Computers and Mathematics with Applications 59 (2010) 2150–2167

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

On the stability structure for lattice Boltzmann schemes

Martin Rheinländer

Fachbereich Mathematik und Statistik, Universität Konstanz, Postfach D194, 78457 Konstanz, Germany

ARTICLE INFO

Keywords:
Lattice Boltzmann methods
Stability
Stability structure
Collision operators
Equilibrium

ABSTRACT

The stability structure for lattice Boltzmann schemes has been introduced in Banda et al. (2006) [16], Junk and Yong (2007) [14] to analyze the stability of numerical algorithms. The first purpose of this paper is to discuss the stability structure from the perspective of matrix analysis. Its second goal is to illustrate and apply the results to different classes of lattice Boltzmann collision operators. In particular we formulate an equivalence condition – just recently also reported in Yong (2008) [18] – that guarantees the existence of a pre-stability structure. It is then illustrated by several examples, how this equivalence condition can be effectively employed for the systematic verification and construction of stable collision operators. Finally, we point out some shortcomings of the stability structure approach arising in certain cases.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

For almost two decades lattice Boltzmann algorithms have been established as a viable numerical method to simulate processes in gas and fluid dynamics (see for example [1–5]). More precisely, lattice Boltzmann schemes can be considered as a specific way to discretize the Navier–Stokes equation [6] and other related equations. This can be shown by a consistency analysis using certain expansion techniques from statistical physics or alternatively from finite difference methods and asymptotic analysis as presented in various works like [5,7,8]. Despite increasing areas of applications there are still few results available providing a mathematically rigorous justification of lattice Boltzmann methods. In particular, stability is much less understood than consistency, although it is equally of practical and theoretical importance to guarantee a predictable and reliable performance of the numerical schemes as well as to fill a gap in convergence proofs.

Some early publications investigate the stability by means of numerical experiments [9,10]. In contrast [11] performs a sort of von Neumann stability analysis, which clearly represents a decisive step towards mathematical rigour. However, this approach suffers from the disadvantage of being restricted to periodic-domain or full-space problems. Furthermore, the stability of several simplified lattice Boltzmann model algorithms is thoroughly discussed in [12,13].

So far the most remarkable approach to treat the stability of lattice Boltzmann methods from a mathematical point of view is given in [14] preceded by [15] and [16]. Especially in [15] the convergence of the D2Q9 finite velocity Boltzmann equation towards the Navier–Stokes equation is proven. [15] also presents a precursor of the stability structure (Lemma 2.1) but it is mainly designed for the symmetric hyperbolic PDE system of the finite velocity Boltzmann equation. The definition of a stability structure for lattice Boltzmann schemes first appears in [16], where it is used to determine the range of the parameters in three parameterized Stokes equilibria combined with the BGK collision operator. However, this notion is exploited only in [14] to prove the stability of the lattice Boltzmann algorithm where several collision operators (BGK,TRT and MRT) are considered together with the Stokes equilibrium. The stability result is not only valid in the case of periodic boundary conditions but also includes bounce back boundary conditions. [15] and [14] became the point of departure to prove the convergence of lattice Boltzmann algorithms approximating the Stokes and Navier–Stokes equations [17].

E-mail address: martin.rheinlaender@uni-konstanz.de.

0898-1221/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2009.08.040

This work is mainly based on [14] and can be considered as an extension. As [14] is focused on the important but specific case of the Stokes equilibrium, we try here to extract those ideas which can be generalized and therefore permit a wider range of application.

Section 2 is mainly intended as a motivation of the stability structure for readers who are not so familiar with the analysis of hyperbolic relaxation problems. Section 3 is devoted to a discussion of the (pre-)stability structure in the light of linear algebra and matrix analysis. After illuminating one of the defining conditions for the (pre-)stability structure, it is observed that symmetric collision operators always admit a pre-stability structure. This suggests the question up to what extent the condition of symmetry could be generalized in order to guarantee the existence of a pre-stability structure. Trying to answer this question finally ends up with an equivalence statement (Theorem 6) which makes it easier to check whether a given collision operator has a pre-stability structure or not.

The main issue of this work is addressed in Section 4 where the results of Section 3 are applied to various lattice Boltzmann collision operators. In particular, the construction of MRT-type collision operators allowing for a stability structure is discussed (see Sections 4.1 and 4.3). The principal messages are

- to illustrate how the stability structure can be employed to verify and construct stable collision operators,
- to point out some shortcomings of the stability structure (cf. Section 2 concerning requirements of transport step, Sections 4.2 and 5).

Recently [18] established a connection between the stability structure for lattice Boltzmann schemes and a certain intrinsic structure of collision operators (Onsager-like relation) coming from nonequilibrium thermodynamics. This result anticipates our equivalence statement (Theorem 6) from another perspective.

Notation: As usual, the elements of $\mathbb{R}^q$, $q \in \mathbb{N}$ are considered as column vectors. The standard scalar product in $\mathbb{R}^q$ is denoted by brackets $\langle \cdot, \cdot \rangle$. Furthermore, if $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^q$ then $\boldsymbol{a}^2 \in \mathbb{R}^q$, $\boldsymbol{ab} \in \mathbb{R}^q$ and $\boldsymbol{a}/\boldsymbol{b} \in \mathbb{R}^q$ denote the vectors which are obtained by the respective componentwise operations.

2. Motivation of the stability structure

By the following reasoning we want to show how the stability structure for lattice Boltzmann schemes comes into play. Although the stability structure is by far not directly evident, its derivation does not involve any sophisticated or tricky arguments. In particular we focus on the following two aspects:

- What conditions on the lattice Boltzmann transport step are necessary in order to employ the existence of a stability structure as criterion for stability?
- Which further (restrictive) assumptions enter the derivation of the stability structure that may reduce its applicability.

It should be pointed out that the (pre-)stability structure was originally proposed to study hyperbolic relaxation problems (cf. [19]). Its appearance in the context of lattice Boltzmann methods is an exemplary application.

In contrast to other numerical methods the discretization by lattice Boltzmann schemes does not proceed in a direct manner but is based on a kinetic approach. Therefore the primary variables are not those occurring in the equation to be discretized but a so-called *population function* $\boldsymbol{f}$, which is considered as $\mathbb{R}^q$-valued function[1] of a discrete time and space domain: $\boldsymbol{f} : \mathcal{T} \times \mathcal{G} \to \mathbb{R}^q$. Classical lattice Boltzmann algorithms operate on regular grids that are characterized by a constant distance between nearest neighbor nodes. With the time step Δt the temporal grid becomes $\mathcal{T} = \{0, \Delta t, 2\Delta t, \dots\}$. In most cases the spatial grid $\mathcal{G} \subset \mathbb{R}^d$ is a square- or cubic-like grid where the minimal distance between two grid nodes is given by the grid spacing Δs.

Unlike $\mathcal{T}$, a complete description of $\mathcal{G}$ might be quite complicated in two and three dimensions ($d = 2, 3$), depending on the geometry of the spatial domain to be discretized. Similarly to finite difference methods for hyperbolic and parabolic partial differential equations, lattice Boltzmann algorithms require a strict coupling between Δt and Δs. The so-called *hyperbolic scaling* is obtained for $\Delta t = \Delta s$ (more generally $\Delta t \sim \Delta s$) while the *parabolic* or *diffusive scaling* is associated with the setting $\Delta t = \Delta s^2$ (or $\Delta t \sim \Delta s^2$ respectively).

Lattice Boltzmann algorithms fit into the class of explicit schemes. Therefore they can be written in the form of the following iteration

$$\boldsymbol{f}(t + \Delta t, \cdot) = E_{\mathcal{G}} \boldsymbol{f}(t, \cdot) \tag{1}$$

starting with an initial value $\boldsymbol{f}(0, \cdot)$ that must be provided somehow from the initial value problem to be discretized. $E_{\mathcal{G}}$ is denoted as *evolution operator* which operates on the set $\mathcal{F}(\mathcal{G}, \mathbb{R}^q)$ of $\mathbb{R}^q$-valued (spatial) grid functions. As the lattice Boltzmann time step decomposes into two substeps – the collision followed by the transport – the evolution operator is obtained as product of the *transport step operator* $T_{\mathcal{G}}$ and the *collision step operator* $C_{\mathcal{G}}$:

$$E_{\mathcal{G}} = \underbrace{T_{\mathcal{G}}}_{\substack{\text{transport}\\ \text{step}}} \cdot \underbrace{C_{\mathcal{G}}}_{\substack{\text{collision}\\ \text{step}}} . \tag{2}$$

[1] The components of $\boldsymbol{f}$ are associated with densities of fictitious species of particles. These particles travel with constant velocities which characterize the species.

Boundary conditions are usually encoded in the transport step operator. In the sequel it is assumed that the evolution operator is linear.

From (1) it is deduced that $\boldsymbol{f}(n\Delta t,\cdot)=E^n_{\mathcal{G}}\boldsymbol{f}(0,\cdot)$. Roughly speaking stability means that the population function $\boldsymbol{f}$ remains bounded on every finite time interval, i.e. whenever $n\Delta t<t_{\max}$. In particular, this is satisfied if

$$\|E^n_{\mathcal{G}}\|_{\mathcal{G}}\le K\quad\text{for all } n\in\mathbb{N}_0 \tag{3}$$

where the constant $K>0$ must be independent of n and the discretization parameter $h:=\Delta s$ which determines the refinement of both $\mathcal{T}$ and $\mathcal{G}$ due to the scaling relations. However K may depend on the chosen norm $\|\cdot\|_{\mathcal{G}}$.

To exploit the product structure (2) of the evolution operator let us apply the standard estimate:

$$\|E^n_{\mathcal{G}}\|_{\mathcal{G}}\le\|E_{\mathcal{G}}\|^n_{\mathcal{G}}=\|T_{\mathcal{G}}C_{\mathcal{G}}\|^n_{\mathcal{G}}\le\left(\|T_{\mathcal{G}}\|_{\mathcal{G}}\|C_{\mathcal{G}}\|_{\mathcal{G}}\right)^n=\|T_{\mathcal{G}}\|^n_{\mathcal{G}}\|C_{\mathcal{G}}\|^n_{\mathcal{G}}.$$

This motivates the following sufficient stability condition:

If one can find a norm $\|\cdot\|_{\mathcal{G}}$ such that

- the transport step operator $T_{\mathcal{G}}$ becomes an *isometry*, i.e. $\|T_{\mathcal{G}}\|_{\mathcal{G}}=1$,
- the collision step operator $C_{\mathcal{G}}$ becomes a *contraction*, i.e. $\|C_{\mathcal{G}}\|_{\mathcal{G}}\le1$,

then the corresponding lattice Boltzmann scheme is stable over any finite time interval with respect to this norm and (3) holds with $K=1$. Thereby the evolution operator $E_{\mathcal{G}}$ itself becomes a contraction.

Classical lattice Boltzmann transport operators act like permutations by shifting the values of the population function between the grid nodes (bulk transport) or swapping the population indices. More concretely, let us consider the class of transport operators being characterized by the following property:

There exists a bijection $\Pi:\mathcal{G}\times\{1,\dots,q\}\to\mathcal{G}\times\{1,\dots,q\}$ depending on $T_{\mathcal{G}}$ such that for all $\mathbb{R}^q$-valued grid functions $\boldsymbol{g}$ and $\tilde{\boldsymbol{g}}$ over $\mathcal{G}$ satisfying $T_{\mathcal{G}}\tilde{\boldsymbol{g}}=\boldsymbol{g}$ we have

$$g_j(\mathbf{y})=\tilde{g}_i(\mathbf{x})\quad\text{if } \Pi(\mathbf{x},i)=(\mathbf{y},j). \tag{4}$$

Above all, this condition includes those transport operators realizing the standard boundary conditions of

- *periodic*,
- *bounce back*,
- and *bounce back* with *flipping of sign*

type.[2] Observe that transport operators encoding periodic or bounce back boundary conditions are represented by permutation matrices. Hence the corresponding transport matrices are orthonormal (unitary) and bistochastic. In the case of bounce back boundary conditions with sign flipping, the entries of the representing matrices also attain -1 beside 0 and 1. Even if the permutation property and bistochasticity is then lost, the transport matrices still stay orthonormal and result into permutation matrices if the entries are replaced by their modulus.

