a simple von Neumann polynomial for any $\xi \in [-\pi, \pi]$.

Proof. The key point is the Dunford-Jordan decomposition of the complex square matrices. The matrix $\boldsymbol{G}$ is equivalent to a block diagonal matrix that reads

$$\begin{pmatrix} J_{\lambda_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{\lambda_b} \end{pmatrix}, \text{ where } J_{\lambda_k} = \overbrace{\begin{pmatrix} \lambda_k & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_k \end{pmatrix}}^{\mu_k \text{ columns}} \Bigg\} \mu_k \text{ rows}, \ 1 \leqslant k \leqslant b.$$

The complex values $\lambda_1, \dots, \lambda_b$ are the eigenvalues of $\boldsymbol{G}$ which are not necessarily distinct. The sequel of matrices $(\boldsymbol{G}^n)_{n\in\mathbb{N}}$ is then bounded if, and only if, we have first $|\lambda_k| \leqslant 1$, for $1 \leqslant k \leqslant b$, and second $\mu_k = 1$ if $|\lambda_k| = 1$. This condition is therefore exactly that the minimal polynomial of $\boldsymbol{G}$ is a simple von Neumann polynomial. □

As the minimal polynomial of the amplification matrix $\boldsymbol{G}$ is not easy to determine, we propose to compute its characteristic polynomial φ. Then, if $\varphi \in \mathscr{N}^s$, the minimal polynomial is also a simple von Neumann polynomial; if $\varphi \notin \mathscr{N}$, the minimal polynomial is not a simple von Neumann polynomial; and if $\varphi \in \mathscr{N} \backslash \mathscr{N}^s$, the case has to be investigated more precisely. Fortunately, the first two case are generic and the last case is not.

Theorem 2.12 – *von Neumann stability of the D_1Q_2*

The linear D_1Q_2 is stable in the sense of von Neumann if, and only if, one of these properties is true:

(i) $s_1 = 0$;
(ii) $0 < s_1 < 2$ and $\lambda \geq |c|$;
(iii) $s_1 = 2$ and $\lambda > |c|$.

Proof. We first define $t = s_1 - 1$ and $\gamma = c/\lambda$. We use the notations of Appendix B. The characteristic polynomial φ of the matrix $\boldsymbol{G}$ reads

$$\begin{aligned} \varphi(z;\xi,\beta,t) &= z^2 - \big((1-t)\cos\xi + i(1+t)\gamma\sin\xi\big)z - t, \\ \varphi^\star(z;\xi,\beta,t) &= -tz^2 - \big((1-t)\cos\xi - i(1+t)\gamma\sin\xi\big)z + 1, \\ \varphi^\circ(z;\xi,\beta,t) &= (1-t^2)z - (1-t^2)(\cos\xi + i\gamma\sin\xi), \\ \varphi^\bullet(z;\xi,\beta,t) &= 2z - (1-t)\cos\xi - i(1+t)\gamma\sin\xi. \end{aligned}$$

We use the characterization of Th. B.8 .

$$\begin{aligned} \varphi \in \mathscr{N}^s &\iff (\varphi \in \mathscr{I} \text{ and } \varphi^\circ \in \mathscr{N}^s) \text{ or } (\varphi \in \mathscr{E} \text{ and } \varphi^\bullet \in \mathscr{S}), \\ \varphi \in \mathscr{N} &\iff (\varphi \in \mathscr{I} \text{ and } \varphi^\circ \in \mathscr{N}) \text{ or } (\varphi \in \mathscr{E} \text{ and } \varphi^\bullet \in \mathscr{N}). \end{aligned}$$

We have $\varphi \in \mathscr{I} \iff |t| < 1$ and $\varphi \in \mathscr{E} \iff |t| = 1$. The first conclusion is that the scheme is not stable for $|t| > 1$.

Assuming $|t| < 1$, we immediately have

$$\begin{aligned} \varphi^\circ \in \mathscr{N}^s \iff \varphi^\circ \in \mathscr{N} &\iff |\cos\xi + i\gamma \sin\xi| \leqslant 1 \\ &\iff \cos^2\xi + \gamma^2 \sin^2\xi \leqslant \cos^2\xi + \sin^2\xi \\ &\iff \gamma^2 \leqslant 1 \text{ or } \xi \in \{0, \pm\pi\}. \end{aligned}$$

We conclude that, if $|t| < 1$, the scheme is stable, if and only if $|\gamma| \leqslant 1$.

For $t = 1$, we have $\varphi^\bullet = 2(z - i\gamma \sin\xi)$, then

$$\begin{aligned} \varphi^\bullet \in \mathscr{S} &\iff |\gamma \sin\xi| < 1 \iff \gamma^2 < 1/\sin^2\xi \text{ or } \xi \in \{0, \pm\pi\}, \\ \varphi^\bullet \in \mathscr{N} &\iff |\gamma \sin\xi| \leqslant 1 \iff \gamma^2 \leqslant 1/\sin^2\xi \text{ or } \xi \in \{0, \pm\pi\}. \end{aligned}$$

We conclude that, if $t = 1$, the scheme is stable for $|\gamma| < 1$ and is not stable for $|\gamma| > 1$. The case $\varphi^\bullet \in \mathscr{N} \backslash \mathscr{S}$ has to be investigated. We directly write the amplification matrix for $t = 1$ ($s_1 = 2$) and $\gamma = 1$:

$$\boldsymbol{G}(\xi) = \begin{pmatrix} -e^{-i\xi} & 2e^{-i\xi} \\ 0 & e^{i\xi} \end{pmatrix} \qquad \Longrightarrow \qquad \boldsymbol{G}^n(\pi/2) = i^n \begin{pmatrix} 1 & -2n \\ 0 & 1 \end{pmatrix}.$$

The power of $\boldsymbol{G}$ are not bounded for $\xi = \pm\pi/2$: the scheme is not stable. The case $\gamma = -1$ is exactly the same.

Finally, for $t = -1$, we have $\varphi^\bullet = 2(z - \cos\xi)$, then

$$\varphi^\bullet \in \mathscr{S} \iff \xi \notin \{0, \pm\pi\}, \qquad \varphi^\bullet \in \mathscr{N} \text{ is always true.}$$

The amplification matrix reads

$$\boldsymbol{G}(\xi) = \begin{pmatrix} e^{-i\xi} & 0 \\ 0 & e^{i\xi} \end{pmatrix}.$$

The power of $\boldsymbol{G}$ are bounded for every ξ: the scheme is stable. That ends the proof. □

Illustrations of the eigenvalues of the amplification matrix are given in Fig. 2.4 and 2.5.

