ON STABILITY, MONOTONICITY, AND CONSTRUCTION OF CENTRAL SCHEMES FOR HYPERBOLIC CONSERVATION LAWS WITH SOURCE TERMS I: THEORY

V. S. BORISOV * AND M. MOND †

Abstract. The monotonicity and stability of difference schemes for, in general, hyperbolic systems of conservation laws with source terms are studied. The basic approach is to investigate the stability and monotonicity of a non-linear scheme in terms of its corresponding scheme in variations. Such an approach leads to application of the stability theory for linear equation systems to establish stability of the corresponding non-linear scheme.

In this first paper, we focus on the theoretical background. The main methodological innovation is the theorems establishing the notion that a non-linear scheme is stable (and monotone) if the corresponding scheme in variations is stable (and, respectively, monotone). Criteria are developed for monotonicity and stability of difference schemes associated with the numerical analysis of systems of partial differential equations. The theorem of Friedrichs (1954) is generalized to be applicable to variational schemes. High-order interpolation and employment of monotone piecewise cubics in construction of monotone central schemes are considered.

Key words. Stability, monotonicity, difference equations, difference schemes, central schemes, schemes in variations, hyperbolic equations, systems of partial differential equations, source terms, numerical solution, spurious oscillations, monotone piecewise cubics

1. Introduction.

We are mainly concerned with the stability and monotonicity [4] of difference schemes for hyperbolic systems of conservation laws with source terms. Such systems are used to describe many physical problems of great practical importance in magneto-hydrodynamics, kinetic theory of rarefied gases, linear and nonlinear waves, viscoelasticity, hydrodynamical models for semiconductors, multiphase flows and phase transitions, radiation hydrodynamics, relaxing gas flows with thermal and chemical non-equilibrium, shallow waters, traffic flows, etc. (see, e.g., [2], [6], [12], [19], [24], [25], [26], [28], [32], [36], [37] and references therein). We will consider a 1-D system of the conservation laws written in the following form

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial}{\partial x}\mathbf{f}(\mathbf{u}) = \frac{1}{\tau}\mathbf{q}(\mathbf{u}), \quad x \in \mathbb{R},\ 0 < t \le T_{\max}; \quad \mathbf{u}(x,t)|_{t=0} = \mathbf{u}^0(x), \tag{1.1}$$

where $\mathbf{u} = \{u_1, u_2, \ldots, u_M\}^T$ is a vector-valued function from $\mathbb{R} \times [0, +\infty)$ into an open subset $\Omega_{\mathbf{u}} \subset \mathbb{R}^M$, $\mathbf{f}(\mathbf{u}) = \{f_1(\mathbf{u}), f_2(\mathbf{u}), \ldots, f_M(\mathbf{u})\}^T$ is a smooth function (flux-function) from $\Omega_{\mathbf{u}}$ into $\mathbb{R}^M$, $\mathbf{q}(\mathbf{u}) = \{q_1(\mathbf{u}), q_2(\mathbf{u}), \ldots, q_M(\mathbf{u})\}^T$ denotes the source term, $\tau > 0$ denotes the stiffness parameter, $\mathbf{u}^0(x)$ is either of compact support or periodic. We will assume that the system (1.1) is hyperbolic and, hence, the Jacobian matrix of $\mathbf{f}(\mathbf{u})$ possesses M linearly independent eigenvectors (see, e.g., [12]). Here and in what follows $\|\mathbf{M}\|_p$ denotes the matrix norm of a matrix $\mathbf{M}$ induced by the vector norm $\|\mathbf{v}\|_p = \left(\sum_i |v_i|^p\right)^{1/p}$, and $\|\mathbf{M}\|$ denotes the matrix norm induced by a prescribed vector norm. $\mathbb{R}$ and $\mathbb{C}$ denote the fields of real and complex numbers, respectively, and $\mathbb{K}$ denotes either of these fields.

In the numerical solution of the, in general, stiff ($\tau \ll 1$) system (1.1), one is seeking to establish a numerical scheme that would be robust enough to eradicate spurious

*The Pearlstone Center for Aeronautical Engineering Studies, Department of Mechanical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel. E-mail: viatslav@bgu.ac.il

†The Pearlstone Center for Aeronautical Engineering Studies, Department of Mechanical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel. E-mail: mond@bgu.ac.il

oscillations, i.e. a monotone scheme [4]. At the present time there are several notions of scheme monotonicity. The notion of 'monotonicity preserving scheme' originally appeared in Godunov [13]. Such a scheme transforms a monotone increasing (or decreasing) function $v\left(x\right)$ of space coordinate x at a time level t into a monotone increasing (or decreasing, respectively) function $\widehat{v}\left(x\right)$ at the next time level $t+\Delta t$. Nowadays monotonicity preserving schemes are also known as, e.g., monotonicity conserving iterations (or methods) [18], monotone schemes (e.g., [1], [22], [30]), monotonicity preserving methods [26], and Godunov-monotone schemes [4]. Harten et al. [14] put forward their own definition of scheme monotonicity as follows: a difference scheme

$$\widehat{v}_i = H\left(v_{i-k}, v_{i-k+1}, \ldots, v_{i+k}\right) \tag{1.2}$$

is said to be monotone if H is a monotone increasing function of each of its arguments. The following definition is due to Samarskiy: a scheme is regarded as monotone if the boundary maximum principle is maintained [41] (see also, e.g., [42, p. 183], [4], [3]). A difference scheme may also be referred to as monotone if a maximum principle, e.g., the boundary maximum principle, the region maximum principle, the maximum principle for inverse column entries, the maximum principle for the absolute values, etc', holds for this scheme [5]. A further important notion of difference scheme monotonicity was, in fact, done in [35] (see also [4]). A scheme will be referred to as monotone if it is monotonicity preserving [13] and transforms a "$\wedge$-function" (or "$\vee$-function") into a "μ - function" (or into an "η - function", respectively). Here and in what follows, a scalar grid function v_i will be referred to as $\wedge$-function (or $\vee$-function) if there exist grid nodes m and n such that $m \leq n$; $v_m > v_{m-1}$ and $v_i \geq v_{i-1}$ ($v_m < v_{m-1}$ and $v_i \leq v_{i-1}$ for the $\vee$-function) if $i < m$; $v_i = const$ if $m \leq i \leq n$; $v_n > v_{n+1}$ and $v_i \geq v_{i+1}$ ($v_n < v_{n+1}$ and $v_i \leq v_{i+1}$ for the $\vee$-function) if $i > n$. Simply stated, the $\wedge$-function (or $\vee$-function) is a scalar grid function v_i that has only one generalized local maximum [35] (or generalized local minimum [35], respectively) The set of μ-functions (or η-functions) is the union of $\wedge$-functions ($\vee$-functions, respectively) and the set of monotone functions.

Hereinafter, for the sake of convenience, a monotonicity preserving scheme will be referred to as Godunov-monotone (or G-monotone for short), a scheme monotone in terms of Harten et al. [14] will be referred to as H-monotone, a scheme will be referred to as Samarskiy-monotone (or S-monotone for short) if a maximum principle holds for this scheme (see, e.g., [41], [42, p. 183], [4], [5]), and a scheme being monotone from the standpoint of Ostapenko [35] will be referred to as Godunov-Ostapenko-monotone (or GO-monotone for short) [4].