Thanks to these properties the transport is an isometry with respect to any operator norm induced by a homogeneous vector norm[3] on $\mathcal{F}(\mathcal{G},\mathbb{R}^q)$. As $T_{\mathcal{G}}$ does not modify the values of $\tilde{\boldsymbol{g}}$ but only their assignment to grid nodes and population indices, $T_{\mathcal{G}}$ is particularly isometric with respect to additive norms like the ℓ^2-norm. Due to (4) we have

$$\|\tilde{\boldsymbol{g}}\|^2_{\mathcal{G}}:=\sum_{\mathbf{x}\in\mathcal{G}}\sum_{i=1}^{q}\tilde{g}_i(\mathbf{x})^2=\sum_{\mathbf{y}\in\mathcal{G}}\sum_{j=1}^{q}g_i(\mathbf{x})^2=:\|\boldsymbol{g}\|^2_{\mathcal{G}}=\|T_{\mathcal{G}}\tilde{\boldsymbol{g}}\|^2_{\mathcal{G}}. \tag{5}$$

To gain more flexibility, we give up homogeneity and introduce a tunable weight vector $\boldsymbol{b}\in\mathbb{R}^q$ with positive components. If $\boldsymbol{b}$ is compatible with $T_{\mathcal{G}}$ so that

$$b_i=b_j\quad\text{for } i\ne j\quad\text{if there exist } \mathbf{x},\mathbf{y}\in\mathcal{G}\quad\text{with } \Pi(\mathbf{x},i)=(\mathbf{y},j) \tag{6}$$

then also

$$|||\tilde{\boldsymbol{g}}|||^2_{\mathcal{G}}:=\sum_{\mathbf{x}\in\mathcal{G}}\sum_{i=1}^{q}b_i\tilde{g}_i(\mathbf{x})^2 \tag{7}$$

turns $T_{\mathcal{G}}$ into an isometry. As in (5) we have $|||\boldsymbol{g}|||^2_{\mathcal{G}}=|||T_{\mathcal{G}}\tilde{\boldsymbol{g}}|||^2_{\mathcal{G}}=|||\tilde{\boldsymbol{g}}|||^2_{\mathcal{G}}$ since the sums corresponding to $|||\boldsymbol{g}|||^2_{\mathcal{G}}$ and $|||\tilde{\boldsymbol{g}}|||^2_{\mathcal{G}}$ differ just by the ordering of summation. (7) represents a weighted ℓ^2-norm which is generated by a scalar product.

[2] Bounce back models homogeneous Dirichlet (or no-slip) boundary conditions if the Stokes or Navier–Stokes equation is approximated by lattice Boltzmann algorithm. If a scalar transport equation is approximated, bounce back imitates no-flux boundary conditions which correspond to Neumann (Robin) boundary conditions in the case of the diffusion (advection–diffusion) equation. The additional flipping of the sign then yields homogeneous Dirichlet conditions.

[3] Taking all entries of the vector equally into account.

Now we have to find out how $\boldsymbol{b}$ should be selected so that the collision step operator becomes a contraction. For our further reasoning it is advantageous to exploit the fact that the collision step operator performs a nodal operation. Hence there exists a matrix $C \in \mathbb{R}^{q\times q}$ (referred to as *nodal collision step matrix*) satisfying

$$\left(C_{\mathcal{G}}\boldsymbol{g}\right)(\mathbf{x}) = C\boldsymbol{g}(\mathbf{x}).$$

In words this means: in order to evaluate the grid function $C_{\mathcal{G}}\boldsymbol{g}$ at the grid node $\mathbf{x} \in \mathcal{G}$, it is enough to know $\boldsymbol{g}$ at $\mathbf{x}$ and to apply C on $\boldsymbol{g}(\mathbf{x}) \in \mathbb{R}^q$. The dimensionally reduced collision step operator is normally written as

$$C = \underbrace{I}_{\text{identity}} + \underbrace{J}_{\substack{\text{(physical)}\\ \text{collision}\\ \text{operator}}}$$

where J stands for the (physical) collision operator[4] not to be confused with the (nodal) collision step operator $C_{\mathcal{G}}$ (C).

If C is a contraction in $\mathbb{R}^q$ with respect to some norm $|||\cdot|||$, then the related norm $\sum_{\mathbf{x}\in\mathcal{G}} |||\cdot|||$ in $\mathcal{F}(\mathcal{G}, \mathbb{R}^q)$ makes $C_{\mathcal{G}}$ become a contraction in $\mathcal{F}(\mathcal{G}, \mathbb{R}^q)$. To find an appropriate norm $|||\cdot|||$ induced by a scalar product we start with the ansatz

$$|||\boldsymbol{p}||| = \|B\boldsymbol{p}\| \quad \text{for } \boldsymbol{p} \in \mathbb{R}^q$$

where $\|\cdot\|$ denotes the Euclidean norm (which is the ℓ^2-norm) in $\mathbb{R}^q$ and $B \in \mathbb{R}^{q\times q}$ is some invertible matrix. So the task is to determine B such that $|||C\boldsymbol{p}||| \le |||\boldsymbol{p}|||$ for all $\boldsymbol{p} \in \mathbb{R}^q$. Now let us make the crucial assumption that B and C satisfy the relation

$$BC = B(I + J) = (I + \Lambda)B \tag{8}$$

with some $\Lambda \in \mathbb{R}^{q\times q}$. Then the above requirement on B reads

$$|||C\boldsymbol{p}||| = \|BC\boldsymbol{p}\| = \|(I+\Lambda)B\boldsymbol{p}\| \le \|B\boldsymbol{p}\| = |||\boldsymbol{p}||| \quad \text{for all } \boldsymbol{p} \in \mathbb{R}^q.$$

As B is supposed to be invertible and thus bijective in $\mathbb{R}^q$, this can be true if and only if

$$\|I + \Lambda\| \le 1. \tag{9}$$

Considering the special case of $\Lambda = -\text{diag}(\lambda_1, \ldots, \lambda_q)$ being diagonal, this equality is satisfied if and only if the diagonal elements are non-positive and of modulus not larger than 2. Thus a sufficient condition on Λ to satisfy (9)

$$\Lambda = -\text{diag}(\lambda_1, \ldots, \lambda_q) \quad \text{with } \lambda_i \in [0, 2] \text{ for } i \in \{1, \ldots, q\}. \tag{10}$$

The corresponding norm[5] which turns $C_{\mathcal{G}}$ into a contraction is

$$|||\boldsymbol{g}|||_{\mathcal{G}}^2 := \sum_{\mathbf{x}\in\mathcal{G}} |||\boldsymbol{g}(\mathbf{x})|||^2 = \sum_{\mathbf{x}\in\mathcal{G}} \langle B\boldsymbol{g}(\mathbf{x}), B\boldsymbol{g}(\mathbf{x})\rangle = \sum_{\mathbf{x}\in\mathcal{G}} \langle B^\top B\boldsymbol{g}(\mathbf{x}), \boldsymbol{g}(\mathbf{x})\rangle \tag{11}$$

for $\boldsymbol{g} \in \mathcal{F}(\mathcal{G}, \mathbb{R}^q)$.

In general, the transport step operator $T_{\mathcal{G}}$ is not an isometry with respect to this norm, as $|||\boldsymbol{g}|||_{\mathcal{G}}^2$ contains mixed quadratic terms like $g_i(\mathbf{x})g_j(\mathbf{x})$ with $i \neq j$. To avoid these terms $B^\top B$ must be diagonal. Comparing (7) with (11) we therefore additionally require $B^\top B$ to be diagonal

$$B^\top B = \text{diag}(b_1, \ldots, b_q) \tag{12}$$

where $\boldsymbol{b} = (b_1, \ldots, b_q)^\top$ also satisfies (6). Observe that the invertibility of B automatically implies that $b_1, \ldots, b_q$ are positive.

Let us summarize the reasoning of this section by the following definition and theorem, which paraphrase corresponding statements in [14] using our notation.

Definition 1. The square matrix $J \in \mathbb{R}^{q\times q}$ is said to have a *pre-stability structure* if there exists an *invertible* matrix $B \in \mathbb{R}^{q\times q}$ and vectors $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^\top \in \mathbb{R}^q$ and $\boldsymbol{b} = (b_1, \ldots, b_q)^\top \in \mathbb{R}^q$ such that

$$\left.\begin{array}{ll}\text{(i)} & BJ = -\text{diag}(\lambda_1, \ldots, \lambda_q)B\\ \text{(ii)} & B^\top B = \text{diag}(b_1, \ldots, b_q)\end{array}\right\}. \tag{13}$$

Moreover, the pre-stability structure becomes a *stability structure* if

$$\lambda_k \in [0, 2] \quad \text{for all } k \in \{1, \ldots, q\}.$$

[4] Each lattice Boltzmann algorithm can be formally derived by discretizing an appropriate finite velocity Boltzmann equation. The operator which then appears in the collision term is J, which has a (heuristically) physical meaning. The collision step operator is introduced as a convenient abbreviation when discretizing the finite velocity Boltzmann equation.

[5] Similarly to (5) and (7) the norm is defined bare of a scaling factor like h^{-d}. Usually, the introduction of such a scaling is reasonable if norms on different grid refinements shall be compared with each other as it is typically done in convergence proofs. However, the associated operator norms are not altered by the scaling factor, wherefore it can be omitted for stability investigations.

Theorem 2. *Let $C = I + J \in \mathbb{R}^{q\times q}$ be the nodal collision step matrix. If J permits a stability structure such that $\boldsymbol{b}$ satisfies the compatibility condition (6) with respect to the transport operator $T_{\mathcal{G}}$, then the lattice Boltzmann algorithm is stable in the weighted ℓ^2-norm defined in (11).*

The above derivation of the stability structure suggests that one should not expect an equivalence between the existence of a stability structure and the stability of the corresponding lattice Boltzmann algorithm. Even for a compatible transport operator, it turns out that the existence of a stability structure is only a sufficient criterion for stability. Examples 5 and 6 in Section 4 illustrate that there are lattice Boltzmann algorithms, indeed, being stable without having any stability structure.

3. General discussion of the pre-stability structure

The definition of the pre-stability structure deserves an abstract discussion in the context of linear algebra. The main point of this section will be Theorem 6 to be applied in the next section.

Considering condition (i) of Definition 1, it is quickly observed that it does not represent anything else than a kind of eigenvalue equation for $J \in \mathbb{R}^{q\times q}$. More precisely, if B and $\boldsymbol{\lambda}$ exist such that $(13)_1$ is satisfied, then J is diagonalizable. Furthermore, the vector $\boldsymbol{\lambda}$ contains the negative eigenvalues of J according to their multiplicity whereas the rows of B correspond to the (transposed) *left eigenvectors*[6] of the matrix J. In so far condition (i), taken alone, is nothing special. Actually, from the viewpoint of linear algebra the *clou* of Definition 1 lies in the second condition. In contrast to the first one, "$B^\top B$ is diagonal" should be interpreted as condition for the column vectors of B which requires them to be pairwise orthogonal. Since the rows and columns of a matrix cannot be controlled independently, it is generally difficult to fulfill condition (i) and (ii) of Definition 1 simultaneously. Therefore it might be easier to be confronted only with conditions referring either to columns or to rows. Replacing (ii) by $BB^\top = \text{diag}$ implies that J has orthogonal left eigenvectors. By normalizing these eigenvectors we obtain a matrix $\tilde{B}$ which satisfies $I = \tilde{B}\tilde{B}^\top = \tilde{B}^\top\tilde{B}$ and diagonalizes J as well. So the modified condition (ii) which involves the row vectors of B like (i), yields a sufficient criterion for the existence of a pre-stability structure. This raises the question about the relation between the respective orthogonality of rows and columns.

In the case of *orthonormality* (which specializes the notion of orthogonality) the situation becomes simple. From the equality of the left and right inverse ensues

$$B^\top B = cI \Leftrightarrow BB^\top = cI \tag{14}$$

for any $c \in \mathbb{R}$. Hence the orthonormality of rows implies the orthonormality of columns and vice versa. In general, however, it is not possible that $B^\top B$ and $BB^\top$ are both diagonal as illustrated by the following proposition.

Proposition 3. *Suppose that $B \in \mathbb{R}^{q\times q}$ is invertible and has a full row or column (where no entry is equal to 0). If one of the matrices $B^\top B$ and $BB^\top$ is diagonal without being a multiple of the identity matrix the other is not diagonal.*

The proof of Proposition 3 is found at the end of the Appendix. Observe that the condition of invertibility does not represent a restriction for our purpose because B is supposed to contain an eigenbasis of J. Furthermore, the condition on B concerning a full column or row is only sufficient but not necessary. Therefore it could be weakened but not completely skipped, to exclude B to be diagonal, where the assertion obviously becomes wrong.

If $J \in \mathbb{R}^{q\times q}$ is symmetric, it is well known that it admits an orthonormal eigenbasis. The corresponding eigenvectors can be joined to obtain an orthonormal matrix that diagonalizes J. Since orthonormal matrices are characterized by satisfying both equations in (14) (for $c = 1$), the existence of a pre-stability structure is expected whenever J is symmetric.