4.2 The maximum principle

Another notion of stability can be useful for hyperbolic scalar conservation law. In this section, we consider the general nonlinear case $\varphi(\mathrm{u})$. Applying the theory of the characteritics, the exact solution of the PDE (2.1) satisfies the maximum principle:

$$\forall t > 0, \forall x \in \mathbb{R}, \quad \underline{\mathrm{u}} \leqslant \mathrm{u}(t,x) \leqslant \overline{\mathrm{u}}, \tag{2.20}$$

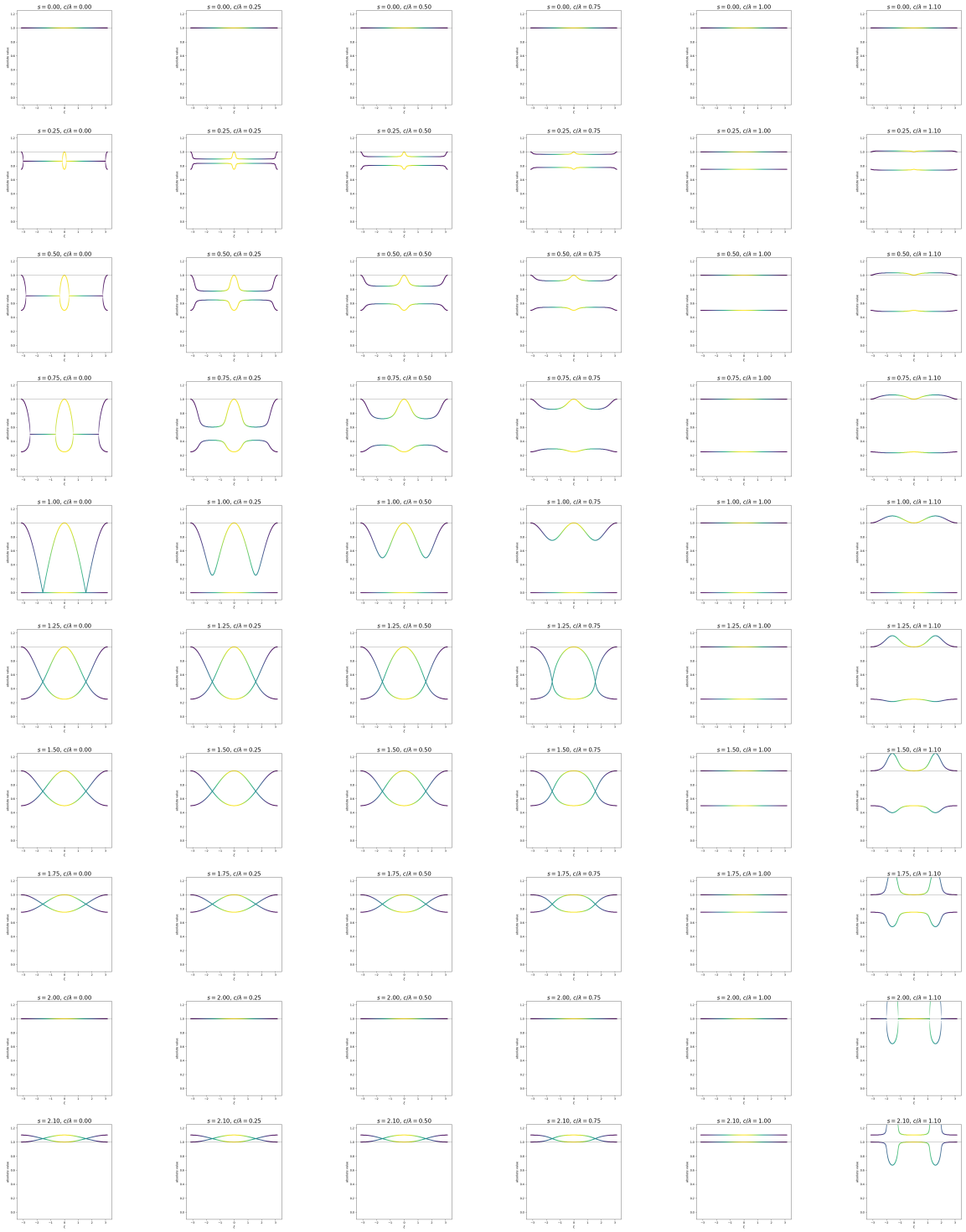


Figure 2.4 – *Modulus of the eigenvalues of the amplification matrix according to the variation of* s_1 *and* c/λ.

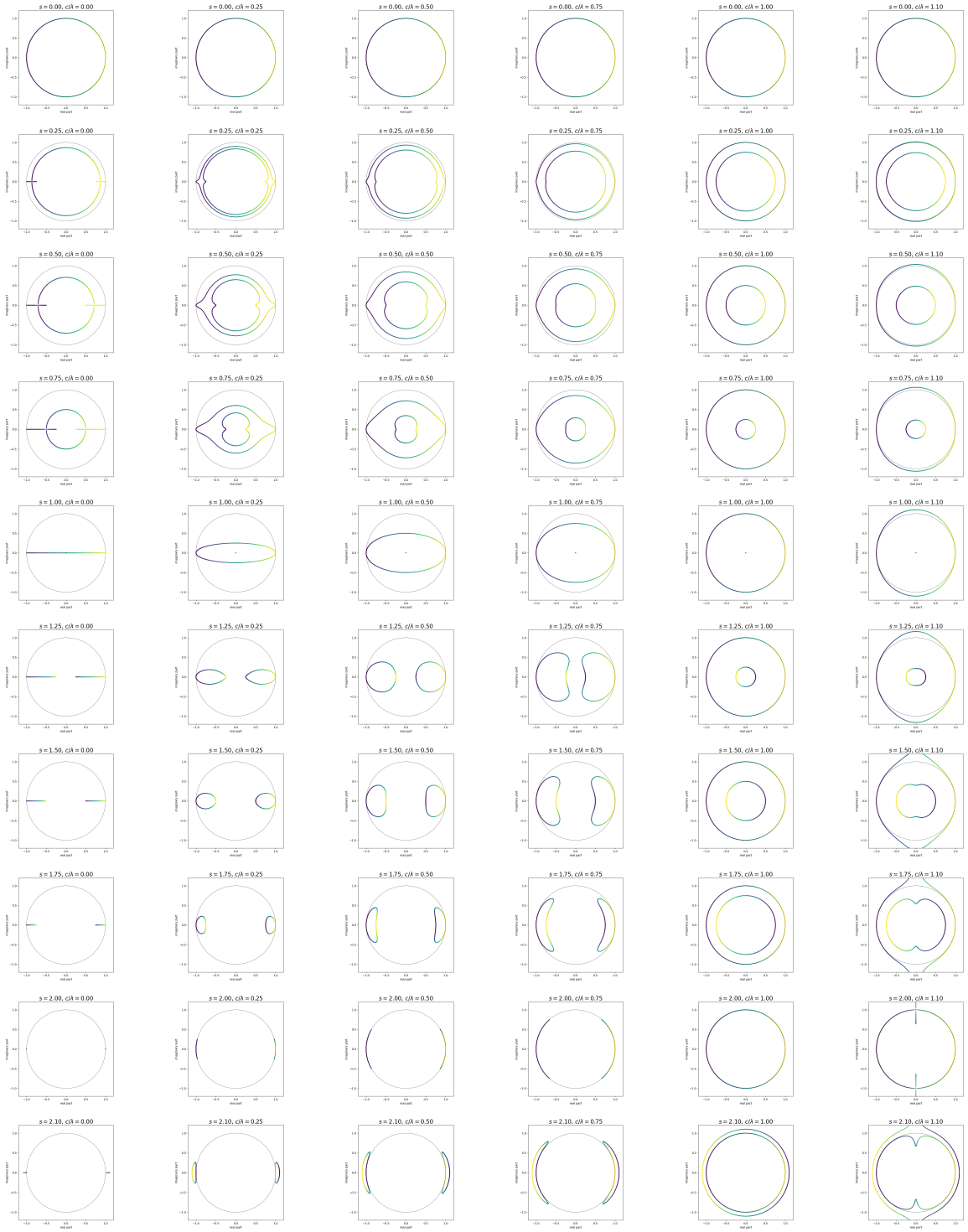


Figure 2.5 – *Eigenvalues of the amplification matrix according to the variation of s_1 and c/λ.*

with

$$\underline{\mathsf{u}} = \min_{y\in\mathbb{R}}(\mathsf{u}(0,y)), \qquad \overline{\mathsf{u}} = \max_{y\in\mathbb{R}}(\mathsf{u}(0,y)).$$