Notice, aiming to construct non-oscillatory numerical schemes for the equations of hydrodynamics, heat transfer problems, etc., Godunov [13] points out that these schemes must possess such an important property of the differential equations as monotonicity preservation. Thus, with the use of a G-monotone scheme, any discontinuity in the initial monotone data can be smeared in succeeding time steps but cannot become oscillatory. Godunov [13, p. 275] has proven that linear schemes with constant coefficients in the form

$$\widehat{v}_i = \sum_{n=-k}^{k} a_n v_{i+n} \tag{1.3}$$

will be G-monotone *iff* (i.e. if and only if) all of the scheme coefficients $a_n \geq 0$. Let us recall that a linear scheme is referred to as having constant coefficients if the scheme coefficients are the same in all equations of the scheme, e.g. a_n in (1.3) does not depend on i. A linear scheme with variable coefficients can be written in the form

$$\widehat{v}_i = \sum_j a_i^j v_j. \tag{1.4}$$

Schemes with variable coefficients are not uncommon in practice. For instance, such a scheme arises if we solve (1.1) on a non-uniform grid. It is easy to see [4] that if a constant will be a solution to (1.4), then the set of equalities

$$\sum_j a_i^j = 1, \quad \forall\, i \tag{1.5}$$

is the necessary condition for G-monotonicity of (1.4).

For studying G-monotonicity of non-linear schemes the notion of total variation (TV, see, e.g., [12], [24], [26]) turns out to be an useful tool. We recall the following definition just for the sake of completeness. Scheme (1.2) is said to be total variation diminishing (TVD) if

$$TV(\widehat{v}) \leq TV(v), \quad TV(v) \equiv \sum_j |v_{j+1} - v_j| \,. \tag{1.6}$$

TVD schemes are attractive for at least the following reasons: (i) The notion of a TVD method is sufficient to prove convergence [26, p. 148]; (ii) H-monotone schemes are TVD [15], and any TVD scheme is G-monotone [12, p. 168], [15], [26, p. 110]; (iii) There exist simple sufficient conditions for a scalar scheme to be TVD [12, p. 169]. Besides, it is widely believed that TVD methods are free from spurious oscillations (e.g., [7], [15], [20], [26], [27], [33]); and thus, for the above-stated reasons, TVD schemes are in common practice (see, e.g., [12], [24], [25], [26], [30], [31], [32], [36], [37], [38], [44] and references therein).

It should be mentioned the long-known fact that TVD schemes can produce spurious oscillations [17]. It is pointed out in [17] that these oscillations are small in the scalar case, and the notion of essentially non-oscillatory (ENO) schemes is introduced. ENO schemes, [17], allow for the production of spurious oscillations on the level of the truncation error, but do not have a Gibbs-like phenomenon at jump-discontinuities, and hence do not involve the generation of spurious oscillations proportional to the size of the jump. Nowadays, a TVD scheme is said to have ENO-type oscillations if their amplitude decreases as the grid is refined [38]. It should be mentioned here that every convergent scheme has such a property, because the numerical solution should converge to the true solution of the differential equation as the grid is refined.

Notice, Harten's theorem relative to G-monotonicity of TVD schemes was proven in [15] for a specific class of schemes approximating a 1-D scalar partial differential equation (PDE), and hence it does not always happen that a TVD scheme is G-monotone. Actually, let us consider the following linear scheme with variable coefficients:

$$\widehat{v}_i = \alpha_i v_{i-1} + \beta_i v_i + \gamma_i v_{i+1}, \tag{1.7}$$

where $\alpha_i = \beta_i = \gamma_i = 1/3$ at $i = -1, \pm 2, \pm 3, \ldots$, and $\alpha_0 = \beta_0 = \varepsilon$, $\gamma_0 = \alpha_1 = 1 - 2\varepsilon$, $\beta_1 = \gamma_1 = \varepsilon$ $(0 < \varepsilon < 0.25)$. Since α_i, β_i, $\gamma_i > 0$ $\forall i$, the scheme (1.7) is H-monotone and, hence, TVD [15]. Considering the monotone increasing function v_i, namely

$v_i = 0$ for all $i \leq 0$ and $v_i = 1$ for all $i > 0$, we obtain that $\widehat{v}_i = 0$ for all $i < 0$, $\widehat{v}_0 = 1-2\varepsilon$, $\widehat{v}_1 = 2\varepsilon$, and $\widehat{v}_i = 1$ for all $i > 1$. We note that the grid function $\widehat{v}_i$ will not be monotone. Hence, a TVD scheme can produce not only small spurious oscillations (cf. [17]). It is demonstrated in [4] that a scalar TVD scheme, be it non-linear or even linear with constant coefficients, can produce spurious oscillations comparable with the size of the jump-discontinuity.

For studying the stability of non-linear schemes the notion of TV turns out to be an effective tool. Actually, the following property

$$\|\mathcal{N}(\mathbf{v}+\delta\mathbf{v}) - \mathcal{N}(\mathbf{v})\| \leq (1+\alpha\Delta t)\,\|\delta\mathbf{v}\| \tag{1.8}$$

is sufficient for stability of a two-step method [26], however it is, in general, difficult to obtain. Here Δt denotes the time increment, α is a constant independent of Δt as $\Delta t \to 0$, $\mathbf{v}$ and $\delta\mathbf{v}$ are any two grid functions ($\delta\mathbf{v}$ will often be referred to as the variation of the grid function $\mathbf{v}$), $\mathcal{N}$ denotes the scheme operator. At the same time, the stability of linearized version of the non-linear scheme is generally not sufficient to prove convergence [16], [26]. Instead, the TV-stability adopted in [16] makes it possible to prove convergence of non-linear scalar schemes with ease. However, the TVD property is a purely scalar notion that cannot, in general, be extended for non-linear systems of equations, as the true solution itself is usually not TVD [12], [26]. Because of it, as noted in [26], in general, no numerical method for non-linear systems of equations has been proven to be stable. There is not even a proof that the first-order Godunov method converges on general systems of non-linear conservation laws [26, p. 340]. Thus, a different approach to testing scheme stability must be adopted to prove convergence of non-linear schemes for systems of PDEs. The notion of variational scheme (or scheme in variations), see [4] and [5], has, in all likelihood, much potential to be an effective tool for studying stability of nonlinear schemes approximating systems of PDEs. An analogous approach suggested by Lyapunov (1892) (see, e.g., [8]), namely to investigate stability by the first approximation, has long been exploited for investigation of the stability of motion (e.g. [8], [10]) as well as the stability of difference equations [11]. For completeness sake we establish the notion that the stability of a variational scheme implies the stability of its original scheme (see Section 2.1, Theorem 2.2 and Remark 2.3).

An approach to investigate non-linear difference schemes for S-monotonicity in terms of corresponding variational schemes was suggested in [4], [5]. The advantage of such an approach is that the variational scheme will always be linear (although it may be emanating from a nonlinear operator) and, hence, enables the investigation of the monotonicity for nonlinear operators using linear patterns. It is proven for the case of explicit schemes that the S-monotonicity of a variational scheme will guarantee that its original scheme also will be S-monotone [4]. Analogous theorem for the case of implicit schemes can be found in Section 2.1, Theorem 2.1.

By way of illustration, let us consider the variational scheme corresponding to the non-linear scheme (1.2):

$$\delta\widehat{v}_i = \sum_{n=-k}^{k} a_i^n \delta v_{i+n};\ a_i^n \equiv \frac{\partial}{\partial v_{i+n}} H(v_{i-k}, v_{i-k+1}, \ldots, v_{i+k}),\ -k \leq n \leq k. \tag{1.9}$$

The necessary and sufficient conditions for the scheme (1.9) to be S-monotone are the following (see Corollary 2.15 in [5])

$$\sum_{n=-k}^{k} |a_i^n| \leq 1, \quad \forall i. \tag{1.10}$$

Hence, (1.10) is sufficient for the scheme (1.2) to be S-monotone [4, p. 1575]. Let a constant ($\delta\widehat{v} = \delta v = const$) will be a solution to the scheme in variations, (1.9). Then we obtain from (1.9) that the coefficients of the variational scheme fulfill (1.5). In such a case, if (1.9) is S-monotone, then $a_i^n \geq 0$ [4, p. 1575]. Thus, H-monotonicity of (1.2) will be the necessary condition for S-monotonicity of its variational scheme.

Since a variational scheme carries such an important properties of its original scheme as S-monotonicity and GO-monotonicity ([4], see also Section 2.1, Theorem 2.1), the following definition will be useful in the investigation on scheme monotonicity.