Proposition 4. *Let $J \in \mathbb{R}^{q\times q}$ be symmetric i.e. $J = J^\top$. Then J admits a pre-stability structure in the sense of Definition 1.*

Proof. According to the spectral theorem for symmetric matrices we can find an orthonormal basis of right eigenvectors of J. Let $E \in \mathbb{R}^{q\times q}$ be the matrix whose rows correspond to these eigenvectors. Then the properties of E are summarized by the subsequent equations

$$\left.\begin{array}{ll} JE = E\Lambda & \text{(eigenbasis)} \\ E^\top E = I & \text{(orthonormality)} \end{array}\right\} \tag{15}$$

where Λ is the diagonal matrix formed by the eigenvalues of J. Transposing Eq. $(15)_1$ yields

$$E^\top J = E^\top \Lambda$$

due to the symmetry of J and Λ. As $(15)_2$ is equivalent to $EE^\top = I$, it is straightforward to see that $B := E^\top$ satisfies property (i) and (ii) of Definition 1. ■

[6] In linear algebra one mostly deals with the right eigenvectors of matrices. However, the eigenvalue equation for a matrix $M \in \mathbb{R}^{q\times q}$ can be written in two forms depending on how the matrix shall act on a vector $\boldsymbol{u} \in \mathbb{R}^q$. If $\boldsymbol{u}$ is considered as a column vector, as it is usually done, then $\boldsymbol{u}$ is a right eigenvector if $M\boldsymbol{u} = \lambda\boldsymbol{u}$ and a left eigenvector if $\boldsymbol{u}^\top M = \mu\boldsymbol{u}^\top$ for some $\lambda, \mu \in \mathbb{R}$. Generally, left eigenvectors of M are right eigenvectors of $M^\top$ and vice versa. Since M and $M^\top$ have the same characteristic polynomial and thus the same eigenvalues, it is not necessary to distinguish between left and right eigenvalues.

Since symmetry is quite a specific feature one may ask whether there is a more general condition which guarantees the existence of a pre-stability structure. Answering this question is the goal of the theorem further below. Actually, it turns out that symmetry is intrinsically contained in the (pre-)stability structure (compare with [18] for physical background). Therefore the variety of matrices that dispose of a pre-stability structure is rather restricted. Nevertheless it is possible to interpret the notion of symmetry in a wider sense with respect to more general scalar products.

In order to formulate Theorem 6 let us recall that

$$\mathbb{R}^q \times \mathbb{R}^q \ni (\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle \boldsymbol{x}, \boldsymbol{y} \rangle_S := \langle S\boldsymbol{x}, \boldsymbol{y} \rangle \quad \text{with } S \in \mathbb{R}^{q\times q}$$

defines a scalar product in $\mathbb{R}^q$ if and only if the matrix S is symmetric and positive definite. In this context the following terminology is also quite useful to avoid misunderstandings.

Definition 5. A matrix $M \in \mathbb{R}^{q\times q}$ is called

$$\begin{aligned} &S\text{-symmetric} \quad \text{if } SM = M^\top S, \\ &S\text{-column-orthonormal} \quad \text{if } M^\top SM = I, \\ &S\text{-row-orthonormal} \quad \text{if } MSM^\top = I. \end{aligned}$$

The naming is justified by the following facts: S-symmetry of a matrix M is equivalent to $\langle M\boldsymbol{x}, \boldsymbol{y} \rangle_S = \langle \boldsymbol{x}, M\boldsymbol{y} \rangle_S$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^q$. Furthermore, S-column-orthonormality implies the columns of M to be *orthonormal* with respect to $\langle \cdot, \cdot \rangle_S$ while S-row-orthonormality analogously refers to the rows. Both notions coincide[7] if $MS = SM$. This is particularly the case for $S = I$; then we simply speak of orthonormality.

Theorem 6 (*See also [18]*). *For $A \in \mathbb{R}^{q\times q}$ the following three statements are equivalent:*

(i) *A admits a pre-stability structure (in the sense of Definition 1).*

(ii) *There exists a diagonal and positive definite matrix $D \in \mathbb{R}^{q\times q}$ such that*

$$A_D := D^{-1}AD \text{ is } D\text{- symmetric} \Leftrightarrow A \text{ is } D^{-1}\text{-symmetric}. \tag{16}$$

(iii) *There exists a diagonal and positive definite matrix $D \in \mathbb{R}^{q\times q}$ such that*

$$AD = DA^\top. \tag{17}$$

The proof of the theorem is given in the Appendix. Note with regard to (ii) that any diagonal matrix is symmetric, thus in particular $D = D^\top$. Being positive definite, D therefore generates a scalar product, so that the statement of (ii) is reasonable. The following points should be also observed:

- Due to the equivalence between (i) and (ii) Theorem 6 is the ultimate extension of Proposition 4. The requirement of symmetry in Proposition 4 is generalized by the theorem but still remains rather specific.
- (17) expresses the fact that AD and $D^{-1}A$ are symmetric[8] with respect to the standard scalar product. Hence A admits the representation $A = DS$ with the symmetric matrix $S := D^{-1}A$ (cf. (25) and the Onsager-like relation mentioned in [18]).
- Any matrix A satisfying one of the three conditions must be diagonalizable.

The importance of Theorem 6 lies in the equivalence statement that considerably simplifies the decision whether a given matrix A admits a stability structure or not. The theorem relegates the original problem (concerning the existence of a stability structure) to the linear matrix equation (17), which has the advantage to be more handy.[9] Thus the search for a stability structure gets basically transformed into the search for a specific solution of a certain linear equation.

For $M, N \in \mathbb{R}^{q\times q}$ it can be shown (see [20]) that the homogeneous matrix equation

$$MX - XN = 0$$

has a nonsingular solution $X \in \mathbb{R}^{q\times q}$ if and only if M and N are similar. As any (diagonalizable) matrix $A \in \mathbb{R}^{q\times q}$ is similar to its transposed $A^\top$, the equation

$$AX - XA^\top = 0$$

allows for a nonsingular solution. However, with regard to Theorem 6 we are only interested in solutions $X = D$ being diagonal and positive definite. Of course, this additional constraint cannot be satisfied generally which becomes also clear if

[7] If a matrix M is both S-column- and S-row-orthonormal, then $M^\top SM = MSM^\top = I$ which implies $(M^\top)^{-1} = SM = MS$ (due to the equality of the left and the right inverse). Hence M and S commute. Now from S-column-orthonormality follows $M^\top MS = I$ whereas S-row-orthonormality yields $SMM^\top = I$. Exploiting the equality of the left and the right inverse again, we conclude that $M^\top M = MM^\top$. Therefore it is a necessary condition for the coincidence of S-column- and S-row-orthonormality that M is *normal*. Let us finally remark that a matrix M commutes with a symmetric matrix S if and only if $M^\top$ commutes with S too.

[8] D and D^{-1} represent a *right* and a *left symmetrizer* for A.

[9] By means of the Kronecker product linear matrix equations can be reformulated into linear equations of standard form.

$AD - DA^\top$ is considered elementwise. Thereby we end up with an overdetermined linear system consisting of q^2 equations where only q unknowns are available corresponding to the q diagonal elements of D. Having much more equations than variables, it is quite probable that the system contains contradictory equations if the matrix A is chosen arbitrarily. Then the existence of any solution would be excluded. In order to avoid this problem A should be of particularly simple but not too simple structure (for instance if A is diagonal then $AD - DA^\top = 0$ for any other diagonal matrix). In opposition to general $q \times q$ matrices, rank-one matrices of the same size are already completely characterized by $2q$ numbers as being the dyadic product of two vectors. Since rows and columns of rank-one matrices are respective multiples of each other, the resulting q^2 equations are related to each other in the same way thus impeding them to conflict with one another. Let us make these things more precise by the next proposition.

Proposition 7. *Let $A \in \mathbb{R}^{q\times q}$ be a rank-one matrix such that there exist two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^q \setminus \{\boldsymbol{0}\}$ with*

$$A\boldsymbol{x} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^q. \tag{18}$$

Then there is a diagonal matrix $D \neq 0$ satisfying

$$AD = DA^\top \tag{19}$$

apart from the zero matrix. Furthermore D can be chosen positive definite if $\operatorname{sgn} u_i = \operatorname{sgn} v_i$ for all $i \in \{1, \dots, q\}$ whereas D becomes unique up to multiples if $u_i \neq 0$ for all $i \in \{1, \dots, q\}$.

Proof. Let us assume that the assertion is true in order to derive conditions which permit us to construct D in dependence of $\boldsymbol{u}$ and $\boldsymbol{v}$. Since $A^\top \boldsymbol{x} = \langle \boldsymbol{v}, \boldsymbol{x}\rangle \boldsymbol{u}$ for all $\boldsymbol{x} \in \mathbb{R}^q$, we get

$$AD\boldsymbol{x} = DA^\top \boldsymbol{x} \Leftrightarrow \langle \boldsymbol{u}, D\boldsymbol{x}\rangle \boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{x}\rangle D\boldsymbol{u}.$$

As $D = D^\top$, thanks to the diagonality, this becomes

$$\langle D\boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{x}\rangle D\boldsymbol{u}. \tag{20}$$

The equality implies that $D\boldsymbol{u}$ is in the span of $\boldsymbol{v}$ and vice versa. Hence there is $\lambda \in \mathbb{R}$ such that $\lambda\boldsymbol{v} = D\boldsymbol{u}$. To determine D, two cases have to be taken into account: $\lambda = 1 \neq 0$ with $\boldsymbol{v} = D\boldsymbol{u}$ and $\lambda = 0$ with $\boldsymbol{0} = D\boldsymbol{u}$. Observe that the general case $\lambda \neq 0$ is contained in the first one, since (20) only defines D up to a scalar factor. So one may consider $\tilde{D} := \lambda^{-1} D$ instead of D.

Case 1: $\forall i \in \{1, \dots, q\} : u_i = 0 \Rightarrow v_i = 0$

If we set

$$d_i := \begin{cases} v_i/u_i & \text{if } u_i \neq 0 \\ 1\,(\text{or arbitrary value}) & \text{if } u_i = 0 \end{cases} \text{for } i \in \{1, \dots, q\}$$

then $D := \operatorname{diag}(\boldsymbol{d})$ evidently satisfies $\boldsymbol{v} = D\boldsymbol{u}$ thus ensuring the validity of (20) and (19). It is clearly seen that D becomes positive definite or uniquely defined under the conditions mentioned in the proposition.

Case 2: $\exists k \in \{1, \dots, q\} : u_k = 0$ but $v_k \neq 0$

By setting

$$d_i := \begin{cases} 0 & \text{if } u_i \neq 0 \\ 1\,(\text{or arbitrary value}) & \text{if } u_i = 0 \end{cases} \text{for } i \in \{1, \dots, q\}$$

we achieve that $D := \operatorname{diag}(\boldsymbol{d})$ satisfies $D\boldsymbol{u} = \boldsymbol{0}$ which also entails (20) and (19). Observe that $D \neq 0$ because we have $d_k = 1$ at least. However, in this case, D can neither become positive definite nor uniquely defined. ■

Case 2 is considered as degenerated while case 1 represents the standard situation. With the convention $0/0 = 1$ the definition of D is more compactly written in the form

$$D = \operatorname{diag}(\boldsymbol{v}/\boldsymbol{u})$$

where / here indicates the componentwise division of the vectors $\boldsymbol{v}$ and $\boldsymbol{u}$.