Require the same property for the numerical solution gives then a natural notion of stability.

Theorem 2.13 – *maximum principle of the D_1Q_2*

We assume $0 < s_1 \leqslant 1$, $m_1^{\text{eq}} = \varphi(m_0)$ and we introduce

$$M = \max\{|\varphi'(\xi)|, \text{ for } \underline{\mathsf{u}} \leqslant \xi \leqslant \overline{\mathsf{u}}\}.$$

Then, if $\lambda \geq M$, the D_1Q_2 satisfies the following maximum principle

$$\forall k \in \mathbb{N}, \forall j \in \mathbb{Z}, \quad \underline{\mathsf{u}} \leqslant m_0(k\Delta t, j\Delta x) \leqslant \overline{\mathsf{u}}.$$

Proof. We first write the relaxation phase on the particle distribution functions. We have

$$f_0^\star = \frac{1}{2}m_0^\star - \frac{1}{2\lambda}m_1^\star = \frac{1}{2}m_0 - \frac{1-s_1}{2\lambda}m_1 - \frac{s_1}{2\lambda}m_1^{\text{eq}} = (1-s_1)f_0 + s_1 f_0^{\text{eq}}(m_0),$$
$$f_1^\star = \frac{1}{2}m_0^\star + \frac{1}{2\lambda}m_1^\star = \frac{1}{2}m_0 + \frac{1-s_1}{2\lambda}m_1 + \frac{s_1}{2\lambda}m_1^{\text{eq}} = (1-s_1)f_1 + s_1 f_1^{\text{eq}}(m_0),$$

with

$$f_0^{\text{eq}}(\xi) = \frac{1}{2}\xi - \frac{1}{2\lambda}\varphi(\xi), \qquad f_1^{\text{eq}}(\xi) = \frac{1}{2}\xi + \frac{1}{2\lambda}\varphi(\xi).$$

The functions f_0^{eq} and f_1^{eq} are $\mathcal{C}^1(\mathbb{R})$ and their derivative is non negative over $[\underline{\mathsf{u}}, \overline{\mathsf{u}}]$. Then, they are non decreasing functions over $[\underline{\mathsf{u}}, \overline{\mathsf{u}}]$. Moreover, we have $f_0^{\text{eq}}(\xi) + f_1^{\text{eq}}(\xi) = \xi$.

A recursive reasoning over $k \in \mathbb{N}$ is done. Since $\mathsf{u}(0,\cdot) \in [\underline{\mathsf{u}}, \overline{\mathsf{u}}]$, the initial step imposes $m_0(0, j\Delta x) \in [\underline{\mathsf{u}}, \overline{\mathsf{u}}]$ for all $j \in \mathbb{Z}$. Since $m_1(0,\cdot) = \varphi(m_0)$, we get

$$f_j(0,\cdot) = f_j^{\text{eq}}(m_0(0,\cdot)) \in [f_j^{\text{eq}}(\underline{\mathsf{u}}), f_j^{\text{eq}}(\overline{\mathsf{u}})], \quad j \in \{0,1\}.$$

We then assume that the three inclusions are true for a certain $k \in \mathbb{N}$. As $s \in (0,1]$, $f_0^\star$ and $f_1^\star$ are respectively convex linear combinations of f_0 and f_0^{eq}, and f_1 and f_1^{eq}, so that $f_j^\star \in [f_j^{\text{eq}}(\underline{\mathsf{u}}), f_j^{\text{eq}}(\overline{\mathsf{u}})]$, $0 \leqslant j \leqslant 1$. The transport phase just shift the distribution functions, so that the inclusions yields for $f_0((k+1)\Delta t,\cdot)$ and $f_1((k+1)\Delta t,\cdot)$. Finally, summing the distribution functions gives

$$m_0((k+1)\Delta t, j\Delta x) \in [f_0^{\text{eq}}(\underline{\mathsf{u}}) + f_1^{\text{eq}}(\underline{\mathsf{u}}), f_0^{\text{eq}}(\overline{\mathsf{u}}) + f_1^{\text{eq}}(\overline{\mathsf{u}})] = [\underline{\mathsf{u}}, \overline{\mathsf{u}}],$$

as $f_0^{\text{eq}}(\xi) + f_1^{\text{eq}}(\xi) = \xi$. That ends the proof. □

Chapter 3

A notion of consistency

Lattice Boltzmann schemes rely on the enlargement of the size of the target problem in order to solve PDEs in a highly parallelizable and efficient kinetic-like fashion, split into a collision and a stream phase. This structure, despite the well-known advantages from a computational standpoint, is not suitable to construct a rigorous notion of consistency with respect to the target equations and to provide a precise notion of stability. In order to alleviate these shortages and introduce a rigorous framework, we demonstrate that any LB scheme can be rewritten as a corresponding multi-step FD scheme (Finite Differences) on the conserved variables. This is achieved by devising a suitable formalism based on operators, commutative algebra and polynomials. Therefore, the notion of consistency of the corresponding FD scheme allows to invoke the Lax-Richtmyer theorem in the case of linear LB schemes. Moreover, we show that the frequently-used von Neumann-like stability analysis for LB schemes entirely corresponds to the von Neumann stability analysis of their FD counterpart. More generally, the usual tools for the analysis of FD schemes are now readily available to study LB schemes.

1 Algebraic form of LB schemes

In this section, we present the formalism used in [1] to build the FD corresponding scheme. We first rewritte the LB scheme by using spatial and temporal shift operators.

1.1 Spatial and temporal discretization

We set the problem in any spatial dimension $d = 1, 2, 3$ considering the whole space $\mathbb{R}^d$, because we are not interested in studying boundary conditions. The space is discretized by a d-dimensional lattice $\mathscr{L} = \Delta x \mathbb{Z}^d$ of constant step $\Delta x > 0$ in all direction. The time is

uniformly discretized with step $\Delta t > 0$. The discrete instants of time shall be indexed by the integer indices $n \in \mathbb{N}$ so that the corresponding time is $t^n = n\Delta t$. We finally introduce the so-called lattice velocity $\lambda > 0$ defined by $\lambda = \Delta x/\Delta t$. Observe that the developing theory is totally discrete and thus fully independent from the scaling between Δx and Δt.