Definition 1.1. *A numerical scheme is termed variationally monotone if its variational scheme is monotone.*

Inasmuch as there are several notions of monotonicity for a numerical scheme, it is evident that there must be several notions of variational monotonicity.

Let us note that the notion of S-monotonicity for an explicit scheme coincides with the notion of scheme operator contractivity [26, p. 144], i.e. the scheme will be S-monotone if (1.8) is valid under $\alpha = 0$. Hence, S-monotonicity of a numerical scheme implies the stability of the scheme.

It is demonstrated in [4] that S-monotonicity implies TVD property of a scalar 3-point scheme, however the scheme can be oscillatory. Furthermore, a conservative scalar scheme consistent with a transport equation, H-monotone (hence TVD [15] and consistent with the entropy condition [14]), S-monotone, and G-monotone can nevertheless produce spurious oscillations [4]. Thus, a different approach to monotonicity must be adopted. As it was demonstrated in [4], the notion of GO-monotonicity is a very helpful tool for the construction of non-oscillatory schemes. However, a GO-monotone scheme can be not TVD [4]. Thus, if a numerical scheme will be GO-monotone as well as S-monotone (GOS-monotone for short), then this scheme will be stable and, in general, free of spurious oscillations [4].

2. Monotonicity and stability of difference schemes.

2.1. Non-linear schemes.

We consider a nonlinear implicit scheme

$$\mathbf{H}_i\left(\mathbf{y}_1, \ldots, \mathbf{y}_M\right) = \mathbf{x}_i, \quad i \in \omega \equiv \{1, 2, \ldots, M\}, \tag{2.1}$$

where $\mathbf{y}_i \in L \equiv \mathbb{K}^N$ denotes the sought-after vector-valued function of grid nodes, $\mathbf{x}_i \in L \equiv \mathbb{K}^N$ denotes the prescribed vector-valued function of grid nodes, $\mathbf{H}_i = \{H_{i,1}, \ldots, H_{i,N}\}^T$ is a vector-valued function with the range belonging to $\mathbb{K}^N$. If we introduce the additional notation $\mathbf{y} = \left\{\mathbf{y}_1^T, \ldots, \mathbf{y}_M^T\right\}^T$, $\mathbf{x} = \left\{\mathbf{x}_1^T, \ldots, \mathbf{x}_M^T\right\}^T$, $\mathbf{H} = \left\{\mathbf{H}_1^T, \ldots, \mathbf{H}_M^T\right\}^T$, then the scheme (2.1) can be represented in the form

$$\mathbf{H}\left(\mathbf{y}\right) = \mathbf{x}, \quad \mathbf{x} \in L^M, \ \mathbf{y} \in L^M. \tag{2.2}$$

Theorem 2.1. *Let a nonlinear implicit scheme (2.1) be written in the form (2.2), where $\mathbf{x} \in \Omega_x \subset L^M$, Ω_x denotes a closed and bounded convex set. Let the mapping $\mathbf{H}$ in (2.2) have a strong Fréchet derivative (strong F-derivative [34, item 3.2.9]), $\mathbf{H}'\left(\mathbf{y}\right)$, at every $\mathbf{y} \in int\left(\Omega_y\right)$ provided that $\mathbf{H}\left(\Omega_y\right) = \Omega_x$, and let $\mathbf{H}'\left(\mathbf{y}\right)$*

be nonsingular. Then, for any $\mathbf{x}$*,* $\mathbf{x}+\delta\mathbf{x}\in\Omega_x$ *the scheme will be S-monotone if its variational difference scheme is S-monotone.*

Proof. The scheme (2.2) can be seen, [5, p. 1126], as a two-node implicit scheme, and its variational scheme becomes

$$\delta\mathbf{x}=\mathbf{H}'\left(\mathbf{y}\right)\cdot\delta\mathbf{y},\quad \mathbf{H}'\left(\mathbf{y}\right)\equiv\frac{\partial\mathbf{H}\left(\mathbf{y}\right)}{\partial\mathbf{y}}. \tag{2.3}$$

As $\mathbf{H}'\left(\mathbf{y}\right)$ is nonsingular, we may rewrite (2.3) in the form

$$\delta\mathbf{y}=\left(\mathbf{H}'\right)^{-1}\left(\mathbf{y}\right)\cdot\delta\mathbf{x}. \tag{2.4}$$

Then, by (2.4)

$$\left\|\delta\mathbf{y}\right\|=\left\|\left(\mathbf{H}'\right)^{-1}\left(\mathbf{y}\right)\cdot\delta\mathbf{x}\right\|\leq\left\|\left(\mathbf{H}'\right)^{-1}\left(\mathbf{y}\right)\right\|\left\|\delta\mathbf{x}\right\|. \tag{2.5}$$

Let (2.3) be S-monotone, i.e. let

$$\left\|\delta\mathbf{y}\right\|\leq\left\|\delta\mathbf{x}\right\|. \tag{2.6}$$

In view of (2.5) and (2.6) we obtain [5, p. 1126] that

$$\left\|\left(\mathbf{H}'\right)^{-1}\left(\mathbf{y}\right)\right\|\leq 1,\ \forall\mathbf{y}\in\Omega_y\subseteq L^M. \tag{2.7}$$

In view of the inverse function theorem [34, item 5.2.1] we obtain from (2.2) that

$$\mathbf{y}=\mathbf{H}^{-1}\left(\mathbf{x}\right),\quad \mathbf{x}\in\Omega_x\subset L^M,\ \mathbf{y}\in\Omega_y\subseteq L^M. \tag{2.8}$$

By virtue of the mean-value theorem [34, item 3.2.3] we obtain from (2.8)

$$\left\|d\mathbf{y}\right\|=\left\|\mathbf{H}^{-1}\left(\mathbf{x}+d\mathbf{x}\right)-\mathbf{H}^{-1}\left(\mathbf{x}\right)\right\|\leq\sup_{0\leq t\leq1}\left\|\left(\mathbf{H}^{-1}\right)'\left(\mathbf{x}+td\mathbf{x}\right)\right\|\left\|d\mathbf{x}\right\|$$

$$\leq\sup_{\mathbf{y}\in\Omega_y}\left\|\left(\mathbf{H}^{-1}\right)'\left(\mathbf{H}\left(\mathbf{y}\right)\right)\right\|\left\|d\mathbf{x}\right\|. \tag{2.9}$$

In view of the inverse function theorem [34, item 5.2.1] we can write

$$\left(\mathbf{H}^{-1}\right)'\left(\mathbf{H}\left(\mathbf{y}\right)\right)=\left(\mathbf{H}'\right)^{-1}\left(\mathbf{y}\right). \tag{2.10}$$

By virtue of (2.10) and (2.7) we obtain from (2.9) that

$$\left\|d\mathbf{y}\right\|\leq\left\|d\mathbf{x}\right\|. \tag{2.11}$$

The inequality (2.11) manifests the prove of the theorem. □

Theorem 2.2. *Let a non-linear explicit scheme be written in the form*

$$\widehat{\mathbf{v}}=\mathbf{H}(\mathbf{v}),\quad \widehat{\mathbf{v}}\in L,\quad \mathbf{v}\in\Omega\subset L, \tag{2.12}$$

where Ω *denotes a closed and bounded convex set in a linear vector space* L*. Then for any* $\mathbf{v},\mathbf{v}+\delta\mathbf{v}\in\Omega$ *the scheme will be stable if its variational scheme is stable.*

Proof. The variational scheme corresponding to the scheme (2.12) reads

$$\delta\widehat{\mathbf{v}} = \mathbf{H}'(\mathbf{v}) \cdot \delta\mathbf{v}, \quad \mathbf{H}'(\mathbf{v}) \equiv \frac{\partial \mathbf{H}(\mathbf{v})}{\partial \mathbf{v}}. \tag{2.13}$$