Generally, every matrix can be written as sum of rank-one matrices. So there exist for any given $A \in \mathbb{R}^{q\times q}$ two sets of vectors $\boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(r)} \in \mathbb{R}^q$ and $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(r)} \in \mathbb{R}^q$ such that

$$A\boldsymbol{x} = \sum_{\ell=1}^{r} \langle \boldsymbol{u}^{(\ell)}, \boldsymbol{x}\rangle \boldsymbol{v}^{(\ell)} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^q,$$

where the rank of A is equal to r whenever $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(r)} \in \mathbb{R}^q$ are linearly independent. A sufficient condition for

$$AD\boldsymbol{x} = DA^\top \boldsymbol{x} \Leftrightarrow \sum_{\ell=1}^{r} \langle D\boldsymbol{u}^{(\ell)}, \boldsymbol{x}\rangle \boldsymbol{v}^{(\ell)} = \sum_{\ell=1}^{r} \langle \boldsymbol{v}^{(\ell)}, \boldsymbol{x}\rangle D\boldsymbol{u}^{(\ell)}$$

with some diagonal matrix D is

$$\operatorname{diag}(\boldsymbol{v}^{(1)}/\boldsymbol{u}^{(1)}) = \cdots = \operatorname{diag}(\boldsymbol{v}^{(r)}/\boldsymbol{u}^{(r)}) = D. \tag{21}$$

However, it must be emphasized that this condition is not necessary. In order to see this let us consider the following example. Setting

$$A\boldsymbol{x} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} + \langle \boldsymbol{v}, \boldsymbol{x}\rangle \boldsymbol{u} \quad \text{such that } A^\top \boldsymbol{x} = \langle \boldsymbol{v}, \boldsymbol{x}\rangle \boldsymbol{u} + \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} = A\boldsymbol{x}$$

for $\boldsymbol{x} \in \mathbb{R}^q$ and fixed $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^q$ defines a symmetric matrix A which satisfies (17) with $D = I$. Now this does by no means oblige $\boldsymbol{u}$ and $\boldsymbol{v}$ to satisfy $\operatorname{diag}(\boldsymbol{v}/\boldsymbol{u}) = \operatorname{diag}(\boldsymbol{u}/\boldsymbol{v}) = I$. Nevertheless, as A is symmetric there exists an orthogonal eigenbasis. Therefore it is possible to find two vectors $\boldsymbol{e}^{(1)}$ and $\boldsymbol{e}^{(2)}$ such that $A\boldsymbol{x} = \langle \boldsymbol{e}^{(1)}, \boldsymbol{x}\rangle \boldsymbol{e}^{(1)} \pm \langle \boldsymbol{e}^{(2)}, \boldsymbol{x}\rangle \boldsymbol{e}^{(2)}$ and $\operatorname{diag}(\boldsymbol{e}^{(1)}/\boldsymbol{e}^{(1)}) = \operatorname{diag}(\boldsymbol{e}^{(2)}/\boldsymbol{e}^{(2)}) = I$.

Let us finally remark that for symmetric A searching a diagonal matrix D complying with (17) means the same as seeking for a diagonal matrix commuting with A. Supposing that A is also *irreducible*, it can be shown that multiples of the identity I are the only diagonal matrices which satisfy $AD = DA$. This permits the following conclusion: If there is for some arbitrary A a diagonal matrix D such that (17) holds and AD is *irreducible* then D is uniquely determined (up to scalar factors).

4. Application to lattice Boltzmann collision operators

In this section we are going to apply the results of the preceding one to prove the existence of stability structures for some frequent and representative types of linear lattice Boltzmann collision operators.

Let us start considering collision operators of the BGK type. Their typical structure is given by

$$J = \omega(G - I) \tag{22}$$

where $\omega \in \mathbb{R}$ is the relaxation parameter (collision frequency) and G denotes the equilibrium operator. Here we assume that G is a (linear) projector thus satisfying $G^2 = G$. Replacing the scalar ω by a relaxation matrix Ω

$$J = \Omega(G - I) \tag{23}$$

generalizes (22) to become a so-called MRT[10] collision operator [21,22]. Usually, the relaxation matrix Ω is required to have the following properties

(I) Ω is *symmetric*, i.e. $\Omega = \Omega^\top$ and *positive semidefinite*.
(II) The (left) nullspace of Ω corresponds to the span of those vectors which generate the moments to be conserved (e.g. $\boldsymbol{e} = (1, 1, \ldots, 1) \in \mathbb{R}^q$ generating the mass moment). Basically this means that Ω and $G - I$ have the same left nullspace.[11]

According to Theorem 6 the collision operator J admits a pre-stability structure if there is a positive definite diagonal matrix D such that

$$JD = DJ^\top \Leftrightarrow \begin{cases} \text{BGK:} & GD = DG^\top \\ \text{MRT:} & \Omega GD - \Omega D = DG^\top \Omega - D\Omega. \end{cases} \tag{24}$$

In many cases the equilibrium matrix G evidently displays the following product structure (cf. Onsager-like relation in [18])

$$G = WS \tag{25}$$

where $W \in \mathbb{R}^{q\times q}$ is a positive definite diagonal matrix and $S = S^\top$ is symmetric. In particular, an equilibrium complies with (25) if it is represented or representable in the form

$$G\boldsymbol{f} = \sum_{i=1}^{m} \gamma_i \langle \boldsymbol{f}, \boldsymbol{s}_i\rangle \boldsymbol{s}_i \boldsymbol{w} \quad \text{for } \boldsymbol{f} \in \mathbb{R}^q \tag{26}$$

with $\gamma_i \in \mathbb{R}$, $\boldsymbol{s}_i \in \mathbb{R}^q$ for $i \in \{1, \ldots, m\}$, the weight vector $\boldsymbol{w} \in \mathbb{R}^q$ and with $\boldsymbol{s}_i\boldsymbol{w} \in \mathbb{R}^q$ denoting the componentwise product. This can easily be seen by setting

$$W = \operatorname{diag}(\boldsymbol{w}) \quad \text{and} \quad S = \sum_{i=1}^{m} \gamma_i \boldsymbol{s}_i \otimes \boldsymbol{s}_i \tag{27}$$

[10] Multiple Relaxation Time.

[11] This condition has a physical background. Moments like $\langle \boldsymbol{e}, \boldsymbol{f}\rangle$, that are obtained by taking the scalar product of the population function $\boldsymbol{f}$ with certain moment generating vectors, shall represent conserved physical quantities as for instance the mass density. In order to derive the desired, physically reasonable conservation equations for these quantities from the lattice Boltzmann equation, it is necessary that the corresponding moments of the collision term vanish, e.g. $\langle \boldsymbol{e}, J\boldsymbol{f}\rangle = 0$ for all $\boldsymbol{f} \in \mathbb{R}^q$. But this implies that the associated moment generating vectors like $\boldsymbol{e}$ must belong to the left kernel of J: $\boldsymbol{e}^\top J = 0$. In the BGK case $\boldsymbol{m} \in \mathbb{R}^q$ generates a conserved moment if $\langle \boldsymbol{m}, G\boldsymbol{f}\rangle = \langle \boldsymbol{m}, \boldsymbol{f}\rangle$ for all $\boldsymbol{f} \in \mathbb{R}^q$. As Ω is supposed to be symmetric it need not distinguish between the left and the right nullspace.

such that $\boldsymbol{s}_i\boldsymbol{w} = W\boldsymbol{s}_i$ and $S\boldsymbol{f} = \sum_{i=1}^{m} \gamma_i \langle \boldsymbol{f}, \boldsymbol{s}_i\rangle \boldsymbol{s}_i$ for all $\boldsymbol{f} \in \mathbb{R}^q$. In fact, W is positive definite as the weight vector $\boldsymbol{w}$ is supposed to have positive components only. Furthermore S is symmetric as a linear combination of symmetric rank-one mappings. Let us additionally remark in this context that the $\boldsymbol{s}_i$'s usually are W-orthogonal ($\langle W\boldsymbol{s}_i, \boldsymbol{s}_j\rangle = 0$ for $i \neq j$) and satisfy $\gamma_i^2 \langle W\boldsymbol{s}_i, \boldsymbol{s}_i\rangle = 1$. In this case G defines a projection.

Thanks to Theorem 6 the existence of a stability structure follows without any effort, if the equilibrium is a projector satisfying (25).

Proposition 8. *Assume that the equilibrium matrix attains the form $G = WS$ with S being symmetric and the positive definite diagonal matrix W. Then the associated BGK collision matrix $J = \omega(G - I)$ has a pre-stability structure. Moreover, if G is a projection, then J admits a stability structure if and only if $0 \leq \omega \leq 2$.*

Proof. Clearly G is W^{-1}-symmetric,or equivalently, G satisfies $(24)_1$ with $D = W$. Since W is supposed to be diagonal and positive definite there is a pre-stability structure according to Theorem 6. Furthermore, if G is a projection then so is $I - G$. Therefore $I - G$ admits only 0 and 1 as eigenvalues which entails that $J = \omega(G - I)$ can only have the eigenvalues 0 and $-\omega$. Hence the negative eigenvalues of J are in $[0, 2]$ – as required for a stability structure – if and only if $\omega \in [0, 2]$. ■

In the sequel we shortly present some examples of equilibrium matrices, which obviously comply with the product structure in (25). In all three cases the equilibrium matrix also represents a projection so that Proposition 8 is fully applicable.

Example 1. In hyperbolic scaling the D1Q3 lattice Boltzmann algorithm leads to a discretization of the one-dimensional wave equation ([13] Section 2.1) if the BGK collision operator is combined with the equilibrium matrix defined by

$$G\boldsymbol{f} = [\langle \boldsymbol{e}, \boldsymbol{f}\rangle \boldsymbol{e} + \theta \langle \boldsymbol{c}, \boldsymbol{f}\rangle \boldsymbol{c}]\,\boldsymbol{w} \quad \text{for}\, \boldsymbol{f} \in \mathbb{R}^3 \tag{28}$$

where $\boldsymbol{e} := (1, 1, 1)^\top$, $\boldsymbol{c} := (-1, 1, 0)^\top$, $\boldsymbol{w} := (\frac{1}{2\theta}, \frac{1}{2\theta}, \frac{\theta-1}{\theta})^\top$ and the multiplication of vectors is understood componentwise (see notational remark on page 3). Note that $\theta > 1$ is necessary to obtain a weight vector $\boldsymbol{w}$ with positive components and thus a positive definite W. While the relaxation frequency ω becomes a purely algorithmic parameter, $\sqrt{\frac{1}{\theta}}$ plays the role of the corresponding propagation speed occurring in the wave equation.

Example 2. To discretize the diffusion equation in two space dimensions using a D2Q4 velocity model (and parabolic scaling) one employs the equilibrium matrix G defined by

$$G\boldsymbol{f} = \frac{1}{4}\langle \boldsymbol{e}, \boldsymbol{f}\rangle \boldsymbol{e} \quad \text{for}\, \boldsymbol{f} \in \mathbb{R}^4. \tag{29}$$

Here G is even symmetric since the weight vector $\boldsymbol{w} := \frac{1}{4}\boldsymbol{e}$ with $\boldsymbol{e} := (1, 1, 1, 1)^\top$ is uniform by reason of isotropy and hence $W = \frac{1}{4}I$. Observe that the weight vector of the D2Q5 model, which ensues from the D2Q4 model by adding a rest population, is not uniform in $\mathbb{R}^5$ wherefore the corresponding equilibrium matrix is not symmetric but still of the form (25). In contrast to example 1 the relaxation frequency ω bears a "macroscopic" meaning as it determines the diffusivity: $\nu = \frac{1}{2}(\frac{1}{\omega} - \frac{1}{2})$.

Example 3. Certainly the most prominent equilibrium to be presented here is given by

$$G\boldsymbol{f} = \left[\langle \boldsymbol{e}, \boldsymbol{f}\rangle \boldsymbol{e} + \frac{1}{3}\langle \boldsymbol{c}_x, \boldsymbol{f}\rangle \boldsymbol{c}_x + \frac{1}{3}\langle \boldsymbol{c}_y, \boldsymbol{f}\rangle \boldsymbol{c}_y\right]\boldsymbol{w} \quad \text{for}\, \boldsymbol{f} \in \mathbb{R}^9 \tag{30}$$

where

$$\boldsymbol{e} := \begin{pmatrix}1\\1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{w} := \begin{pmatrix}4/9\\1/9\\1/9\\1/9\\1/9\\1/36\\1/36\\1/36\\1/36\end{pmatrix}, \quad \boldsymbol{c}_x := \begin{pmatrix}0\\1\\0\\-1\\0\\1\\-1\\-1\\1\end{pmatrix}, \quad \boldsymbol{c}_y := \begin{pmatrix}0\\0\\1\\0\\-1\\1\\1\\-1\\-1\end{pmatrix}. \tag{31}$$

In combination with the D2Q9 BGK lattice Boltzmann algorithm (again with parabolic scaling) this leads to a discretization of the Stokes equation. Similarly to the preceding example ω is related to the viscosity: $\nu = \frac{1}{3}(\frac{1}{\omega} - \frac{1}{2})$.

4.1. Construction of MRT collision operators with stability structure

The authors of [14] investigate the existence of stability structures if the D2Q9 equilibrium (31) for the Stokes equation is combined with MRT collision operators. More exactly, by means of this specific example it is described how an MRT collision operator can be constructed which admits a stability structure. This subsection has the purpose to make this approach more clear and concise as well as to distill it into a more abstract context that enables further applications.

Proposition 9 (*Construction of MRT Collision Operators*). *Let the equilibrium matrix $G \in \mathbb{R}^{q\times q}$ be given by (25) and let $P_{\mathcal{K}}$ denote the orthogonal projector onto the left nullspace of $G - I$. Assume that for $p \le q$ there exists a family $(P_i)_{i\in\{1,\dots,p\}} \subset \mathbb{R}^{q\times q}$ of symmetric projector matrices ($P_i = P_i^\top$) satisfying for all $i, j \in \{1, \dots, p\}$:*

- $P_iP_j = \delta_{ij}P_i$ *(mutual orthogonality)*
- $\sum_{i=1}^{p} P_i = I$ *(completeness)*
- $P_iP_{\mathcal{K}} = P_{\mathcal{K}}P_i$
- $P_iS = SP_i$
- $P_iW = WP_i$.