Definition 3.1 – *lattice velocity*

The lattice velocity is defined by

$$\lambda = \frac{\Delta x}{\Delta t}. \tag{3.1}$$

1.2 Discrete velocities and shift operators

The first choice to be made when devising a lattice Boltzmann scheme concerns the discrete velocities c_j, $0 \leqslant j < q$ with $q \in \mathbb{N}^\star$, which are multiples of the lattice velocity, namely $c_j \in \lambda\mathbb{Z}^d$ for any $0 \leqslant j < q$. Therefore, particles are stuck to move–at each time step–on the lattice $\mathscr{L}$.

We denote the particle distribution function moving with velocity c_j by f_j for every $0 \leqslant j < q$. The shift operators associated with the discrete velocities are an important element of the following analysis.

Definition 3.2 – *Shift operator*

Let $z \in \mathbb{Z}^d$, then the associated shift operator on the lattice $\mathscr{L}$, denoted $\mathsf{T}^z_{\Delta x}$, is defined in the following way. Take $\varphi : \mathscr{L} \to \mathbb{R}$ be any function defined on the lattice, then the action of $\mathsf{T}^z_{\Delta x}$ is

$$(\mathsf{T}^z_{\Delta x}\varphi)(x) = \varphi(x - z\Delta x), \qquad \forall x \in \mathscr{L}.$$

We also introduce $\mathscr{T}^d_{\Delta x} = \{\mathsf{T}^z_{\Delta x} \text{ with } z \in \mathbb{Z}^d\} \cong \mathbb{Z}^d$.

The shift yields information sought in the upwind direction with respect to the considered velocity. Let us introduce the natural binary operation between shifts.

Definition 3.3 – *Product*

Let the "product" $\circ : \mathscr{T}^d_{\Delta x} \times \mathscr{T}^d_{\Delta x} \to \mathscr{T}^d_{\Delta x}$ be the binary operation defined as $\mathsf{T}^z_{\Delta x} \circ \mathsf{T}^w_{\Delta x} = \mathsf{T}^{z+w}_{\Delta x}$, for any $z, w \in \mathbb{Z}^d$.

Henceforth, the product $\circ$ is understood whenever no ambiguity is possible. This operation provides an algebraic structure to the shifts, directly inherited from that of $\mathbb{Z}^d$.

Proposition 3.4

$(\mathscr{T}^d_{\Delta x}, \circ)$ forms an Abelian group.

Moreover, there is only "one movement" for each Cartesian direction which generates the shifts. More precisely

$$\begin{aligned}
&\text{for } d=1, \text{ let } \mathsf{x}=\mathsf{T}_{\Delta x}^{1}, && \text{then} \quad \mathscr{T}_{\Delta x}^{1}=\langle\{\mathsf{x}\}\rangle,\\
&\text{for } d=2, \text{ let } \mathsf{x}=\mathsf{T}_{\Delta x}^{(1,0)}, \mathsf{y}=\mathsf{T}_{\Delta x}^{(0,1)}, && \text{then} \quad \mathscr{T}_{\Delta x}^{2}=\langle\{\mathsf{x},\mathsf{y}\}\rangle,\\
&\text{for } d=3, \text{ let } \mathsf{x}=\mathsf{T}_{\Delta x}^{(1,0,0)}, \mathsf{y}=\mathsf{T}_{\Delta x}^{(0,1,0)}, \mathsf{z}=\mathsf{T}_{\Delta x}^{(0,0,1)}, && \text{then} \quad \mathscr{T}_{\Delta x}^{3}=\langle\{\mathsf{x},\mathsf{y},\mathsf{z}\}\rangle,
\end{aligned}$$

where $\langle\cdot\rangle$ is the customary notation for the generating set of a group.

We can add one more binary operation, which is non-internal to $\mathscr{T}_{\Delta x}^{d}$. This yields the cornerstone of this work, namely the set of FD operators, finite combinations of weighted shifts operators *via* a sum. It is defined as follows.

Definition 3.5 – *FD operators*

The set of FD operators on the lattice $\mathscr{L}$ is defined as

$$\mathscr{D}_{\Delta x}^{d}=\mathbb{R}\mathscr{T}_{\Delta x}^{d}=\left\{\sum\nolimits_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}\alpha_{\mathsf{T}}\mathsf{T}, \quad \text{where} \quad \alpha_{\mathsf{T}}\in\mathbb{R} \text{ and } \alpha_{\mathsf{T}}=0 \text{ a.e.}\right\},$$

the group ring (or group algebra) of $\mathscr{T}_{\Delta x}^{d}$ over $\mathbb{R}$. The sum $+:\mathscr{D}_{\Delta x}^{d}\times\mathscr{D}_{\Delta x}^{d}\to\mathscr{D}_{\Delta x}^{d}$ the product $\circ:\mathscr{D}_{\Delta x}^{d}\times\mathscr{D}_{\Delta x}^{d}\to\mathscr{D}_{\Delta x}^{d}$ of two elements are defined by

$$\begin{aligned}
\left(\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}\alpha_{\mathsf{T}}\mathsf{T}\right)+\left(\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}\beta_{\mathsf{T}}\mathsf{T}\right)&=\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}(\alpha_{\mathsf{T}}+\beta_{\mathsf{T}})\mathsf{T},\\
\left(\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}\alpha_{\mathsf{T}}\mathsf{T}\right)\circ\left(\sum_{\tilde{\mathsf{T}}\in\mathscr{T}_{\Delta x}^{d}}\beta_{\tilde{\mathsf{T}}}\tilde{\mathsf{T}}\right)&=\sum_{\mathsf{T},\tilde{\mathsf{T}}\in\mathscr{T}_{\Delta x}^{d}}\alpha_{\mathsf{T}}\beta_{\tilde{\mathsf{T}}}\mathsf{T}\circ\tilde{\mathsf{T}}.
\end{aligned}$$

Furthermore, the product of $\sigma\in\mathbb{R}$ with elements of $\mathscr{D}_{\Delta x}^{d}$ is given by

$$\sigma\left(\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}\alpha_{\mathsf{T}}\mathsf{T}\right)=\sum_{\mathsf{T}\in\mathscr{T}_{\Delta x}^{d}}(\sigma\alpha_{\mathsf{T}})\mathsf{T}.$$

With the two binary operations, $\mathscr{D}_{\Delta x}^{d}$ behaves closely to $\mathbb{Z}$, $\mathbb{R}$ or $\mathbb{C}$ as stated by the following result.

Proposition 3.6 – *Ring of FD operators*

$(\mathscr{D}_{\Delta x}^{d},+,\circ)$ is a commutative ring.