The linear scheme (2.13) will be stable [26, p. 145] if $\|\mathbf{H}'(\mathbf{v})\| \leq 1+\alpha\Delta t$ for all $\mathbf{v} \in \Omega$, that is

$$\sup_{\mathbf{v}\in\Omega} \|\mathbf{H}'(\mathbf{v})\| \leq 1 + \alpha\Delta t. \tag{2.14}$$

By virtue of the mean-value theorem [34, item 3.2.3] we obtain from (2.12)

$$\|\mathbf{H}(\mathbf{v} + \delta\mathbf{v}) - \mathbf{H}(\mathbf{v})\| \leq \sup_{0\leq\zeta\leq1} \|\mathbf{H}'(\mathbf{v} + \zeta\delta\mathbf{v})\| \, \|\delta\mathbf{v}\| \leq \sup_{\mathbf{v}\in\Omega} \|\mathbf{H}'(\mathbf{v})\| \, \|\delta\mathbf{v}\|. \tag{2.15}$$

In view of (2.14) we conclude from (2.15) that the inequality (1.8) for (2.12) will be fulfilled, and hence the original non-linear scheme (2.12) will be stable. □

Remark 2.3. *Theorem 2.2 can be reformulated for implicit schemes with ease. The proof of this theorem for implicit schemes is identical to the proof of Theorem 2.1.*

2.2. Linear schemes. Let $\mathbf{v} \equiv \left\{\ldots, \mathbf{v}_i^T, \ldots\right\}^T$ (or $\mathbf{A} = \{\mathbf{A}_{ij}\}$) be a partitioned [29] vector (or a matrix, respectively), then we shall denote by $\langle\mathbf{v}\rangle$ the ordinary vector obtained from $\mathbf{v}$ (or by $\langle\mathbf{A}\rangle$ the ordinary matrix obtained from $\mathbf{A}$, respectively) by removing its partitions. It is easy to see that

$$\|\mathbf{v}\|_\infty \equiv \max_i \|\mathbf{v}_i\|_\infty = \|\langle\mathbf{v}\rangle\|_\infty . \tag{2.16}$$

To start with, we obtain necessary conditions for some class of linear schemes to be GOS-monotone. We consider the following explicit homogeneous scheme

$$\mathbf{z}_i = \sum_j \mathbf{B}_{ij} \cdot \mathbf{y}_j, \quad \mathbf{z}_i, \mathbf{y}_j \in L, \tag{2.17}$$

where L denotes the linear vector space with the orthonormal basis $\{\mathbf{b}_l\}_1^M$, $\mathbf{b}_1 = \{1, 0, \ldots, 0\}^T$, $\mathbf{b}_2 = \{0, 1, \ldots, 0\}^T$, ..., $\mathbf{b}_M = \{0, 0, \ldots, 1\}^T$; $\mathbf{B}_{ij} \equiv \left\{B_{ij}^{kl}\right\}$ is a square matrix. It is assumed that any constant (i.e., $\mathbf{z}_i = \mathbf{y}_j \equiv \mathbf{c} = const$) is a solution to (2.17). Then, in view of (2.17), we find that

$$\sum_j \mathbf{B}_{ij} = \mathbf{I}, \quad \forall i \tag{2.18}$$

will be the necessary condition for (2.17) to be G-monotone (cf. [4, p. 1560]). Here and in what follows, $\mathbf{I}$ denotes the identity operator.

Theorem 2.4. *Let an explicit linear scheme be written in the form (2.17), and let any constant be a solution to (2.17). If (2.17) is GOS-monotone, then the diagonal elements, B_{ij}^{kk}, of the matrices $\mathbf{B}_{ij} \equiv \left\{B_{ij}^{kl}\right\}$ are non-negative, i.e.*

$$B_{ij}^{kk} \geq 0, \quad \forall i, j, k, \tag{2.19}$$

and

$$B_{ij}^{kk} \ is \ a \ \mu - function \ of \ i \ for \ all \ j \ and \ k. \tag{2.20}$$

Proof. We consider (2.17) when $\mathbf{y}_j = y_j \mathbf{b}_l$, $l = 1,\ 2,\ ...,M$, where y_j is a scalar value. Let $\left\{z^l_{1,i}, z^l_{2,i}, \ldots, z^l_{M,i}\right\}^T$ be the left-hand side of (2.17) under $\mathbf{y}_j = y_j \mathbf{b}_l$. Then we obtain from (2.17) the following system of decoupled scalar equalities

$$z^l_{k,i} = \sum_j B^{kl}_{ij} y_j, \quad k, l = 1, 2, \ldots, M. \tag{2.21}$$

In view of Corollary 3.14 in [4], the scheme (2.21) will be S-monotone *iff*

$$\sum_j \left|B^{kl}_{ij}\right| \leq 1, \quad \forall i, k, l. \tag{2.22}$$

In view of (2.18) we have

$$\sum_j B^{kk}_{ij} = 1, \quad \forall i, k, \tag{2.23}$$

and hence

$$\sum_j \left|B^{kk}_{ij}\right| \geq 1, \quad \forall i, k. \tag{2.24}$$

By virtue of (2.24) and using (2.22) under $k = l$ we obtain that

$$\sum_j \left|B^{kk}_{ij}\right| = 1, \quad \forall i, k. \tag{2.25}$$

Thus, (2.23) and (2.25) must be valid simultaneously. It is possible *iff* all coefficients B^{kk}_{ij} comply with (2.19).

To prove (2.20) we consider (2.21) under $k = l$. Let m be the scheme matrix column number and δ_{mj} denote the Kronecker delta. Assuming that the scheme will be GO-monotone, it will transform $y_j = \delta_{mj}$ into a μ-function as δ_{mj} is a $\wedge$-function of j. Then we obtain from (2.21), under $k = l$, that $z^k_{k,i} = B^{kk}_{im}$. Hence, B^{kk}_{im} is a μ-function of i, $\forall k, m$. □

Consider the special case of the scheme (2.17), namely, $\mathbf{B}_{ij}$ in (2.17) depends on a square matrix $\mathbf{A}_i$

$$\mathbf{B}_{ij} = \varphi_{ij}\left(\mathbf{A}_i\right), \quad \forall i, j, \tag{2.26}$$

where $\mathbf{A}_i$ is similar [29, p. 119] to a diagonal matrix $\mathbf{\Lambda}_i$, i.e. there exists a non-singular matrix $\mathbf{S}_i$ such that

$$\mathbf{S}^{-1}_i \cdot \mathbf{A}_i \cdot \mathbf{S}_i = \mathbf{\Lambda}_i \equiv diag\left\{\lambda^1_i, \lambda^2_i, \ldots, \lambda^M_i\right\}. \tag{2.27}$$

It is assumed that $\mathbf{B}_{ij} = 0$ if $j \notin J_i \equiv \{j : \ i - k_L \leq j \leq i + k_R\}$, where k_L, $k_R = const \geq 0$. Notice, it is not assumed here that any constant (i.e., $\mathbf{z}_i = \mathbf{y}_j \equiv \mathbf{c} = const$) is bound to be a solution to (2.17). The following notation is used:

$$\mathbf{y} = \left\{\ldots, \mathbf{y}^T_j, \ldots\right\}^T, \ \overline{\mathbf{y}}_j = \mathbf{S}^{-1}_j \cdot \mathbf{y}_j, \ \overline{\mathbf{y}} = \left\{\ldots, \overline{\mathbf{y}}^T_j, \ldots\right\}^T, \ \overline{\mathbf{y}}_{i,j} = \mathbf{S}^{-1}_i \cdot \mathbf{y}_j,$$

$$\widetilde{\mathbf{y}}_i = \left\{\ldots, \overline{\mathbf{y}}^T_{i-k_L-1, i-k_L-1}, \overline{\mathbf{y}}^T_{i, i-k_L}, \ldots, \overline{\mathbf{y}}^T_{i, i+k_R}, \overline{\mathbf{y}}^T_{i+k_R+1, i+k_R+1}, \ldots\right\}^T,$$

$$\mathbf{B}=\{\mathbf{B}_{ij}\}, \quad \overline{\mathbf{B}}_{ij}=\mathbf{S}_i^{-1}\cdot\mathbf{B}_{ij}\cdot\mathbf{S}_i, \quad \overline{\mathbf{B}}_i=\left\{\ldots,\overline{\mathbf{B}}_{ij-1},\overline{\mathbf{B}}_{ij},\overline{\mathbf{B}}_{ij+1},\ldots\right\}. \tag{2.28}$$