If the relaxation matrix $\Omega \in \mathbb{R}^{q\times q}$ is of the form

$$\Omega = \sum_{i=1}^{p} \mu_i P_i (I - P_{\mathcal{K}}) \quad \text{with } \mu_i \ge 0 \text{ for } i \in \{1, \dots, p\} \tag{32}$$

then the following statements hold true:

(1) *Ω satisfies property (I) and (II) in Section 4; in particular Ω is symmetric and positive semidefinite.*
(2) *The MRT collision matrix $J = \Omega(G - I)$ satisfies (24) with $D = W^{-1}$ which means that J admits a pre-stability structure.*
(3) *Moreover, if G represents a projection, the above pre-stability structure of $J = \Omega(G - I)$ turns into a stability structure if and only if $0 \le \mu_i \le 2$ for $i = \{1, \dots, p\}$ in Eq. (32).*

Proof. ad (1) Generally, the product of two symmetric matrices is again a symmetric matrix if the factor matrices commute with each other. By hypothesis these conditions are satisfied by $P_{\mathcal{K}}$ and each of the P_i's, as orthogonal projections are symmetric (and vice versa). Therefore Ω is the sum of symmetric matrices and thus symmetric itself. Due to its symmetry Ω is diagonalizable with the eigenvalues $\mu_1, \dots, \mu_p$ (see Lemma 15). Hence Ω is positive semidefinite if the μ_i's are non-negative.

ad (2) Let us first establish the following representation of the collision operator:

$$J = \sum_{i=1}^{p} \mu_i P_i (G - I). \tag{33}$$

Actually, this follows if we can show that

$$P_{\mathcal{K}}(G - I) = 0. \tag{34}$$

Since $P_{\mathcal{K}}$ corresponds to the orthogonal projection onto the left nullspace $\mathcal{K}$ (left kernel) of $G - I$, which is defined as

$$\begin{aligned} \mathcal{K} \equiv \text{left}\ker(G - I) &:= \left\{\boldsymbol{z} \in \mathbb{R}^q : \boldsymbol{z}^\top (G - I) = \boldsymbol{0}^\top\right\} \\ &= \left\{\boldsymbol{z} \in \mathbb{R}^q : \langle \boldsymbol{z}, (G - I)\boldsymbol{y}\rangle = 0 \text{ for all } \boldsymbol{y} \in \mathbb{R}^q\right\}, \end{aligned}$$

we have $P_{\mathcal{K}}\boldsymbol{x} \in \mathcal{K}$ for all $\boldsymbol{x} \in \mathbb{R}^q$. So we get

$$\langle P_{\mathcal{K}}\boldsymbol{x}, (G - I)\boldsymbol{y}\rangle = \langle \boldsymbol{x}, P_{\mathcal{K}}^\top (G - I)\boldsymbol{y}\rangle = 0 \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^q$$

enforcing $P_{\mathcal{K}}^\top(G - I) = 0$. With $P_{\mathcal{K}} = P_{\mathcal{K}}^\top$ we obtain (34). The equation $WJ^\top = JW$ is verified by the subsequent computation taking (33) and $G = WS$ into account:

$$\begin{aligned} WJ^\top &= W \sum_i \mu_i (WS - I)^\top P_i^\top \\ &= W \sum_i \mu_i (SW - I)P_i | \text{ symmetry of the } P_i\text{'s and } S \\ &= W \sum_i \mu_i P_i (SW - I) | \text{ commutativity of the } P_i\text{'s with } S \text{ and } W \\ &= \sum_i \mu_i P_i W (SW - I) | \text{ again commutativity of the } P_i\text{'s with } W \end{aligned}$$

$$
\begin{aligned}
&= \sum_i \mu_i P_i (WSW - W) \\
&= \sum_i \mu_i P_i (WS - I) W \\
&= JW.
\end{aligned}
$$

Hence the statement (ii) of Theorem 6 is true with respect to J and W which yields the existence of a pre-stability structure.

ad (3) The assertion becomes clear if $-\mu_1, \ldots, -\mu_p$ turn out to be the eigenvalues of J. To show this let us remark that each $\tilde{P}_i := P_i(I - G)$ with $i \in \{1, \ldots, p\}$ represents a projection. Indeed, we have

$$
\begin{aligned}
\tilde{P}_i^2 &= P_i(I-G)P_i(I-G) = P_i^2(I-G)^2 \\
&= P_i(I - 2G + G^2) = P_i(I-G) \\
&= \tilde{P}_i
\end{aligned}
$$

where it has been exploited that P_i and G are commuting projections. Now we conclude from Lemma 15 that the eigenvalues of $-J = \sum_{i=1}^p \mu_i \tilde{P}_i$ are just the μ_i's. Then, however, J attains the desired eigenvalues. ■

If the equilibrium is given by a projection satisfying (25), then – according to the previous proposition – the construction of MRT collision matrices with stability structure essentially reduces to the choice of an appropriate family $(P_i)_{i\in\{1,\ldots,p\}}$ of projections. Here the major difficulty lies in fulfilling the three commutativity relations required in the hypothesis of Proposition 9. By mimicking the approach pursued in [14] the commutativity of the P_i's with $P_{\mathcal{K}}$ and S may be rather easily obtained.

Lemma 10. *Let $\mathcal{U}$ be a subspace of $\mathbb{R}^q$ with the basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ and let $P_{\mathcal{U}}$ be the orthogonal projection onto $\mathcal{U}$. Furthermore let $(P_i)_{i\in\{1,\ldots,p\}}$ be a family of symmetric, mutually orthogonal projections. Assume that for any $k \in \{1, \ldots, m\}$ there is an index $i_k \in \{1, \ldots, p\}$ such that*

$$P_{i_k} \boldsymbol{u}_k = \boldsymbol{u}_k.$$

Then $P_{\mathcal{U}}$ commutes with all P_i's, i.e.

$$P_i P_{\mathcal{U}} = P_{\mathcal{U}} P_i \quad \text{for all } i \in \{1, \ldots, p\}.$$

Proof. Let us start with two convenient remarks. Firstly, recall that $P_{\mathcal{U}}$ acts like the identity on $\mathcal{U}$ while its orthogonal complement $\mathcal{U}^\perp$ is mapped to $\mathbf{0}$ owing to the fact that $P_{\mathcal{U}}$ is an orthogonal projection. Secondly, as $\mathbb{R}^q = \mathcal{U} \oplus \mathcal{U}^\perp$ is the direct sum of $\mathcal{U}$ and $\mathcal{U}^\top$, it is sufficient to verify that

$$P_i P_{\mathcal{U}} \boldsymbol{v} = P_{\mathcal{U}} P_i \boldsymbol{v} \tag{35}$$

for all $i \in \{1, \ldots, p\}$ and for all v being either in $\mathcal{U}$ or $\mathcal{U}^\perp$.

Case 1: "$\boldsymbol{v} \in \mathcal{U}$". Since $\mathcal{U}$ is spanned by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ it is enough to show that (35) holds for these basis vectors. To do this let us fix some $k \in \{1, \ldots, m\}$. Then (35) is obviously satisfied for $\boldsymbol{v} = \boldsymbol{u}_k$ and $i = i_k$ since $P_{i_k}\boldsymbol{u}_k = P_{\mathcal{U}}\boldsymbol{u}_k = \boldsymbol{u}_k$. Now let $i \neq i_k$ such that $\delta_{ii_k} = 0$. Exploiting the mutual orthogonality of the P_i's we get

$$P_i \boldsymbol{u}_k = P_i P_{i_k} \boldsymbol{u}_k = \delta_{ii_k} P_{i_k} \boldsymbol{u}_k = \delta_{ii_k} \boldsymbol{u}_k = 0 \boldsymbol{u}_k = \mathbf{0}.$$

Therefore $P_i P_{\mathcal{U}} \boldsymbol{u}_k = \mathbf{0} = P_{\mathcal{U}} P_i \boldsymbol{u}_k$ which verifies (35) also for $\boldsymbol{v} = \boldsymbol{u}_k$ and $i \neq i_k$.

Case 2: "$\boldsymbol{v} \in \mathcal{U}^\perp$". We first demonstrate that the P_i's leave

$$\mathcal{U}^\perp = \left\{ \boldsymbol{v} \in \mathbb{R}^q : \langle \boldsymbol{v}, \boldsymbol{u}_k \rangle = 0 \text{ for } k = 1, \ldots, m \right\}$$

invariant which means $P_i \mathcal{U}^\perp \subset \mathcal{U}^\perp$. So we must verify that

$$\langle P_i \boldsymbol{v}, \boldsymbol{u}_k \rangle = 0 \tag{36}$$

for all $\boldsymbol{v} \in \mathcal{U}^\perp$ and all $k \in \{1, \ldots, m\}$. Thanks to the symmetry of the P_i's (36) is equivalent to $\langle \boldsymbol{v}, P_i^\top \boldsymbol{u}_k \rangle = 0$. From case 1 we can conclude that either $P_i^\top \boldsymbol{u}_k = \boldsymbol{u}_k$ or $P_i^\top \boldsymbol{u}_k = \mathbf{0}$. Therefore the right hand side of (36) becomes either $\langle \boldsymbol{v}, \boldsymbol{u}_k \rangle$ or $\langle \boldsymbol{v}, \mathbf{0} \rangle$, where the latter expression vanishes trivially while the first one gets zero by the defining property of $\mathcal{U}^\perp$. ■

Let us assume that the equilibrium G is a projector and fits into the general form of Eq. (26), with mutually orthogonal $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$. Then the left nullspace of $G - I$ corresponds to $\mathrm{span}(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m)$. To obtain an MRT collision operator, a relaxation matrix $\Omega \in \mathbb{R}^q$ can be constructed along the following steps:

- Normalize $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ and complement the resulting vectors to obtain an orthonormal basis $\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_q$ in $\mathbb{R}^q$.

- Group the $\tilde{\boldsymbol{s}}_1, \ldots, \tilde{\boldsymbol{s}}_q$ in $\mathbb{R}^q$ into p different W-invariant $\mathcal{W}_1, \ldots, \mathcal{W}_p$ subspaces with $p \le q$, where each subspace is spanned by some of the $\tilde{\boldsymbol{s}}_i$'s and each of the $\boldsymbol{s}_k$'s that belongs only to one of the subspaces. Then $\mathbb{R}^q$ is obtained as the direct sum of the subspaces, i.e. $\mathbb{R}^q = \mathcal{W}_1 \oplus \cdots \oplus \mathcal{W}_p$. In particular $\mathcal{W}_i \cap \mathcal{W}_j = \{\boldsymbol{0}\}$ for $i \neq j$.

Usually the second step may be a bit tricky as no precise recipe for the construction of the subspaces $\mathcal{W}_1, \ldots, \mathcal{W}_p$ can be given. But it might be reasonable to consider for each $\tilde{\boldsymbol{s}}_i$, $i \in \{1, \ldots, q\}$ the smallest W-invariant subspace $\tilde{\mathcal{W}}_i := \{\tilde{\boldsymbol{s}}_i, W\tilde{\boldsymbol{s}}_i, W^2\tilde{\boldsymbol{s}}_i, \ldots\}$ which contains $\tilde{\boldsymbol{s}}_i$. These spaces cannot have all a pairwise trivial intersection except if $W\tilde{\boldsymbol{s}}_i = \tilde{\boldsymbol{s}}_i$ for all $i \in \{1, \ldots, q\}$. However, if $\tilde{\mathcal{W}}_i \cap \tilde{\mathcal{W}}_j \neq \{\boldsymbol{0}\}$ then (usually) $\tilde{\mathcal{W}}_i = \tilde{\mathcal{W}}_j$. This can be seen by the following argument: If there is a common nonzero vector of $\mathcal{W}_i$ and $\mathcal{W}_j$ we can find two polynomials p_i and p_j such that $p_i(W)\tilde{\boldsymbol{s}}_i = p_j(W)\tilde{\boldsymbol{s}}_j$. Let us now additionally assume that p_i does not vanish for one of the eigenvalues (diagonal elements) of W. Then $p_i(W)$ is invertible. As $p_i(W)^{-1}$ has the same eigenspaces as $p_i(W)$ and thus as W, there is another polynomial[12] q_i with $p_i(W)^{-1} = q_i(W)$. So it follows $\tilde{\boldsymbol{s}}_i = q_i(W)p_j(W)\tilde{\boldsymbol{s}}_j$ which shows that $\tilde{\boldsymbol{s}}_i \in \tilde{\mathcal{W}}_j$ and therefore $\tilde{\mathcal{W}}_i \subset \tilde{\mathcal{W}}_j$. If p_j also has no zeros at the eigenvalues of W, then we conversely infer that $\tilde{\mathcal{W}}_i \supset \tilde{\mathcal{W}}_j$ which yields the asserted equality. Hence one should take for $\mathcal{W}_1, \ldots, \mathcal{W}_p$ a maximal number of $\tilde{\mathcal{W}}_i$'s such that the chosen $\tilde{\mathcal{W}}_i$'s are mutually disjoint apart from $\boldsymbol{0}$. It may happen that there is one $i \in \{1, \ldots, q\}$ such that $\tilde{\mathcal{W}}_i = \mathbb{R}^q$. In this case it is not possible to construct a non-trivial MRT collision operator by this approach.