Observe that $(\mathscr{D}_{\Delta x}^{d},+,\circ)$ is not a field: not every element of $\mathscr{D}_{\Delta x}^{d}$ has multiplicative inverse, take for example the centered approximation of the derivative along x: $(\mathsf{T}_{\Delta x}^{-1}-\mathsf{T}_{\Delta x}^{1})/(2\Delta x)$. The elements having inverse are called "units" and divide all the other elements. It can be easily seen that the units are the product of a non-zero real number and a shift in $\mathscr{T}_{\Delta x}^{d}$. Indeed $(\alpha\mathsf{T}_{\Delta x}^{z})^{-1}=(1/\alpha)\mathsf{T}_{\Delta x}^{-z}$ for any $\alpha\in\mathbb{R}\setminus\{0\}$ and $z\in\mathbb{Z}^{d}$.

1.3 Monolithic scheme

We now introduce the vast class of lattice Boltzmann schemes we consider in the present chapter and for which we shall explain how to find the corresponding Finite Difference scheme. Any lattice Boltzmann scheme consists in an algorithm made up of two phases: a local collision phase performed on each site of the lattice and a stream phase, where particles are exchanged between different sites of the lattice. Let us recall each of them.

The N conserved moments, forming the variables of interest, are linear functions of the distributions densities $\boldsymbol{f} = (f_0, \dots, f_{q-1})$. Equally, the remaining $q - N$ non-conserved moments are linear functions of $\boldsymbol{f}$ as well. For this reason, we introduce a change of basis called moment matrix $\boldsymbol{M} \in \mathscr{Gl}_q(\mathbb{R})$. Thus, the moments are recovered by $\boldsymbol{m} = \boldsymbol{M}\boldsymbol{f}$ and *viceversa*. The entries of $\boldsymbol{M}$ can depend on Δx and/or on Δt, as we have stated that the scaling does not play any role here, but cannot be a function of the space and time variables.

We follow the general formalism of D'Humières [9], which easily and effectively accounts for multiple-relaxation-times (MRT) schemes. In this framework, the collision is written as a diagonal relaxation in the moments basis. Thus, the collision phase reads, denoting by $\star$ any post-collision state

$$\boldsymbol{m}^\star(x) = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{m}(x) + \boldsymbol{S}\boldsymbol{m}^{\mathrm{eq}}(x), \qquad x \in \mathscr{L}. \tag{3.2}$$

This part of the algorithm is local to each site of the lattice. Here

- $\boldsymbol{I} \in \mathscr{Gl}_q(\mathbb{R})$ is the identity matrix of size q;
- $\boldsymbol{S} \in \mathcal{M}_q(\mathbb{R})$ is the relaxation matrix which is a singular diagonal matrix with $\mathrm{rank}(\boldsymbol{S}) = q - N$, where $N \in \{1, \dots, q-1\}$ is the number of conserved moments:

  $$\boldsymbol{S} = \mathrm{diag}(0, \dots, 0, s_N, \dots, s_{q-1}),$$

  where the first N entries are zero and correspond to the conserved moments, the following $q - N$ are such that $s_k \in (0, 2]$ for $N \leqslant i < q$;
- we employ the notation $\boldsymbol{m}^{\mathrm{eq}}(x) = \boldsymbol{m}^{\mathrm{eq}}(m_0(x), \dots, m_{N-1}(x))$ for $x \in \mathscr{L}$, where $\boldsymbol{m}^{\mathrm{eq}} : \mathbb{R}^N \to \mathbb{R}^q$ are the equilibria for the moments, which are possibly non-linear functions of the N conserved moments. Since these equilibria are then multiplied by $\boldsymbol{S}$, the first N components do not need to be defined.

In the collision phase Eq. (3.2), the entries of $\boldsymbol{S}$ can depend on Δx or Δt, but not on space and time. The equilibria are allowed to follow the same dependencies plus those on space and time and can also depend on some "external variable" as in the case of vectorial schemes [17].

As previously said, the post-collisional distributions are recovered at each node by computing $\boldsymbol{f}^\star = \boldsymbol{M}^{-1}\boldsymbol{m}^\star$. The stream phase is diagonal in the space of the distributions densities. It can be written as

$$\boldsymbol{f}(t + \Delta t, x) = \mathrm{diag}(\mathsf{T}_{\Delta x}^{e_0}, \dots, \mathsf{T}_{\Delta x}^{e_{q-1}})\boldsymbol{f}^\star(t, x), \qquad x \in \mathscr{L}, \tag{3.3}$$

where the unit velocity vectors $e_j \in \mathbb{Z}^d$, $0 \leqslant j < q$ are defined by $c_j = \lambda e_j$. For the first time, the matrices have entries in a commutative ring $\mathscr{D}_{\Delta x}^d$ instead than in the field $\mathbb{R}$. The

set $\mathcal{M}_q(\mathscr{D}_{\Delta x}^d)$ of square matrices of size q with entries belonging to $\mathscr{D}_{\Delta x}^d$ forms a ring under the usual operations between matrices. Even if $\mathscr{D}_{\Delta x}^d$ is commutative from Proposition 1.2, $\mathcal{M}_q(\mathscr{D}_{\Delta x}^d)$ is not commutative for $q \geqslant 2$, as for real matrices and matrices of first-order differential operators.

The stream phase Eq. (3.3) can be rewritten in a non-diagonal form in the space of moments by introducing the matrix $\boldsymbol{T} = \boldsymbol{M}\,\mathrm{diag}(\mathsf{T}_{\Delta x}^{e_0}, \ldots, \mathsf{T}_{\Delta x}^{e_{q-1}})\boldsymbol{M}^{-1}$. One time step of the global scheme can then be written

$$\boldsymbol{m}(t+\Delta t, x) = \boldsymbol{A}\boldsymbol{m}(t,x) + \boldsymbol{B}\boldsymbol{m}^{\mathrm{eq}}(t,x), \qquad x \in \mathscr{L}, \tag{3.4}$$

with $\boldsymbol{A} = \boldsymbol{T}(\boldsymbol{I} - \boldsymbol{S})$ and $\boldsymbol{B} = \boldsymbol{T}\boldsymbol{S}$.

Consider a D_1Q_2 as example (see chapter 2), we compute the matrix $\boldsymbol{A}$ and $\boldsymbol{B}$. We have

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 \\ -\lambda & \lambda \end{pmatrix}, \quad \boldsymbol{M}^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2\lambda} \\ \frac{1}{2} & \frac{1}{2\lambda} \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} 0 & 0 \\ 0 & s_1 \end{pmatrix}.$$

We then obtain, denoting $\mathsf{x} = \mathsf{T}_{\Delta x}^1$ and $\bar{\mathsf{x}} = \mathsf{T}_{\Delta x}^{-1}$

$$\boldsymbol{T} = \begin{pmatrix} \frac{\mathsf{x}+\bar{\mathsf{x}}}{2} & \frac{\mathsf{x}-\bar{\mathsf{x}}}{2\lambda} \\ \lambda\frac{\mathsf{x}-\bar{\mathsf{x}}}{2} & \frac{\mathsf{x}+\bar{\mathsf{x}}}{2} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \frac{\mathsf{x}+\bar{\mathsf{x}}}{2} & (1-s_1)\frac{\mathsf{x}-\bar{\mathsf{x}}}{2\lambda} \\ \lambda\frac{\mathsf{x}-\bar{\mathsf{x}}}{2} & (1-s_1)\frac{\mathsf{x}+\bar{\mathsf{x}}}{2} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 0 & s_1\frac{\mathsf{x}-\bar{\mathsf{x}}}{2\lambda} \\ 0 & s_1\frac{\mathsf{x}+\bar{\mathsf{x}}}{2} \end{pmatrix}.$$

2 Corresponding multi-step scheme

The goal of this section is to explain the transformation of any lattice Boltzmann scheme into a multi-step scheme which involves only the conserved moments. Indeed, the study of the consistency is difficult because the non conserved moments are pure numeric values with no corresponding physical quantities.