Notice, $\overline{\mathbf{B}}_{ij}=0$ in (2.28) if $j\notin J_i$. Thus, it can be written that $\overline{\mathbf{B}}_{ij}=\mathbf{S}_i^{-1}\cdot\mathbf{B}_{ij}\cdot\mathbf{S}_j$ if $j\notin J_i$, and hence $\overline{\mathbf{y}}_{i,j}=\overline{\mathbf{y}}_{j,j}=\mathbf{S}_j^{-1}\cdot\mathbf{y}_j=\overline{\mathbf{y}}_j$ ($\forall j\notin J_i$). The stability for (2.17) provided (2.26)-(2.27) can be addressed by the following.

Lemma 2.5. *Consider an explicit scheme that can be written in the form (2.17) under (2.26), (2.27). Let $s_i=s(\mathbf{A}_i)$ be the spectrum of $\mathbf{A}_i$, and $\varphi_{ij}(\lambda)$ be represented by an absolutely convergent power series at each point $\lambda\in s_i$. Let $\mathbf{B}_{ij}=0$ in (2.17) if $j\notin J_i=\{j:\ i-k_L\le j\le i+k_R\}$. Then the scheme will be stable if*

$$\max_{\lambda\in s_i}\sum_j\left|\varphi_{ij}(\lambda)\right|\le 1, \quad \forall i, \tag{2.29}$$

$$\left\|\left(\mathbf{S}_i^{-1}-\mathbf{S}_j^{-1}\right)\cdot\mathbf{S}_j\right\|_\infty\le\Theta=const, \quad \forall i,\ \forall j\in J_i. \tag{2.30}$$

Proof. It is easy to see that

$$\mathbf{S}_i^{-1}\equiv\left\{\mathbf{I}+\left(\mathbf{S}_i^{-1}-\mathbf{S}_j^{-1}\right)\cdot\mathbf{S}_j\right\}\cdot\mathbf{S}_j^{-1}. \tag{2.31}$$

Then, in view of (2.30), we find $\forall i$, $\forall j\in J_i$

$$\left\|\mathbf{S}_i^{-1}\cdot\mathbf{y}_j\right\|_\infty\le\left\|\mathbf{I}+\left(\mathbf{S}_i^{-1}-\mathbf{S}_j^{-1}\right)\cdot\mathbf{S}_j\right\|_\infty\left\|\mathbf{S}_j^{-1}\cdot\mathbf{y}_j\right\|_\infty\le$$

$$\left\{1+\left\|\left(\mathbf{S}_i^{-1}-\mathbf{S}_j^{-1}\right)\cdot\mathbf{S}_j\right\|_\infty\right\}\left\|\mathbf{S}_j^{-1}\cdot\mathbf{y}_j\right\|_\infty\le(1+\Theta)\left\|\mathbf{S}_j^{-1}\cdot\mathbf{y}_j\right\|_\infty. \tag{2.32}$$

By virtue of (2.28), we rewrite (2.17) to read

$$\overline{\mathbf{z}}_i\equiv\mathbf{S}_i^{-1}\cdot\mathbf{z}_i=\sum_j\overline{\mathbf{B}}_{ij}\cdot\left(\mathbf{S}_i^{-1}\cdot\mathbf{y}_j\right)\equiv\overline{\mathbf{B}}_i\cdot\widetilde{\mathbf{y}}_i\equiv\left\langle\overline{\mathbf{B}}_i\right\rangle\cdot\left\langle\widetilde{\mathbf{y}}_i\right\rangle, \quad \forall i, \tag{2.33}$$

where $\left\langle\overline{\mathbf{B}}_i\right\rangle$ and $\left\langle\widetilde{\mathbf{y}}_i\right\rangle$ denote the ordinary matrix and vector obtained from $\overline{\mathbf{B}}_i$ and $\widetilde{\mathbf{y}}_i$, respectively, by removing the partitions. In view of (2.33) we obtain that

$$\left\|\overline{\mathbf{z}}_i\right\|_\infty\equiv\left\|\mathbf{S}_i^{-1}\cdot\mathbf{z}_i\right\|_\infty\le\left\|\left\langle\overline{\mathbf{B}}_i\right\rangle\right\|_\infty\left\|\left\langle\widetilde{\mathbf{y}}_i\right\rangle\right\|_\infty, \quad \forall i, \tag{2.34}$$

The norm $\left\|\left\langle\widetilde{\mathbf{y}}_i\right\rangle\right\|_\infty$ in (2.34) can be estimated by virtue of (2.16) and (2.32):

$$\left\|\left\langle\widetilde{\mathbf{y}}_i\right\rangle\right\|_\infty=\left\|\widetilde{\mathbf{y}}_i\right\|_\infty\le(1+\Theta)\max_j\left\|\mathbf{S}_j^{-1}\cdot\mathbf{y}_j\right\|_\infty=(1+\Theta)\left\|\overline{\mathbf{y}}\right\|_\infty, \ \forall i. \tag{2.35}$$

Let us estimate $\left\|\left\langle\overline{\mathbf{B}}_i\right\rangle\right\|_\infty$ in (2.34). In view of (2.27) $\boldsymbol{\Lambda}_i=\mathbf{S}_i^{-1}\cdot\mathbf{A}_i\cdot\mathbf{S}_i$. It can be verified, by induction with respect to n, that $(\boldsymbol{\Lambda}_i)^n=\mathbf{S}_i^{-1}\cdot(\mathbf{A}_i)^n\cdot\mathbf{S}_i$. Then, in view of Theorem 11.2.2 and Theorem 11.2.4 in [29], we find

$$\overline{\mathbf{B}}_{ij}\equiv\mathbf{S}_i^{-1}\cdot\mathbf{B}_{ij}\cdot\mathbf{S}_i=\varphi_{ij}\left(\mathbf{S}_i^{-1}\cdot\mathbf{A}_i\cdot\mathbf{S}_i\right)=\varphi_{ij}\left(\boldsymbol{\Lambda}_i\right). \tag{2.36}$$

Thus, $\overline{\mathbf{B}}_{ij}$ can be written in the form

$$\overline{\mathbf{B}}_{ij}=diag\left\{\Lambda_{ij}^1,\Lambda_{ij}^2,\ldots,\Lambda_{ij}^M\right\}, \quad \Lambda_{ij}^k=\varphi_{ij}\left(\lambda_j^k\right), \quad k=1,2,\ldots,M. \tag{2.37}$$

In view of (2.37), we find that

$$\left\|\langle\overline{\mathbf{B}}_i\rangle\right\|_\infty = \max_k \sum_j \left|\Lambda_{ij}^k\right| = \max_{\lambda\in s_i} \sum_j \left|\varphi_{ij}(\lambda)\right|, \quad \forall i. \tag{2.38}$$

By virtue of (2.35), (2.38), and (2.29) we obtain from (2.34) that

$$\left\|\mathbf{S}_i^{-1}\cdot\mathbf{z}_i\right\|_\infty \leq (1+\Theta)\left\|\overline{\mathbf{y}}\right\|_\infty, \quad \forall i. \tag{2.39}$$