If $1 < p \le q$ invariant subspaces $\mathcal{W}_1, \ldots, \mathcal{W}_p$ are found whose direct sum is $\mathbb{R}^q$, then the associated family of orthogonal projectors $P_1, \ldots, P_p$ onto these subspaces has the properties required in Proposition 9. Indeed, the mutual orthogonality and the commutativity with $P_{\mathcal{K}}$ and S directly ensue from the fact, that each P_k, $k \in \{1, \ldots, p\}$ is the sum of some one-dimensional orthogonal projectors associated with the $\tilde{\boldsymbol{s}}_i$'s ($\boldsymbol{x} \mapsto \langle \tilde{\boldsymbol{s}}_i, \boldsymbol{x} \rangle \tilde{\boldsymbol{s}}_i$). The commutativity with W results from the subsequent lemma.

Lemma 11. *Let $(P_i)_{i\in\{1,\ldots,p\}}$ be a complete family of mutually orthogonal projectors such that $P_iP_j = \delta_{ij}P_i$ and $\sum_{i=1}^p P_i = I$. If the image of each P_i is invariant under W, i.e. $W\mathrm{im}(P_i) \subset \mathrm{im}(P_i)$ for $i \in \{1, \ldots, p\}$, then*

$$P_iW = WP_i \quad \textit{for all } i \in \{1, \ldots, p\}.$$

Proof. Let $i \in \{1, \ldots, p\}$ be fixed. Let us assume that W does not only leave invariant the image but also the kernel (nullspace) of P_i. Then we have $P_iW\boldsymbol{x} = WP_i\boldsymbol{x}$ for every $\boldsymbol{x} \in \mathrm{im}(P_i) \cup \ker(P_i)$. Being a projector it follows that $\mathrm{im}(P_i) \oplus \ker(P_i) = \mathbb{R}^q$. Hence $P_iW\boldsymbol{x} = WP_i\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^q$ which just means that W and P_i commute.

The completeness and mutual orthogonality imply

$$\ker(P_i) = \bigoplus_{j=1, j\neq i}^{p} \mathrm{im}(P_j).$$

Now the W-invariance of each $\mathrm{im}(P_j)$ entails the W-invariance of the direct sum and thus of $\ker(P_i)$ which justifies the above assumption. ■

Example 4. For illustration let us consider the D1Q3 case. The equilibrium (28) in example 1 contains the orthogonal vectors $\boldsymbol{e}, \boldsymbol{c} \in \mathbb{R}^q$ spanning the left nullspace of G. By normalizing these vectors we obtain $\tilde{\boldsymbol{s}}_1$ and $\tilde{\boldsymbol{s}}_2$, while the unit vector $\tilde{\boldsymbol{s}}_3$ is chosen orthogonal to $\boldsymbol{e}$ and $\boldsymbol{c}$

$$\tilde{\boldsymbol{s}}_1 := \frac{1}{\sqrt{3}}\boldsymbol{e} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \qquad \tilde{\boldsymbol{s}}_2 := \frac{1}{\sqrt{2}}\boldsymbol{c} = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\\0\end{pmatrix}, \qquad \tilde{\boldsymbol{s}}_3 := \frac{1}{\sqrt{6}}(3\boldsymbol{c}^2 - 2\boldsymbol{e}) = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix},$$

so that $(\tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_2, \tilde{\boldsymbol{s}}_3)$ form an orthonormal basis of $\mathbb{R}^3$. It is convenient to express the weight vector $\boldsymbol{w}$ in terms of $\boldsymbol{e}$ and $\boldsymbol{c}$:

$$\boldsymbol{w} = \begin{pmatrix}1/(2\theta)\\1/(2\theta)\\(\theta-1)/\theta\end{pmatrix} = \langle \tilde{\boldsymbol{s}}_1, \boldsymbol{w}\rangle \tilde{\boldsymbol{s}}_1 + \langle \tilde{\boldsymbol{s}}_3, \boldsymbol{w}\rangle \tilde{\boldsymbol{s}}_3 = \frac{1}{3}\boldsymbol{e} + \frac{1}{6\theta}(3 - 2\theta)(3\boldsymbol{c}^2 - 2\boldsymbol{e}).$$

Taking into account that

$$\boldsymbol{e}^n = \boldsymbol{e} \quad \text{and} \quad \boldsymbol{c}^{2n+1} = \boldsymbol{c}, \qquad \boldsymbol{c}^{2n} = \boldsymbol{c}^2 \quad \text{for all } n \in \mathbb{N}$$

as well as $\boldsymbol{e}\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^3$, it is easily seen that $W\tilde{\boldsymbol{s}}_1 = \boldsymbol{w}\tilde{\boldsymbol{s}}_1$ and $W^2\tilde{\boldsymbol{s}}_1 = \boldsymbol{w}^2\tilde{\boldsymbol{s}}_1$ lie in the span of $\boldsymbol{e}$ and $\boldsymbol{c}^2$ which implies $W^n\tilde{\boldsymbol{s}}_1 \in \mathrm{span}(\boldsymbol{e}, \boldsymbol{c}^2)$. In contrast, $\tilde{\boldsymbol{s}}_2$ is an eigenvector of W. Hence we obtain two W-invariant subspaces

$$\mathcal{W}_1 := \mathrm{span}(\tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_3), \qquad \mathcal{W}_2 := \mathrm{span}(\tilde{\boldsymbol{s}}_2)$$

with the associated orthogonal projectors

$$P_1 : \boldsymbol{x} \mapsto \langle \tilde{\boldsymbol{s}}_1, \boldsymbol{x}\rangle \tilde{\boldsymbol{s}}_1 + \langle \tilde{\boldsymbol{s}}_3, \boldsymbol{x}\rangle \tilde{\boldsymbol{s}}_3, \qquad P_2 : \boldsymbol{x} \mapsto \langle \tilde{\boldsymbol{s}}_2, \boldsymbol{x}\rangle \tilde{\boldsymbol{s}}_2.$$

[12] q_i is determined by the requirement $q_i(w) = 1/p_i(w)$ for all eigenvalues (diagonal elements) w of W.

4.2. Collision operators with asymmetric rank-one equilibria

Proposition 7 is exactly tailored to treat the case of asymmetric rank-one equilibria that typically arise in discretizing the transient diffusion–advection equation by lattice Boltzmann schemes. The following lemma complements Proposition 7 with regard to our application.

Lemma 12. *Let $A \in \mathbb{R}^{q\times q}$ be defined by $A\boldsymbol{x} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} \quad \text{for } \boldsymbol{x} \in \mathbb{R}^q$. If $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 1$ then A represents a projection, i.e. $A^2 = A$.*

Proof. The little computation below confirms that A indeed satisfies the projector property:

$$A^2\boldsymbol{x} = \langle \boldsymbol{u}, \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v}\rangle\, \boldsymbol{v} = \overbrace{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}^{=1}\langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{x}\rangle \boldsymbol{v} = A\boldsymbol{x}. \quad \blacksquare$$

Proposition 13. *Let the rank of the equilibrium matrix $G \in \mathbb{R}^{q\times q}$ be equal to 1. Then there is a diagonal matrix D such that the BGK collision operator $J = \omega(G - I)$ satisfies $JD = DJ^\top$. Furthermore, provided that G is a projection and D positive definite, J admits a stability structure if and only if $\omega \in [0, 2]$.*

Proof. The existence of D follows from Proposition 7. The second statement is shown in the same way as in the proof of Proposition 8. ■

The subsequent examples illustrate that stability does not only depend on ω but may be influenced also by further parameters occurring in the equilibrium. Stability conditions for these parameters result from the requirement that D must be positive definite.

Example 5. The diffusion–advection equation in one space dimension (with periodic boundary conditions) can be discretized by the D1Q2 lattice Boltzmann algorithm in parabolic scaling.[13] Setting $\boldsymbol{e} := (1, 1)^\top$ and $\boldsymbol{c} := (-1, 1)^\top$, the necessary equilibrium matrix $G \in \mathbb{R}^{2\times 2}$ is defined by

$$G\boldsymbol{f} := \frac{1}{2}\langle \boldsymbol{e}, \boldsymbol{f}\rangle(\boldsymbol{e} + ah\boldsymbol{c}) \quad \text{for } \boldsymbol{f} \in \mathbb{R}^2. \tag{37}$$

By a consistency analysis $a \in \mathbb{R}$ is found to be the advection velocity appearing in the diffusion–advection equation. In opposition, h is a purely algorithmic parameter which corresponds to the spacing of the grid that the lattice Boltzmann algorithm is supposed to run on. Finally, the diffusivity is related to the relaxation parameter ω by $\nu = \frac{1}{\omega} - \frac{1}{2}$.

Since $\langle \frac{1}{2}\boldsymbol{e}, \boldsymbol{e} + ah\boldsymbol{c}\rangle = 1$, the equilibrium (37) represents a projection by Lemma 12. Therefore Proposition 13 can be applied. The diagonal matrix D is given by

$$D = \operatorname{diag}\left((\boldsymbol{e} + ah\boldsymbol{c})/\boldsymbol{e}\right) = \operatorname{diag}(\boldsymbol{e} + ah\boldsymbol{c}) = \operatorname{diag}\left(\begin{pmatrix} 1 - ah \\ 1 + ah \end{pmatrix}\right).$$

It is clearly seen that the diagonal elements of D are positive if and only if

$$|ah| < 1. \tag{38}$$

So we get a stability structure by Proposition 7 and thus stability by Theorem 2 if the above condition on a and the condition on ω in Proposition 13 are satisfied.

Condition (38) is a bit stronger than the corresponding result in [24] (Theorem 2) and [13] (Theorem 6.6) though this was formulated in a slightly different context[14] there. Actually, in these references stability is also proved for $|ah| = 1$. In this case, however, a stability structure cannot exist because of the following reason: According to Proposition 7, the diagonal matrix D is uniquely defined up to multiples as $\boldsymbol{e}$ (which corresponds to $\boldsymbol{u}$ in Proposition 7) does not vanish in any of its components. But as D is *not* positive definite for $|ah| = 1$, the existence of a stability structure is excluded by Theorem 6. So this example proves that the existence of a stability structure is only a sufficient criterion for stability.

Example 6. The D1Q3 velocity model provides another possibility to discretize the same diffusion–advection equation as in the previous example (for a comprehensive consistency analysis cf. [13], chapter 4). With the definition of $\boldsymbol{e}$, $\boldsymbol{c}$ and $\boldsymbol{w}$ as in example 1, the equilibrium matrix corresponding to (37) now becomes

$$G\boldsymbol{f} := \langle \boldsymbol{e}, \boldsymbol{f}\rangle(\boldsymbol{e} + ah\theta\boldsymbol{c})\boldsymbol{w} \quad \text{for } \boldsymbol{f} \in \mathbb{R}^3. \tag{39}$$

The additional scalar quantity θ, which essentially parameterizes the weight vector $\boldsymbol{w}$, determines together with ω the diffusivity by the formula $\nu = \frac{1}{\theta}(\frac{1}{\omega} - \frac{1}{2})$. Let us remark that for $\theta = 1$ the rest population gets decoupled from the other two populations. Then, in principle, the D1Q3 scheme reduces to the D1Q2 algorithm ([13] Section 6.5).

[13] A detailed consistency analysis of virtually the same lattice Boltzmann scheme in hyperbolic scaling can be found in [23]. Owing to the hyperbolic scaling, the resulting macroscopic target equation is not the diffusion–advection but the advection equation.

[14] The analysis in [24] and [13] refers to practically the same algorithm. The only difference is that the factor ah in the equilibrium (37) is replaced by a due to the hyperbolic scaling which also prescribes a different interpretation of the output data of the algorithm.