In order to avoid this difficulty, we try to write a scheme only on the conserved moments. The idea is to write the matrix version of the scheme (see previsou section) and then to diagonalize it.

2.1 Characteristic polynomial and Cayley-Hamilton theorem

Polynomials with coefficients in $\mathscr{D}_{\Delta x}^d$ and matrices with entries in $\mathscr{D}_{\Delta x}^d$ play a central role in what we are going to develop.

Definition 3.7 – *Characteristic polynomial*

Let $\mathscr{R}$ be a commutative ring and $\boldsymbol{C} \in \mathcal{M}_r(\mathscr{R})$ for some $r \in \mathbb{N}^\star$. The characteristic polynomial of $\boldsymbol{C}$, denoted $\chi_{\boldsymbol{C}} \in \mathscr{R}[X]$, is given by $\chi_{\boldsymbol{C}} = \det(X\boldsymbol{I} - \boldsymbol{C})$, where $\det(\cdot)$ is the determinant and $\boldsymbol{I}$ is the $r \times r$ identity matrix.

A central result used in this work is the Cayley-Hamilton theorem for matrices over a commutative ring, see [3] for the proof, generalizing the same result holding for matrices on a field.

Theorem 3.8 – *Cayley-Hamilton*

Let $\mathscr{R}$ be a commutative ring and $\boldsymbol{C} \in \mathcal{M}_r(\mathscr{R})$ for some $r \in \mathbb{N}^\star$. Then $\chi_{\boldsymbol{C}}$ is a monic polynomial in the ring $\mathscr{R}[X]$ in the indeterminate X, under the form

$$\chi_{\boldsymbol{C}} = X^r + \gamma_{r-1}X^{r-1} + \dots \gamma_1 X + \gamma_0,$$

with $(\gamma_k)_{0\leqslant k\leqslant r} \subset \mathscr{R}$. Then $\chi_{\boldsymbol{C}}(\boldsymbol{C}) = 0$.

This result states that any square matrix with entries in a commutative ring verifies its characteristic equation.

2.2 Main result

In this section, we give only the main theorem for one conserved moments $N = 1$. The general case can be found in [1] and is just more technical to write.

Theorem 3.9 – *Corresponding multi-step scheme*

Let $N = 1$, then the lattice Boltzmann scheme given in (3.4) can be rewritten as a multi-step explicit scheme on the conserved moment m_0 under the form

$$m_0(t+\Delta t) = -\sum_{k=0}^{q-1} \gamma_k m_0(t-(q-1-k)\Delta t) + \left(\sum_{k=0}^{q-1}\left(\sum_{l=0}^{k}\gamma_{q+l-k}\boldsymbol{A}^l\right)\boldsymbol{B}\boldsymbol{m}^{\mathrm{eq}}(t-k\Delta t)\right)_0,$$

where $(\gamma_k)_{0\leqslant k\leqslant q}$ are the coefficients of the characteristic polynomial of $\boldsymbol{A}$: $\chi_{\boldsymbol{A}} = \sum_{k=0}^{q}\gamma_k X^k$.

Proof. The proof only consists in using the Cayley-Hamilton theorem to the matrix $\boldsymbol{A}$ and in combining the relations to eliminate all the terms with $m_1, \dots, m_{q-1}$. □

3 Example of the D_1Q_2

We already obtained the matrix formulation of the D_1Q_2 with

$$\boldsymbol{A} = \begin{pmatrix} \frac{\mathrm{x}+\bar{\mathrm{x}}}{2} & (1-s_1)\frac{\mathrm{x}-\bar{\mathrm{x}}}{2\lambda} \\ \lambda\frac{\mathrm{x}-\bar{\mathrm{x}}}{2} & (1-s_1)\frac{\mathrm{x}+\bar{\mathrm{x}}}{2} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 0 & s_1\frac{\mathrm{x}-\bar{\mathrm{x}}}{2\lambda} \\ 0 & s_1\frac{\mathrm{x}+\bar{\mathrm{x}}}{2} \end{pmatrix}.$$

Defining $\iota = (\mathsf{x} + \bar{\mathsf{x}})/2$ and $\delta = (\mathsf{x} - \bar{\mathsf{x}})/2$, we have $\iota^2 - \delta^2 = 1$. A direct calculation gives

$$\chi_{\boldsymbol{A}} = X^2 - (2 - s_1)\iota X - 1.$$

We then compute the corresponding two-step scheme directly as for a particular proof. We have

$$\begin{aligned} m_0(t + \Delta t) &= \iota m_0(t) + (1 - s_1)\frac{1}{\lambda}\delta m_1(t) + s_1\frac{1}{\lambda}\delta m_1^{\mathrm{eq}}(t), \\ \frac{1}{\lambda}\delta m_1(t) &= \delta^2 m_0(t - \Delta t) + (1 - s_1)\frac{1}{\lambda}\iota\delta m_1(t - \Delta t) + s_1\frac{1}{\lambda}\iota\delta m_1^{\mathrm{eq}}(t - \Delta t), \\ \iota m_0(t) &= \iota^2 m_0(t - \Delta t) + (1 - s_1)\frac{1}{\lambda}\iota\delta m_1(t - \Delta t) + s_1\frac{1}{\lambda}\iota\delta m_1^{\mathrm{eq}}(t - \Delta t). \end{aligned}$$

Combining these relations, we have

$$\frac{1}{\lambda}\delta m_1(t) = \iota m_0(t) - m_0(t - \Delta t).$$

And finally

$$m_0(t + \Delta t) = (2 - s_1)\iota m_0(t) - (1 - s_1)m_0(t - \Delta t) + s_1\frac{1}{\lambda}\delta m_1^{\mathrm{eq}}(t).$$

We can then compute the equivalent equation in the sense of the finite differences by using a Taylor expansion. We use the expansions:

$$\begin{aligned} \mathsf{x} m_0 &= m_0 - \lambda\Delta t\partial_x m_0 + \frac{\lambda^2\Delta t^2}{2}\partial_{xx} m_0 + \mathcal{O}(\Delta t^3), \\ \bar{\mathsf{x}} m_0 &= m_0 + \lambda\Delta t\partial_x m_0 + \frac{\lambda^2\Delta t^2}{2}\partial_{xx} m_0 + \mathcal{O}(\Delta t^3). \end{aligned}$$

The moment m_0 satisfies

$$\begin{aligned} m_0 + \Delta t\partial_t m_0 + \frac{\Delta t^2}{2}\partial_{tt} m_0 = (2 - s_1)\Big(m_0 + \frac{\lambda^2\Delta t^2}{2}\partial_{xx} m_0\Big) - s_1\frac{1}{\lambda}\Big(\lambda\Delta t\partial_x m_1^{\mathrm{eq}}\Big) \\ - (1 - s_1)\Big(m_0 - \Delta t\partial_t m_0 + \frac{\Delta t^2}{2}\partial_{tt} m_0\Big) + \mathcal{O}(\Delta t^3). \end{aligned}$$