Since

$$\left\|\overline{\mathbf{z}}\right\|_\infty \equiv \max_i \left\|\mathbf{S}_i^{-1}\cdot\mathbf{z}_i\right\|_\infty, \tag{2.40}$$

we obtain, in view of (2.39), (2.40), that

$$\left\|\mathbf{z}\right\|_* \equiv \left\|\overline{\mathbf{z}}\right\|_\infty \leq (1+\Theta)\left\|\overline{\mathbf{y}}\right\|_\infty \equiv (1+\Theta)\left\|\mathbf{y}\right\|_*. \tag{2.41}$$

The last inequality establishes Lemma 2.5. □

We consider the following explicit linear scheme on a uniform grid with time step Δt and spatial mesh size Δx

$$\mathbf{v}_i^{n+1} = \sum_j \mathbf{B}_{ij}^n\cdot\mathbf{v}_j^n, \quad n\geq 0, \tag{2.42}$$

where

$$\mathbf{B}_{ij}^n = \begin{cases} \varphi_{ij}^n\left(\mathbf{A}_j^n\right), & j\in J_i \\ 0, & j\notin J_i \end{cases}, \quad \forall i,j,n, \tag{2.43}$$

$$J_i = \{j:\ i-k_L\leq j\leq i+k_R\}, \quad k_L,k_R = const, \quad \forall i,n, \tag{2.44}$$

k_L, k_R, denote the non-negative integer constants being independent of t, x, Δx, and Δt. It is assumed that the matrix-valued function $\mathbf{A}=\mathbf{A}(x,t)$ is Lipschitz-continuous and $\mathbf{A}$ is diagonizable, i.e. for $\mathbf{A}_i^n = \mathbf{A}(x_i,t_n)$ there exists a non-singular matrix $\mathbf{S}_i^n = \mathbf{S}(x_i,t_n)$ such that

$$\left(\mathbf{S}_i^n\right)^{-1}\cdot\mathbf{A}_i^n\cdot\mathbf{S}_i^n = \mathbf{\Lambda}_i^n \equiv diag\left\{\lambda_i^{n,1},\lambda_i^{n,2},\ldots,\lambda_i^{n,M}\right\}, \quad \forall i,n. \tag{2.45}$$

Let us note that even if $\mathbf{B}_{ij}^n = \varphi_{ij}^n\left(\mathbf{A}_i^n\right)$ in (2.43) and Lemma 2.5 be valid for the linear scheme (2.42) with $\Theta = O(\Delta t)$ at every time step, the scheme (2.42) will be "locally stable" only, i.e. any growth in error is, at most, order $O(\Delta t)$ in one time step. However, we cannot, in general, show on the basis of (2.41) that

$$\left\|\mathbf{v}^{N_T}\right\|_{**} \leq C_T\left\|\mathbf{v}^0\right\|_*, \quad C_T = const, \tag{2.46}$$

where $\|\cdot\|_{**}$ and $\|\cdot\|_*$ denote some norms, $\mathbf{v}^n = \left\{\ldots,\left(\mathbf{v}_i^n\right)^T,\ldots\right\}^T$, N_T denotes the time level corresponding to time $T = N_T\Delta t$ over which we wish to compute. The reason is that the vector norm in (2.41) depends on the time level t_n, and hence we maynot apply (2.41) recursively to obtain (2.46). The stability of the system (2.42), can be addressed by the following.

THEOREM 2.6. *Consider an explicit scheme that can be written in the form (2.42) under (2.43)-(2.45), where the functions* $\varphi_{ij}^n(\mathbf{A})$ *and* $\mathbf{A}(x,t)$ *are both Lipschitz-continuous. Let there exist* $\Delta x_0 > 0$ *such that the function* $\varphi_{ij}^n(\lambda)$ *in (2.43) can be represented by an absolutely convergent power series at each point of the spectrum* $s_i^n = s(\mathbf{A}_i^n)$ $\forall i,n,$ $\forall j \in J_i,$ $\forall \Delta x \le \Delta x_0,$ *and let the matrix-valued functions* $\mathbf{S}(x,t)$*, and* $\mathbf{S}^{-1}(x,t)$ *in (2.45) can be taken such that the matrix-valued functions* $\left[(\mathbf{S}_i^n)^{-1} - (\mathbf{S}^n)^{-1}(x)\right] \cdot \mathbf{S}^n(x)$ *and* $\left[(\mathbf{S}_i)^{-1}(t) - (\mathbf{S}_i^n)^{-1}\right] \cdot \mathbf{S}_i^n$ *will be Lipschitz-continuous in space and, respectively, time* $\forall i,n$*. Let*

$$\left\| (\mathbf{S}_j^n)^{-1} \right\|_\infty \le \beta_{-1} = const, \quad \left\| \mathbf{S}_j^n \right\|_\infty \le \beta_0 = const, \quad \forall j,n. \tag{2.47}$$

Then the scheme (2.42) will be stable, i.e. (2.46) will be valid, if

$$\max_{\lambda \in s_i^n} \sum_j \left| \varphi_{ij}^n(\lambda) \right| \le 1, \quad \forall i,n. \tag{2.48}$$

Proof. Let $\check{\mathbf{B}}_{ij}^n = \varphi_{ij}^n(\mathbf{A}_i^n)$, and let us rewrite (2.42) to read

$$\mathbf{v}_i^{n+1} = \check{\mathbf{v}}_i^n + \hat{\mathbf{v}}_i^n, \quad \forall i,n, \tag{2.49}$$

where

$$\check{\mathbf{v}}_i^n = \sum_j \check{\mathbf{B}}_{ij}^n \cdot \mathbf{v}_j^n, \quad \forall i,n, \ \forall j \in J_i, \tag{2.50}$$

$$\hat{\mathbf{v}}_i^n = \sum_j \left(\mathbf{B}_{ij}^n - \check{\mathbf{B}}_{ij}^n \right) \cdot \mathbf{v}_j^n, \quad \forall i,n, \ \forall j \in J_i. \tag{2.51}$$

First, let us estimate the norm, $h_n(\cdot)$, of $\check{\mathbf{v}}^n$:

—

$$h_n(\check{\mathbf{v}}^n) \equiv \left\| \overline{\check{\mathbf{v}}}^n \right\|_\infty \equiv \max_i \left\| (\mathbf{S}_i^n)^{-1} \cdot \check{\mathbf{v}}_i^n \right\|_\infty. \tag{2.52}$$

Since $\left[(\mathbf{S}_i^n)^{-1} - (\mathbf{S}^n)^{-1}(x)\right] \cdot \mathbf{S}^n(x)$ and $\left[(\mathbf{S}_i)^{-1}(t) - (\mathbf{S}_i^n)^{-1}\right] \cdot \mathbf{S}_i^n$ are Lipschitz-continuous in space and time, respectively, we may write

$$\left\| \left[(\mathbf{S}_i^n)^{-1} - \left(\mathbf{S}_{i+1}^n\right)^{-1} \right] \cdot \mathbf{S}_{i+1}^n \right\|_\infty \le \beta_1 \Delta x, \ \boldsymbol{\beta}_1{=}const, \quad \forall i,n, \tag{2.53}$$

$$\left\| \left[\left(\mathbf{S}_i^{n+1}\right)^{-1} - (\mathbf{S}_i^n)^{-1} \right] \cdot \mathbf{S}_i^n \right\|_\infty \le \beta_2 \Delta t, \ \beta_2 = const, \quad \forall i,n. \tag{2.54}$$

By virtue of (2.53), we find

$$\left\| \left[(\mathbf{S}_i^n)^{-1} - \left(\mathbf{S}_j^n\right)^{-1} \right] \cdot \mathbf{S}_j^n \right\|_\infty \le \beta_3 \Delta x, \ \beta_3 = const, \quad \forall i,n, \ \forall j \in J_i, \tag{2.55}$$

where $\beta_3 = \beta_1 \max(k_L, k_R)$. In view of the CFL condition [26], we assume for the explicit scheme (2.42), that $\Delta x = O(\Delta t)$ (i.e. $\exists\ \Delta t_0 > 0$, $\exists\ \alpha_0 > 0$ such that $\Delta x \le \alpha_0 \Delta t$ $\forall \Delta t \le \Delta t_0$). Then we find by virtue of (2.55) that

$$\left\| \left\{ (\mathbf{S}_i^n)^{-1} - \left(\mathbf{S}_j^n\right)^{-1} \right\} \cdot \mathbf{S}_j^n \right\|_\infty \le \beta_4 \Delta t, \ \beta_4 = \alpha_0 \beta_3, \ \forall i,n, \ \forall j \in J_i. \tag{2.56}$$