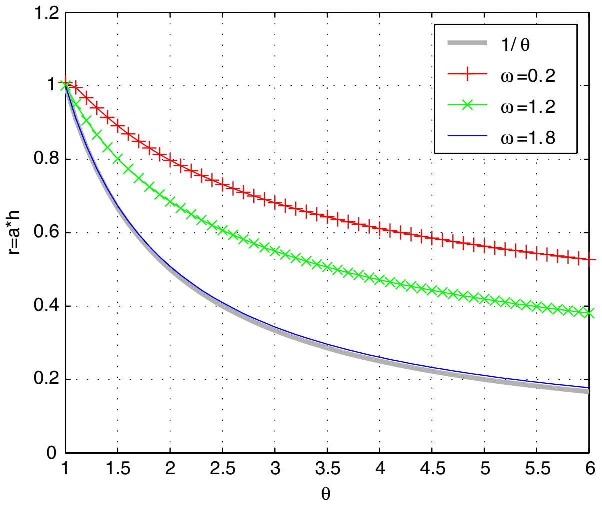

Fig. 1. Several loci of the onset of instability in the θ–ah plane for fixed ω. The curves represent the values of θ and ah where the spectral radius of the evolution operator is equal to 1. Above the curves the spectral radius becomes larger than 1 provided that ω remains constant.

Thanks to the relations $\langle \boldsymbol{e}, \boldsymbol{ew}\rangle = 1$ and $\langle \boldsymbol{e}, \boldsymbol{cw}\rangle = 0$ the equilibrium (39) turns out to be a projector as well, so that Proposition 13 applies too. Furthermore we find a unique symmetrizer for G

$$D = \operatorname{diag}\left((\boldsymbol{ew} + ah\boldsymbol{cw})/\boldsymbol{e}\right) = \operatorname{diag}(\boldsymbol{ew} + ah\boldsymbol{cw}) = \operatorname{diag}\left(\begin{pmatrix} \frac{1}{2\theta}(1 - ah\theta) \\ \frac{1}{2\theta}(1 + ah\theta) \\ \frac{\theta - 1}{\theta} \end{pmatrix}\right).$$

In order to extract sufficient stability conditions for θ and a we have to ensure again that D has a positive diagonal. Evidently, this is the case if and only if

$$\theta > 1 \quad \text{and} \quad |ah\theta| < 1. \tag{40}$$

Notice that the stability condition on the advection velocity a becomes the more restrictive the greater θ is chosen. Hence the largest admissible interval for a is reached for $\theta \to 1$. Therefore, from the viewpoint of stability, the D1Q2 algorithm should be preferred to the D1Q3 algorithm.

Numerical tests corroborate that the second condition in (40), $|ah| < \theta^{-1}$, is not necessary for stability and can be relaxed in spite of the non-existence of a stability structure. The deficiency of a stability structure can be seen by the same argument as in example 5. Keeping the relaxation frequency ω constant, the curves in Fig. 1 indicate for which values of θ and $|ah|$ instability starts to occur if the quantities θ or $|ah|$ are increased. Only for $\omega = 2$ the critical threshold value seems to agree with $|ah| = \theta^{-1}$ as suggested by the stability structure. Furthermore it should be observed that the threshold value gets independent of ω for $\theta \to 1$ (D1Q2 case).

Example 7. The applicability of Proposition 13 is not restricted to one-dimensional lattice Boltzmann schemes. Actually the spatial dimension does not matter as long as the equilibrium is a rank-one projector. As an example let us consider the D2Q4 scheme which discretizes the diffusion–advection equation in two dimensions if the following equilibrium

$$G\boldsymbol{f} = \frac{1}{4}\langle \boldsymbol{e}, \boldsymbol{f}\rangle(\boldsymbol{e} + 2a_x h\boldsymbol{c}_x + 2a_y h\boldsymbol{c}_y) \quad \text{for } \boldsymbol{f} \in \mathbb{R}^4 \tag{41}$$

is employed. Here $\mathbf{a} = (a_x, a_y)^\top \in \mathbb{R}^2$ denotes the (constant) advection velocity and $\boldsymbol{e}, \boldsymbol{c}_x, \boldsymbol{c}_y$ stand for $\boldsymbol{e} := (1, 1, 1, 1)^\top$, $\boldsymbol{c}_x := (1, 0, -1, 0)^\top$ and $\boldsymbol{c}_y := (0, 1, 0, -1)^\top$.

Due to the orthogonality relations $\langle \boldsymbol{e}, \boldsymbol{c}_x\rangle = 0$, $\langle \boldsymbol{e}, \boldsymbol{c}_y\rangle = 0$ and $\frac{1}{4}\langle \boldsymbol{e}, \boldsymbol{e}\rangle = 1$ the equilibrium (41) is revealed as a projection again. The diagonal matrix D is found to be

$$\begin{aligned} D &= \operatorname{diag}\left((\boldsymbol{e} + 2a_x h\boldsymbol{c}_x + 2a_y h\boldsymbol{c}_y)/\boldsymbol{e}\right) \\ &= \operatorname{diag}\left((\boldsymbol{e} + 2a_x h\boldsymbol{c}_x + 2a_y h\boldsymbol{c}_y)\right) = \operatorname{diag}\left(\begin{pmatrix} 1 + 2a_x h \\ 1 + 2a_y h \\ 1 - 2a_x h \\ 1 - 2a_y h \end{pmatrix}\right). \end{aligned}$$

In order to guarantee the positivity of the diagonal we have to require in analogy to (38) and (40) that

$$|2a_x h| < 1 \quad \text{and} \quad |2a_y h| < 1.$$

4.3. Alternative construction of MRT collision operators with stability structure

By the following procedure it is possible to construct MRT collision operators allowing for (pre-)stability structures if the associated BGK collision operator admits such a structure. So let us assume that there is a diagonal and positive definite matrix $\tilde{D}$ such that $J_{\text{BGK}}\tilde{D} = \tilde{D}J_{\text{BGK}}^{\top}$ or equivalently

$$(G - I)\tilde{D} = \tilde{D}(G - I)^{\top}. \tag{42}$$

We would like to find a diagonal, positive definite matrix D together with a relaxation matrix Ω satisfying the conditions (I) and (II) on page 14 such that $J_{\text{MRT}}D = DJ_{\text{MRT}}^{\top}$ or more explicitly

$$\Omega(G - I)D = D(G - I)^{\top}\Omega, \tag{43}$$

where the desired symmetry of Ω has already been taken into account. The idea is now to choose for D the matrix $\tilde{D}$ from the pre-stability structure of the BGK case. Exploiting (42), which states nothing else than the symmetry of $(G - I)\tilde{D}$, the resulting condition on Ω reads

$$\Omega(G - I)\tilde{D} = (G - I)\tilde{D}\Omega.$$

Hence the searched symmetric Ω is required to commute with the symmetric matrix $(G-I)\tilde{D}$. Since two symmetric matrices commute if and only if they have all eigenspaces in common, Ω is determined by $(G-I)\tilde{D}$ up to its eigenvalues. This, however, rises the question whether Ω can simultaneously fulfill condition (II) on page 14. Fortunately, it turns out that (II) becomes satisfied almost automatically. As $G - I$ and $(G - I)D$ have the same left nullspace, the symmetry of $(G - I)D$ entails[15] that the left nullspace of $G-I$ is included in the eigenspace of $(G-I)D$ pertaining to the eigenvalue 0. Therefore (II) is guaranteed if $(G-I)D$ and Ω have the same nullspace (or equivalently the same eigenspace with respect to the eigenvalue 0). The other eigenvalues of Ω can be chosen freely such that the obtained pre-stability structure for $J_{\text{MRT}} = \Omega(G - I)$ finally becomes a stability structure. This necessitates the eigenvalues of $\Omega(G-I)$, which must be real due to the existence of the pre-stability structure, to lie in $[-2, 0]$.

5. Conclusion

Let us briefly highlight and comment the main aspects that have been illuminated in the previous section.

- Thanks to the equivalence statement in Theorem 6 it becomes easy in many situations (sometimes even obvious) to check whether a given collision operator admits a stability structure. In the case of parameterized equilibrium or collision operators (depending for instance on physical parameters like the advection velocity) this circumstance considerably simplifies the search for sufficient stability conditions on the parameters which would need much more effort otherwise.
- Two possibilities to construct MRT collision operators with a stability structure have been described. Often it has been claimed that MRT collision operators exhibit a better stability behavior than BGK collision operators. Please observe that our investigation is neither appropriate to back nor to contest this statement. The existence of a stability structure permits only a statement about stability in principle.
- Example 5 and even more 6 clearly show that a lattice Boltzmann algorithm can be stable although a stability structure does not exist.
- In Examples 5–7 the stability structure is only compatible with the transport operator for periodic boundary conditions. Due to the asymmetry of the equilibria (37), (39) and (41) for the diffusion–advection equation the resulting weight vector $\boldsymbol{b}$ – defined by $B^{\top}B = D^{-1} = \text{diag}(\boldsymbol{b})$ – generally has even not two equal components if $a \neq 0$ (or $\mathbf{a} \neq \mathbf{0}$). Hence condition (6) is violated in this case for transport operators containing the bounce back boundary condition with or without sign flipping. However, in the purely diffusive case where $a = 0$ (or $\mathbf{a} = 0$) this problem does not arise. For instance $\boldsymbol{b}$ is a multiple of $\boldsymbol{e}$ in the case of example 5 and 7, such that all components are equal and condition (6) is trivially satisfied for each of the considered transport operators.
- The construction of an MRT collision operator in Section 4.3 leads to the same diagonal matrix D and thus to the same weight vector $\boldsymbol{b}$ of the stability structure as in the BGK case. Therefore it is not possible to include further boundary conditions in this way. Maybe another construction of MRT collision operators offers more flexibility, to tune the weight vector $\boldsymbol{b}$ in such a way that it becomes compatible with more boundary conditions than in the BGK case.

[15] Observe that left and right nullspaces coincide for symmetric matrices.

- From the viewpoint of application it is interesting to consider also spatially varying advection velocities. This leads to a weight vector $\boldsymbol{b}(\mathbf{x})$ which depends on the grid node. However, condition (6), generalized to this case, implies that $\boldsymbol{b}(\mathbf{x})$ must be constant in the direction of the links to the neighbor nodes. But then the advection velocity must be constant everywhere.

The last two points clearly underline some limitations of the stability structure approach if the equilibrium is not directly given in the form (25) or if it contains spatially dependent parameters. It seems that the framework of the stability structure is not suitable to treat these kinds of problems.

Appendix

For completeness we recall the following two lemmas. The first has been implicitly used to prove Propositions 8 and 13. It is also employed to verify the second lemma, which is cited in the proof of Proposition 9.

Lemma 14. *Let $P \in \mathbb{R}^{q\times q}$ be a projector matrix, i.e. $P^2 = P$. Then P is diagonalizable with* $\operatorname{spec}(P) \subset \{0, 1\}$.

Lemma 15. *Let $(P_i)_{i\in\{1,\dots,p\}} \subset \mathbb{R}^{q\times q}$ be a set of projector matrices such that*

$$P_iP_j = \delta_{ij}P_i \quad \text{for all } i, j \in \{1, \dots , p\} \tag{A.1}$$

which particularly subsumes the projector property $P_i^2 = P_i$ for all $i \in \{1, \dots , p\}$. Then $M \in \mathbb{R}^{q\times q}$ defined by

$$M := \sum_{i=1}^{p} \mu_i P_i \quad \text{with some } \mu_1, \dots , \mu_p \in \mathbb{R}$$

is diagonalizable and $\operatorname{spec}(M) \subset \{0, \mu_1, \dots , \mu_p\}$.

For the proofs of these lemmas we refer to the literature.

Now let us come to the proof of Proposition 3.

Proof. Let us assume that $B^\top B =: \Delta$ is diagonal.