At zeroth order, it reads

$$m_0 = (2 - s_1)m_0 - (1 - s_1)m_0 + \mathcal{O}(\Delta t) \iff m_0 = m_0 + \mathcal{O}(\Delta t).$$

At first order, we obtain

$$\partial_t m_0 = -s_1\partial_x m_1^{\mathrm{eq}} + (1 - s_1)\partial_t m_0 + \mathcal{O}(\Delta t) \iff s_1(\partial_t m_0 + \partial_x m_1^{\mathrm{eq}}) = \mathcal{O}(\Delta t).$$

Finally, at second-order, we recover, but rigourously, the second order macroscopic equation (2.19) previously obtained by the Taylor expansion method:

$$\partial_t m_0 + \partial_x m_1^{\mathrm{eq}} = \Delta t\Big(\frac{1}{s_1} - \frac{1}{2}\Big)(\lambda^2\partial_{xx} - \partial_{tt})m_0 + \mathcal{O}(\Delta t^2).$$

Appendix A

Fourier Transform

It is possible to analyze a linear scheme from its Fourier transform. The obvious benefit is that the derivation is then transformed into a multiplication.

We start by introducing the set of discrete summable square sequences scaled by the space step.

Definition A.1 – *space* $\ell^2(\Delta x\mathbb{Z})$

We define the space $\ell^2(\Delta x\mathbb{Z})$ by

$$\ell^2(\Delta x\mathbb{Z}) = \left\{ v = (v_j)_{j\in\mathbb{Z}} : \sum_{j\in\mathbb{Z}} |v_j|^2 < +\infty \right\}$$

which is provided with the norm

$$\|v\|_{2,\Delta x}^2 = \Delta x \sum_{j\in\mathbb{Z}} |v_j|^2.$$

On this space, we define the discret Fourier transform.

Definition A.2 – *discret Fourier transform*

For $v \in \ell^2(\Delta x\mathbb{Z})$, we define the function $\mathscr{F}(v)$ on $[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}]$ by

$$\mathscr{F}(v)(\xi) = \Delta x \sum_{j\in\mathbb{Z}} v_j e^{-ij\Delta x\xi}.$$

With this formula, the operator $\mathscr{F}$ is well defined for $v \in \ell^1$ and we extend as usual the definition to ℓ^2 by density.

Proposition A.3 – *Parseval formula*

The Fourier transform $\mathscr{F} : \ell^2(\Delta x\mathbb{Z}) \to L^2([-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}])$ satisfies the Parseval formula

$$\frac{1}{2\pi}\int_{-\pi/\Delta x}^{\pi/\Delta x} |\mathscr{F}(v)(\xi)|^2 \,\mathrm{d}\xi = \Delta x \sum_{j\in\mathbb{Z}} |v_j|^2.$$

Proof. We first assume that $v \in \ell^1 \cap \ell^2$. We then finish the proof by using the density of $\ell^1 \cap \ell^2$ in ℓ^2.

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\pi/\Delta x}^{\pi/\Delta x} |\mathscr{F}(v)(\xi)|^2 \,\mathrm{d}\xi &= \frac{1}{2\pi}\int_{-\pi/\Delta x}^{\pi/\Delta x} \mathscr{F}(v)(\xi)\overline{\mathscr{F}(v)(\xi)} \,\mathrm{d}\xi, \\
&= \frac{\Delta x^2}{2\pi}\int_{-\pi/\Delta x}^{\pi/\Delta x} \sum_{j,k\in\mathbb{Z}} v_j \bar{v}_k e^{-ij\Delta x\xi} e^{ik\Delta x\xi} \,\mathrm{d}\xi, \\
&= \frac{\Delta x^2}{2\pi}\sum_{j,k\in\mathbb{Z}} v_j \bar{v}_k \int_{-\pi/\Delta x}^{\pi/\Delta x} e^{-ij\Delta x\xi} e^{ik\Delta x\xi} \,\mathrm{d}\xi, \\
&= \Delta x \sum_{j\in\mathbb{Z}} |v_j|^2.
\end{aligned}$$

Of course, the exchange of summations and integrations has to be justified. □

Appendix B

Schur and von Neumann polynomials

Investigate the stability of a given scheme requires us to determine the location of roots of amplification polynomials, and in this appendix we present an algorithm for checking the roots of such polynomials. Let $\varphi \in \mathbb{C}[z]$ be a complex polynomial of degree d,

$$\varphi = a_d z^d + \ldots + a_0 = \sum_{\ell=0}^{d} a_\ell z^\ell.$$

We say that φ is of exact degree d if a_d is not zero.

1 Definitions

These definitions are based on the paper of Miller [25] and the book of Strikwerda [28].

Definition B.1 – *Schur polynomial*

The polynomial φ is a *Schur polynomial* if all its roots r satisfy $|r| < 1$.
We denote by $\mathscr{S}$ the set of the Schur polynomials.

Definition B.2 – *von Neumann polynomial*

The polynomial φ is a *von Neumann polynomial* if all its roots r satisfy $|r| \leqslant 1$.
We denote by $\mathscr{N}$ the set of the von Neumann polynomials.

Definition B.3 – *simple von Neumann polynomial*

The polynomial φ is a *simple von Neumann polynomial* if φ is a von Neumann polynomial and its roots on the unit circle are simple roots.
We denote by $\mathscr{N}^s$ the set of the simple von Neumann polynomials.

We finally define three operators on $\mathbb{C}[z]$, where the bar denotes the complex conjugate.

Definition B.4 – *operator* $\star$

For any polynomial $\varphi \in \mathbb{C}[z]$ of exact degree d we define the polynomial $\varphi^\star$ by

$$\varphi^\star(z) = \bar{a}_0 z^d + \ldots + \bar{a}_d = \sum_{\ell=0}^{d} \bar{a}_{d-\ell} z^\ell.$$

Definition B.5 – *operator* $\circ$

For any polynomial $\varphi \in \mathbb{C}[z]$ of exact degree d we define the polynomial φ° by

$$\varphi^\circ(z) = \frac{\varphi^\star(0)\varphi(z) - \varphi(0)\varphi^\star(z)}{z}.$$

Remark that φ° is indeed a polynomial as the constant term of $\varphi^\star(0)\varphi - \varphi(0)\varphi^\star$ vanishes. It is easy to see that the degree of φ° is less than that of φ. The operator $\circ$ is a kind of a derivative but for complex polynomials. We also define the classical derivative for polynomials.

Definition B.6 – *operator* $\bullet$

For any polynomial $\varphi \in \mathbb{C}[z]$ of exact degree d we define the polynomial $\varphi^\bullet$ by

$$\varphi^\bullet(z) = \sum_{\ell=0}^{d-1} (\ell+1) a_{\ell+1} z^\ell.$$

2 Theorems

The next three theorems give recursive tests for Schur, simple von Neumann and von Neumann polynomials. The proofs depend on Rouche's theorem from complex analysis.