The inequality (2.56) coincides with the assumption (2.30) in Lemma 2.5 under $\Theta = \beta_4 \Delta t$. Then, in view of Lemma 2.5, we obtain for the scheme (2.50), that

$$h_n\left(\check{\mathbf{v}}^n\right) \equiv \left\|\overline{\check{\mathbf{v}}}^n\right\|_\infty \equiv \max_i \left\|\left(\mathbf{S}_i^n\right)^{-1} \cdot \check{\mathbf{v}}_i^n\right\|_\infty \leq$$

$$[1+\beta_4\Delta t] \max_i \left\|\left(\mathbf{S}_i^n\right)^{-1} \cdot \mathbf{v}_i^n\right\|_\infty \equiv [1+\beta_4\Delta t]\, h_n\left(\mathbf{v}^n\right). \tag{2.57}$$

Let us now estimate the norm $h_n\left(\hat{\mathbf{v}}^n\right)$. Since $\varphi_{ij}^n\left(\mathbf{A}\right)$, $\mathbf{A}\left(x,t\right)$ are both Lipschitz continuous, we may write

$$\left\|\varphi_{ij}^n\left(\mathbf{A}_j^n\right) - \varphi_{ij}^n\left(\mathbf{A}_i^n\right)\right\|_\infty \leq \alpha_1 \left\|\mathbf{A}_i^n - \mathbf{A}_j^n\right\|_\infty, \quad \forall i,n,\ \forall j \in J_i, \tag{2.58}$$

$$\left\|\mathbf{A}_i^n - \mathbf{A}_j^n\right\|_\infty \leq \alpha_2 \left|x_j - x_i\right| \leq \alpha_3 \Delta x, \quad \forall i,n,\ \forall j \in J_i, \tag{2.59}$$

where $\alpha_3 = \alpha_2 \max(k_L, k_R)$, α_1, $\alpha_2 = const$. By virtue of (2.58), (2.59), and assuming that $\Delta x = O\left(\Delta t\right)$, we obtain

$$\left\|\mathbf{B}_{ij}^n - \check{\mathbf{B}}_{ij}^n\right\|_\infty \equiv \left\|\varphi_{ij}^n\left(\mathbf{A}_j^n\right) - \varphi_{ij}^n\left(\mathbf{A}_i^n\right)\right\|_\infty \leq \alpha_4 \Delta t, \quad \forall i,n,\ \forall j \in J_i, \tag{2.60}$$

where $\alpha_4 = \alpha_0 \alpha_1 \alpha_3 = const$. We obtain from (2.51) that

$$\left(\mathbf{S}_i^n\right)^{-1} \cdot \hat{\mathbf{v}}_i^n = \sum_j \left(\mathbf{S}_i^n\right)^{-1} \cdot \left(\mathbf{B}_{ij}^n - \check{\mathbf{B}}_{ij}^n\right) \cdot \mathbf{S}_j^n \cdot \left(\mathbf{S}_j^n\right)^{-1} \cdot \mathbf{v}_j^n. \tag{2.61}$$

Whence, by virtue of (2.60) and (2.47), we obtain

$$\left\|\left(\mathbf{S}_i^n\right)^{-1} \cdot \hat{\mathbf{v}}_i^n\right\|_\infty \leq \alpha_5 \Delta t \max_j \left\|\left(\mathbf{S}_j^n\right)^{-1} \cdot \mathbf{v}_j^n\right\|_\infty \equiv \alpha_5 \Delta t h_n\left(\mathbf{v}^n\right), \quad \forall i,n, \tag{2.62}$$

where $\alpha_5 = \beta_{-1}\beta_0\alpha_4 \max(k_L, k_R) = const$. By virtue of (2.49), (2.57), and (2.62), we obtain

$$h_n\left(\mathbf{v}^{n+1}\right) \leq [1+\beta\Delta t]\, h_n\left(\mathbf{v}^n\right),\ \beta = \beta_4 + \alpha_5 = const, \quad \forall n. \tag{2.63}$$

It is easy to see that

$$\left(\mathbf{S}_i^{n+1}\right)^{-1} \equiv \left\{\mathbf{I} + \left(\left(\mathbf{S}_i^{n+1}\right)^{-1} - \left(\mathbf{S}_i^n\right)^{-1}\right) \cdot \mathbf{S}_i^n\right\} \cdot \left(\mathbf{S}_i^n\right)^{-1}, \tag{2.64}$$

whence, by virtue of (2.54), we find

$$h_{n+1}\left(\mathbf{v}^{n+1}\right) = \max_i \left\|\left(\mathbf{S}_i^{n+1}\right)^{-1} \cdot \mathbf{v}_i^{n+1}\right\|_\infty \leq$$

$$\max_i \left\|\mathbf{I} + \left(\left(\mathbf{S}_i^{n+1}\right)^{-1} - \left(\mathbf{S}_i^n\right)^{-1}\right) \cdot \mathbf{S}_i^n\right\|_\infty \left\|\left(\mathbf{S}_i^n\right)^{-1} \cdot \mathbf{v}_i^{n+1}\right\|_\infty \leq$$

$$\left(1+\beta_2\Delta t\right) \max_i \left\|\left(\mathbf{S}_i^n\right)^{-1} \cdot \mathbf{v}_i^{n+1}\right\|_\infty = \left(1+\beta_2\Delta t\right) h_n\left(\mathbf{v}^{n+1}\right). \tag{2.65}$$

In view of (2.63) and (2.65), we find

$$h_{n+1}\left(\mathbf{v}^{n+1}\right) \leq \left[1+\gamma\Delta t\right]^2 h_n\left(\mathbf{v}^n\right),\ \gamma = \max\left(\alpha, \beta_2\right), \quad \forall n. \tag{2.66}$$

Applying (2.66) recursively gives

$$h_{N_T}\left(\mathbf{v}^{N_T}\right)\leq(1+\gamma\Delta t)^{2N_T}h_0\left(\mathbf{v}^0\right)\leq C_T h_0\left(\mathbf{v}^0\right),\quad C_T=\exp\left(2\gamma T\right).\tag{2.67}$$

The inequalities in (2.67) prove the theorem. □

Remark 2.7. *Theorem 2.6 can be generalized for the case when $\varphi_{ij}^n(\lambda)$ in (2.43) can be represented by a convergent Laurent series.*

Let us consider the case when the operator $\mathbf{B}_{ij}^n$ in (2.42) depends on a matrix $\mathbf{A}_j^n$ belonging to a set of pairwise commutative diagonizable matrices:

$$\mathbf{B}_{ij}^n=\varphi_{ij}^n\left(\mathbf{A}_j^n\right),\ \mathbf{A}_j^n\cdot\mathbf{A}_k^m=\mathbf{A}_k^m\cdot\mathbf{A}_j^n,\quad\forall i,j,n,k,m.\tag{2.68}$$

In such a case, the S-monotonicity of the system (2.42), can be addressed by the following.