The proof is done by contradiction where we assume $BB^\top := D$ to be diagonal too. From this follows by means of the associativity of the matrix product

$$B\Delta = B(B^\top B) = (BB^\top)B = DB. \tag{A.2}$$

Due to the invertibility of B this equivalent to the *similarity* of Δ and D which means $\Delta = B^{-1}DB$. Therefore Δ and D have the same eigenvalues and thus the same diagonal elements (with equal multiplicity). However, it may happen that the diagonal elements occur in different order. In any case there is a permutation matrix $\Pi \in \mathbb{R}^{q\times q}$ such that[16]

$$\Delta = \Pi D \Pi^{-1}.$$

Inserting this into (A.2) we obtain after some rearrangements $D = \Pi^{-1}B^{-1}DB\Pi$. This entails that $\tilde{B} := B\Pi$ commutes with the diagonal matrix D:

$$\tilde{B}D = D\tilde{B}. \tag{A.3}$$

Supposing that B has a full column let us demonstrate by the subsequent argumentation that (A.3) conflicts with the hypothesis. Since $B\Pi = \tilde{B}$ contains the columns of B in permuted order, there is an index $n \in \{1, \dots , q\}$ so that the n'th column $\tilde{\boldsymbol{b}}_n \in \mathbb{R}^q$ of $\tilde{B}$ has no element equal to 0. Furthermore, as $\Delta \notin \operatorname{span}(I)$ implies $D \notin \operatorname{span}(I)$, it exists a diagonal element of D, say d_m with $m \in \{1, \dots , q\}$ and $m \neq n$ such that

$$d_m \neq d_n. \tag{A.4}$$

[16] There is a permutation $\pi : \{1, \dots , q\} \to \{1, \dots , q\}$ such that $\delta_{ii} = d_{\pi(i)\pi(i)}$ for all $i \in \{1, \dots , q\}$. It is a general fact that any permutation can be written as composition of finitely many transpositions being special permutations that just swap two numbers while the others are kept fixed. Therefore we get $\pi = \tau_s \circ \cdots \circ \tau_1$ for some $s \in \mathbb{N}$. Let T_k denote the permutation matrix associated with τ_k such that multiplication with T_k from left (right) effectuates the swap of rows (columns) with the corresponding indices. As each transposition is inverse to itself, we have $T_k^2 = I$ or equivalently $T_k = T_k^{-1}$. Taking these remarks into account we conclude that

$$\Delta = T_s \cdot \cdots \cdot T_1 D T_1 \cdot \cdots \cdot T_s = T_s \cdot \cdots \cdot T_1 D T_1^{-1} \cdot \cdots \cdot T_s^{-1} = \Pi D \Pi^{-1}$$

where we have set $\Pi := T_s \cdot \ldots \cdot T_1$.

Now the diagonal matrix D acts on the columns if multiplied from right (multiplying the ith column with d_i), while it affects the rows if multiplied from left. In particular we get for the matrix elements with index (m, n) of the two product matrices in (A.3)

$$(\tilde{B}D)_{mn} = \tilde{\beta}_{mn} d_n \quad \text{and} \quad d_m \tilde{\beta}_{mn} = (D\tilde{B})_{mn}$$

where $\tilde{\beta}_{mn}$ denotes the corresponding matrix element of $\tilde{B}$. As $\tilde{\beta}_{mn}$ belongs to $\boldsymbol{b}_n$ it does not vanish. If (A.3) holds then

$$(\tilde{B}D)_{mn} = (D\tilde{B})_{mn} \Leftrightarrow \tilde{\beta}_{mn} d_n = d_m \tilde{\beta}_{mn}.$$

However, dividing the right equation by $\tilde{\beta}_{mn} \neq 0$ contradicts (A.4) which directly emanates from the hypothesis. Hence Eq. (A.3) and thus the diagonalizability of $BB^\top$, from which (A.3) was derived, cannot be true.

If B has a full row instead of a full column the argument proceeds analogously. As multiplication with Π from the left changes only the order of columns, $B\Pi$ has a full row(column) if and only if B has one.

The alternative case, where $BB^\top$ is diagonal, can be handled by setting $C := B^\top$ and applying the above arguments to $C^\top C = BB^\top$. ■

Finally, let us provide the proof of Theorem 6. Another proof is contained in [18].

Proof. "(ii) $\Leftrightarrow$ (iii)" According to Definition 5, the D-symmetry of A_D means explicitly: $DA_D = A_D^\top D$. Replacing A_D by its definition we get (17) and thus (iii). Multiplying (17) by D^{-1} from the left and right gives $D^{-1}A = A^\top D^{-1}$ which says that A is D^{-1}-symmetric.

"(i) $\Rightarrow$ (ii)" Let us suppose that a matrix $B \in \mathbb{R}^{q\times q}$ is given such that $BA = \Lambda B$ with $B^\top B$ and Λ being diagonal. The goal is to find a positive definite, diagonal matrix D which satisfies (ii).

The invertibility of B implies in particular that B has a trivial nullspace which means $B\boldsymbol{x} = 0 \Leftrightarrow \boldsymbol{x} = 0$. Hence it follows for all $\boldsymbol{x} \neq 0$

$$0 < \|Bx\|^2 = \langle B\boldsymbol{x}, B\boldsymbol{x}\rangle = \langle B^\top B\boldsymbol{x}, \boldsymbol{x}\rangle.$$

Due to this equation $B^\top B$ is positive definite and *a fortiori* invertible. As the positive definiteness and the assumed diagonality of $B^\top B$ entails on the inverse, we may set

$$D := (B^\top B)^{-1}.$$

Thanks to the equivalence between (16) and (17) it is now sufficient to verify that AD is a symmetric matrix, i.e. $AD = (AD)^\top = DA^\top$. From the above definition of D follows

$$AD = AB^{-1}(B^\top)^{-1} = B^{-1}BAB^{-1}(B^{-1})^\top$$

where the fact has been employed that transposition and inversion can be swapped. Taking recourse to the hypothesis $BA = \Lambda B$ this simplifies to

$$AD = B^{-1}\Lambda BB^{-1}(B^{-1})^\top = B^{-1}\Lambda(B^{-1})^\top.$$

As Λ is diagonal, it is obvious that the right expression for AD is symmetric indeed.

"(ii) $\Rightarrow$ (i)" By assumption there exists a diagonal, positive definite and thus invertible matrix D such that $A_D := D^{-1}AD$ is symmetric with respect to the scalar product $\langle\cdot,\cdot\rangle_D = \langle D\cdot,\cdot\rangle$ generated by D. Now we have to find a matrix B satisfying (i).

The symmetry of A_D implies[17] that A_D admits a basis of right eigenvectors in $\mathbb{R}^q$ which is orthonormal with respect to $\langle\cdot,\cdot\rangle_D$. Let $E \in \mathbb{R}^{q\times q}$ be the matrix whose columns correspond to these eigenvectors. Analogously to (15) the properties of E are now summarized by:

$$\left.\begin{array}{ll} A_D E = E\Lambda & \text{(eigenbasis)} \\ E^\top DE = I & \text{(D-column-orthonormality)} \end{array}\right\}, \tag{A.5}$$

where Λ is diagonal (containing the eigenvalues of A_D). Having full rank, E and its transposed are invertible.

Eq. $(A.5)_1$ transforms to

$$ADE = DE\Lambda \tag{A.6}$$

after substituting $A_D = D^{-1}AD$ and multiplying with D from the left. In virtue of the invertibility of D and E we may set

$$B := (DE)^{-1}. \tag{A.7}$$

Multiplying (A.6) by B from both sides results into the equation $BA = \Lambda B$ which verifies the first half of (13).

[17] Cf. the *spectral theorem* for symmetric endomorphisms in finite-dimensional vector spaces.

To complete the proof it remains to check that $B^\top B$ as defined in (A.7) is diagonal. Using (A.7) we get

$$B^\top B = \left((DE)^{-1}\right)^\top (DE)^{-1}.$$

Reversing transposition and inversion and noticing that $D = D^\top$, we obtain

$$\begin{aligned} B^\top B &= \left((DE)^\top\right)^{-1} E^{-1}D^{-1} = \left(E^\top D\right)^{-1} E^{-1}D^{-1} = D^{-1}(E^\top)^{-1}E^{-1}D^{-1} \\ &= D^{-1}(EE^\top)^{-1}D^{-1}. \end{aligned}$$

Rearranging $(A.5)_2$ yields $D = (EE^\top)^{-1}$ which just appears above as a factor. Inserting this we finally end up with

$$B^\top B = D^{-1}(EE^\top)^{-1}D^{-1} = D^{-1}DD^{-1} = D^{-1}.$$

Hence $B^\top B$ is equal to a diagonal matrix as diagonality is conserved under matrix inversion. ■

Concerning the second part of the proof it might seem more appropriate to consider the left eigenbasis. This, however, is not D-orthonormal but D^{-1}-orthonormal.[18] If this circumstance is observed the modified proof leads to the same B.

References

[1] U. Frisch, D. d'Humières, B. Hasslacher, P. Lallemand, Y. Pomeau, J.P. Rivet, Lattice gas hydrodynamics in two and three dimensions, Complex Systems 1 (1987) 649–707.

[2] G. McNamara, G. Zanetti, Use of Boltzmann equation to simulate lattice-gas automata, Phys. Rev. Lett. 61 (1988) 2332.

[3] G. Doolen, U. Frisch, B. Hasslacher, S. Orzag, S. Wolfram (Eds.), Lattice-gas Methods for Partial Differential Equations, Addison-Wesley, Redwood City, CA, 1990.

[4] Y. Qian, D. d'Humières, P. Lallemand, Lattice BGK models for Navier–Stokes equation, Europhys. Lett. 17 (1992) 479–484.

[5] D.A. Wolf-Gladrow, Lattice-gas Cellular Automata and Lattice Boltzmann models, in: Lecture Notes in Mathematics, Springer, 2000.

[6] H. Chen, S. Chen, W. Matthaeus, Recovery of the Navier–Stokes equations using a lattice-gas Boltzmann method, Phys. Rev. A 45 (1992) R5339–5342.

[7] M. Junk, A finite difference interpretation of the lattice-Boltzmann method, Numer. Methods Partial Differential Equations 17 (2001) 383–402.

[8] M. Junk, A. Klar, L.-S. Luo, Asymptotic analysis of the lattice Boltzmann equation, J. Comput. Phys. 210 (2005) 676–704.

[9] J.D. Sterling, S. Chen, Stability analysis of the lattice-Boltzmann method, J. Comput. Phys. 123 (1996) 196–206.

[10] R.A. Worthing, J. Mozer, G. Seeley, Stability of lattice-Boltzmann methods in hydrodynamic regime, Phys. Rev. E 56 (1997) 2243–2253.

[11] P. Lallemand, L.-S. Luo, Theory of the lattice Boltzmann method: Dispersion, dissipation, isotropy, Galilean invariance and stability, Phys. Rev. E 61 (6) (2000) 6546–6562.

[12] J. Weiß, Numerical analysis of lattice-Boltzmann methods for the heat equation on a bounded interval, Ph.D. thesis, Universität Karlsruhe, 2006.

[13] M. Rheinländer, Analysis of lattice Boltzmann methods — asymptotic and numeric investigation of a singularly perturbed system, Ph.D. thesis, Universität Konstanz, 2007, URL: http://www.ub.uni-konstanz.de/kops/volltexte/2007/3635/.

[14] M. Junk, W.-A. Yong, Weighted l^2-stability of the lattice Boltzmann method, SIAM J. Numer. Anal. 47 (2009) 1651–1665.

[15] M. Junk, W.-A. Yong, Rigorous Navier–Stokes limit of the lattice Boltzmann equation, Asymptot. Anal. 35 (2003) 165–185.

[16] M. Banda, W.-A. Yong, A. Klar, A stability notion for lattice-Boltzmann equations, SISC 27 (6) (2006) 2098–2111.

[17] M. Junk, Z. Yang, Convergence of Boltzmann methods for Stokes flows in periodic and bounded domains, Comput. Math. Appl. 55 (2008) 1481–1491.

[18] W.-A. Yong, An Onsager-like relation for the lattice Boltzmann method arXiv:0805.1483v1 [physics.comp-ph], 10 May 2008.

[19] W.-A. Yong, An interesting class of partial differential equations, arXiv:0707.3708v2 [math-ph], 28 Aug 2007.

[20] R. Horn, C. Johnson, Topics in matrix analysis, Cambridge University press, 1991.

[21] D. d'Humières, Generalized lattice-Boltzmann equations, in: B.D. Shizgal, D.P. Weave (Eds.), Rarefied Gas Dynamics: Theory and Simulations, in: Prog. Astronaut. Aeronaut., vol. 159, AIAA, Washington, DC, 1992, pp. 450–458.

[22] D. d'Humières, I. Ginzburg, M. Krafczyk, P. Lallemand, L.-S. Luo, Multiple-relaxation-time lattice Boltzmann models in three dimensions, Philos. Trans. R. Soc. Lond. A 360 (1972) (2002) 437–451.

[23] M. Junk, M. Rheinländer, Consistency and multiscale analysis of a lattice-Boltzmann method, Prog. CFD 8 (1–4) (2008) 25–37.

[24] M. Rheinländer, Stability and multiscale analysis of an advective lattice-Boltzmann scheme, Prog. CFD 8 (1–4) (2008) 56–68.

[18] Multiplying $(A.5)_1$ with E^{-1} from the left and right yields $E^{-1}A_D = \Lambda E^{-1}$, which shows that the rows of E^{-1} are left eigenvectors of A_D. By inverting $(A.5)_2$ we get $E^{-1}D^{-1}(E^{-1})^\top$. This establishes the D^{-1}-orthonormality of E^{-1}. Therefore the left eigenvectors of A_D are orthogonal with respect to the scalar product generated by D^{-1}.