Definition B.7

We denote

$$\mathscr{I} = \Big\{\varphi \in \mathbb{C}[z] \text{ s.t. } |\varphi(0)| < |\varphi^\star(0)|\Big\} \quad \text{and} \quad \mathscr{E} = \Big\{\varphi \in \mathbb{C}[z] \text{ s.t. } \varphi^\circ = 0\Big\}.$$

Theorem B.8 – *Cohn-Schur*

With these definitions, if $\varphi \in \mathbb{C}[z]$ of exact degree d, we have

$$\varphi \in \mathscr{S} \iff \varphi \in \mathscr{I} \text{ and } \varphi^\circ \in \mathscr{S}, \tag{B.1}$$

$$\varphi \in \mathscr{N}^s \iff \big(\varphi \in \mathscr{I} \text{ and } \varphi^\circ \in \mathscr{N}^s\big) \text{ or } \big(\varphi \in \mathscr{E} \text{ and } \varphi^\bullet \in \mathscr{S}\big), \tag{B.2}$$

$$\varphi \in \mathscr{N} \iff \big(\varphi \in \mathscr{I} \text{ and } \varphi^\circ \in \mathscr{N}\big) \text{ or } \big(\varphi \in \mathscr{E} \text{ and } \varphi^\bullet \in \mathscr{N}\big). \tag{B.3}$$

References

[1] Thomas Bellotti, Benjamin Graille, and Marc Massot. "Finite Difference formulation of any lattice Boltzmann scheme". *Numerische Mathematik* **152**.1 (Sept. 1, 2022), pp. 1–40.

[2] L. Boltzmann. "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen". *Sitzungsberichte Akademie der Wissenschaften* **66** (1872), pp. 275–370.

[3] James W. Brewer, John W. Bunce, and Fred S. Van Vleck. *Linear Systems over Commutative Rings*. Lecture Notes in Pure and Applied Mathematics. CRC Press, Apr. 22, 1986. 220 pp.

[4] S. Chapman. "On the Law of Distribution of Molecular Velocities, and on the Theory of Viscosity and Thermal Conduction, in a Non-Uniform Simple Monoatomic Gas". *Phil. Trans. Roy. Soc. Lond. A* **216** (1916), pp. 279–348.

[5] S. Chapman. "The Kinetic Theory of Simple and Composite Monoatomic Gases: Viscosity, Thermal Conduction, and Diffusion". *Proc. Roy. Soc. Lond. A* **93** (1916), pp. 1–20.

[6] S. Chapman. "On the Kinetic Theory of a Gas. Part II: A Composite Monoatomic Gas: Diffusion, Viscosity, and Thermal Conduction". *Phil. Trans. Roy. Soc. Lond. A* **217** (1918), pp. 115–197.

[7] S. Chapman and T. G. Cowling. *The Mathematical Theory of Non-Uniform Gases*. third. Cambridge Mathematical Library, 1970.

[8] S. Chen and G. D. Doolen. "Lattice Boltzmann Method for Fluid Flow". *Ann. Rev. Fluid Mech.* **30** (1998), pp. 329–364.

[9] D. d'Humière. "Generalized Lattice-Boltzmann equations". *Rarefied Gas Dynamics: Theory and simulation*. Vol. 159. AIAA Progress in astronomics and aeronautics, 1992, pp. 450–458.

[10] D. d'Humières, P. Lallemand, and U. Frisch. "Lattice gas models for 3D hydrodynamics". *Europhys. Lett.* **2**.4 (1986), pp. 291–297.

[11] F. Dubois. "Une introduction au schéma de Boltzmann sur réseau". *ESAIM Proc.* **18** (2007), pp. 181–215.

[12] F. Dubois. "Equivalent partial differential equations of a lattice Boltzmann scheme". *Computers and Mathematics with Applications.* **55** (2008), pp. 1441–1449.

[13] D. Enskog. "Kinetiche Theorie der Vorgänge in Mässig Verdünnten Gases". Inaugural Dissertation. Uppsala, 1917.

[14] D. Enskog. "The numerical calculation of phenomena in fairly dense gases". *Arkiv Mat. Astr. Fys* **16**.1 (1921), pp. 1–60.

[15] J. H. Ferziger and H. G. Kaper. *Mathematical Theory of Transport Processes in Gases*. North-Holland Publishing Company, Amsterdam, 1972.

[16] U. Frisch, B. Hasslacher, and Y. Pomeau. "Lattice gas automata for the Navier-Stokes equations". *Physical Review Letters* **56**.14 (1986), pp. 1505–1508.

[17] Benjamin Graille. "Approximation of mono-dimensional hyperbolic systems: a lattice Boltzmann scheme as a relaxation method". *Journal of Computational Physics* **266** (2014), pp. 74–88.

[18] J. Hardy, O. de Pazzis, and Y. Pomeau. "Time evolution of a two-dimensional model system I: invariant states and time correlation functions". *J. Math. Phys.* **14** (1973), 1746(14pp).

[19] J. Hardy, O. de Pazzis, and Y. Pomeau. "Molecular dynamics of a lattice gas: transport properties and time correlation functions". *Physical review A* **13** (1976), pp. 1949–1961.

[20] F. Higuera, S. Succi, and R. Benzi. "Lattice gas dynamics with enhanced collisions". *Europhys. Lett.* **9**.7 (1989), pp. 663–668.

[21] D. Hilbert. *Grundzüge einer Allgemeinen Theorie der Linearen Integralgleichungen*. Chelsea Publishing Co., New York, 1953.

[22] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird. *Molecular Theory of Gases and Liquids*. John Wiley & Sons, Inc., New York, 1954.

[23] R. L. Liboff. *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*. Second Edition. John Wiley & Sons, Inc., New York, 1998.

[24] G. McNamara and G. Zanetti. "Use of the Boltzmann equation to simulate lattice gas automata". *Phys. Rev. Lett.* **61**.20 (1988), pp. 2332–2335.

[25] John J. H. Miller. "On the Location of Zeros of Certain Classes of Polynomials with Applications to Numerical Analysis". *IMA Journal of Applied Mathematics* **8**.3 (1971), pp. 397–406.

[26] S. Orszag and V. Yakhot. "Reynolds number scaling of cellular automaton hydrodynamics". *Phys. Rev. Lett.* **56**.16 (1986), pp. 1691–1693.

[27] Y. H. Qian, D. d'Humières, and P. Lallemand. "Lattice BGK Models for Navier-Stokes equation". *Europhys. Lett.* **17**.6 (1992), pp. 479–484.

[28] John C. Strikwerda. *Finite difference schemes and partial differential equations*. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 2004. 435 pp.

[29] S. Succi. *The Lattice Boltzmann Equation: For Fluid Dynamics and Beyond*. Numerical Mathematics and Scientific Computation. Clarendon Press, 2001.

[30] W. G. Vincenti and C. H. Jr Kruger. *Introduction to Physical Gas Dynamics*. Second Printing. John Wiley & Sons, Inc., New York, 1967.