Theorem 2.8. *Consider an explicit scheme that can be written in the form (2.42) provided (2.68). Let $\varphi_{ij}^n(\lambda)$ in (2.68) can be represented by an absolutely convergent power series at each point of the spectrum $s_j^n=s\left(\mathbf{A}_j^n\right)$ $\forall i,j,n$. Then the scheme (2.42) will be S-monotone* iff

$$\max_{\lambda\in s_j^n}\sum_i\left|\varphi_{ij}^n(\lambda)\right|\leq 1,\quad\forall j,n.\tag{2.69}$$

Proof. As $\mathbf{A}_j^n$ belongs to the set of pair-wise permutable diagonizable matrices, the matrices of the set are simultaneously similar to diagonal matrices [29, p. 318], i.e., there exists a non-singular matrix $\mathbf{S}$ such that

$$\mathbf{S}^{-1}\cdot\mathbf{A}_j^n\cdot\mathbf{S}=\mathbf{\Lambda}_j^n\equiv diag\left\{\lambda_j^{n,1},\lambda_j^{n,2},\ldots,\lambda_j^{n,M}\right\},\quad\forall j,n.\tag{2.70}$$

where $\lambda_j^{n,m}$ denotes the m-th eigenvalue of $\mathbf{A}_j^n$. The following notation is used:

$$\overline{\mathbf{v}}_j^n=\mathbf{S}^{-1}\cdot\mathbf{v}_j^n,\ \overline{\mathbf{B}}_{ij}^n=\mathbf{S}^{-1}\cdot\mathbf{B}_{ij}^n\cdot\mathbf{S},\ \overline{\mathbf{B}}^n=\left\{\overline{\mathbf{B}}_{ij}^n\right\},\ \mathbf{B}^n=\left\{\mathbf{B}_{ij}^n\right\}.\tag{2.71}$$

By virtue of (2.71), we rewrite (2.42) to read

$$\overline{\mathbf{v}}_i^{n+1}\equiv\mathbf{S}^{-1}\cdot\mathbf{v}_i^{n+1}=\sum_j\overline{\mathbf{B}}_{ij}^n\cdot\overline{\mathbf{v}}_j^n.\tag{2.72}$$

Using $\overline{\mathbf{v}}^n\equiv\left\{\ldots,\left(\overline{\mathbf{v}}_j^n\right)^T,\ldots\right\}^T$, we rewrite (2.72) to read

$$\overline{\mathbf{v}}^{n+1}=\overline{\mathbf{B}}^n\cdot\overline{\mathbf{v}}^n,\quad\overline{\mathbf{v}}^{n+1}\equiv\left\{\ldots,\left(\overline{\mathbf{v}}_i^{n+1}\right)^T,\ldots\right\}^T.\tag{2.73}$$

In view of (2.73) we obtain that

$$h\left(\mathbf{v}^{n+1}\right)\equiv\left\|\left\langle\overline{\mathbf{v}}^{n+1}\right\rangle\right\|_1\leq\left\|\left\langle\overline{\mathbf{B}}^n\right\rangle\right\|_1\left\|\left\langle\overline{\mathbf{v}}^n\right\rangle\right\|_1\equiv h\left(\mathbf{v}^n\right),\tag{2.74}$$

where $\left\langle\overline{\mathbf{B}}^n\right\rangle$ and $\left\langle\overline{\mathbf{v}}^n\right\rangle$ denote the ordinary matrix and vector obtained from $\overline{\mathbf{B}}^n$ and $\overline{\mathbf{v}}^n$, respectively, by removing the partitions. Let us estimate the norm of $\left\langle\overline{\mathbf{B}}^n\right\rangle$ in (2.74). Since $\varphi_{ij}^n(\lambda)$ can be represented by an absolutely convergent power series at each point $\lambda \in s_j^n = s\left(\mathbf{A}_j^n\right)$, we find, in view of Theorem 11.2.2 and Theorem 11.2.4 in [29], that

$$\overline{\mathbf{B}}_{ij}^{n} \equiv \mathbf{S}^{-1} \cdot \mathbf{B}_{ij}^{n} \cdot \mathbf{S} = \varphi_{ij}^{n}\left(\mathbf{S}^{-1} \cdot \mathbf{A}_{j}^{n} \cdot \mathbf{S}\right) = \varphi_{ij}^{n}\left(\mathbf{\Lambda}_{j}^{n}\right). \tag{2.75}$$

Thus, $\overline{\mathbf{B}}_{ij}^{n}$ can be written in the form

$$\overline{\mathbf{B}}_{ij} = diag\left\{\Lambda_{ij}^{n,1}, \Lambda_{ij}^{n,2}, \ldots, \Lambda_{ij}^{n,M}\right\},\ \ \Lambda_{ij}^{n,k} = \varphi_{ij}^{n}\left(\lambda_{j}^{n,k}\right),\ \ k = 1, 2, \ldots, M. \tag{2.76}$$

In view of (2.76), we obtain that

$$\left\|\left\langle \overline{\mathbf{B}}^{n}\right\rangle\right\|_{1} = \max_{j}\left(\max_{k=1,\ldots,M}\sum_{i}\left|\Lambda_{ij}^{n,k}\right|\right) = \max_{j}\left(\max_{\lambda\in s_{j}^{n}}\sum_{i}\left|\varphi_{ij}^{n}\left(\lambda\right)\right|\right). \tag{2.77}$$

Whence, in view of (2.69), we find

$$\left\|\left\langle \overline{\mathbf{B}}^{n}\right\rangle\right\|_{1} \leq 1, \quad \forall n. \tag{2.78}$$

The vector norm $h\left(\cdot\right)$ in (2.74) does not depend on time level. Then, in view of Proposition 3.2 in [5], the inequality (2.78) proves Theorem 2.8 □

Remark 2.9. *Theorem 2.8 can be generalized for the case when $\varphi_{ij}^{n}\left(\lambda\right)$ in (2.68) can be represented by a convergent Laurent series. Moreover, this theorem can be generalized for the case when the operator $\mathbf{B}_{ij}^{n}$ in (2.42) depends on several pair-wise commutative diagonizable matrices (analogous theorems for normal matrices are proven in [4], [5]).*

Proposition 2.10. *If (2.17) is a variational scheme, then (2.18) is, in general, not valid. Notice, Lemma 2.5 and Theorem 2.6 are proven without assumption (2.18). However, in addition to the Lipschitz-continuity of $\mathbf{A}\left(x,t\right)$ (see (2.26) and (2.43)), it is assumed in Lemma 2.5 and Theorem 2.6 that some functions of $\mathbf{S}\left(x,t\right)$ (see (2.27), (2.45)) are also Lipschitz-continuous. Let us note that the stability of a linear scheme can often be proven without assumption (2.18) as well as without additional assumptions on the continuity. To demonstrate it, let us generalize the theorem of Friedrichs (1954) (see, e.g., [40, p. 120], [43, p. 374]) to be applicable to variational schemes. Following Friedrichs, we consider the following difference scheme*

$$\mathbf{y}_{n+1}\left(x\right) = \sum_{k=-m}^{m}\mathbf{B}_{k}\left(x\right)\cdot\mathbf{y}_{n}\left(x + k\Delta x\right), \quad x \in \left(-\infty, \infty\right), \tag{2.79}$$

where $\mathbf{y}_n \in \mathbb{R}^M$, $\mathbf{B}_k \in \mathbb{R}^{M\times M}$ is a symmetric and non-negative matrix, $\mathbf{y}_n\left(x\right)$ and $\mathbf{B}_k\left(x\right)$ are periodic (with the period equal to 1) functions of x. In view of the CFL condition [26], it is assumed that there exists $\alpha_0 = const$ such that $\Delta x \leq \alpha_0 \Delta t$ for a sufficiently small Δt. Let $\mathbf{F}^k \equiv \mathbf{F}\left(x + k\Delta x\right)$, and let

$$\left(\mathbf{u}, \mathbf{v}^{k}\right) \equiv \int_{0}^{1}\left[\mathbf{u}^{T}\left(x\right)\cdot\mathbf{v}\left(x + k\Delta x\right)\right]dx, \quad \left\|\mathbf{u}\right\| \equiv \sqrt{\left(\mathbf{u}, \mathbf{u}\right)}. \tag{2.80}$$

If there exist c_1, $c_2 = const$ such that

$$\left\|\sum_{k}\mathbf{B}_{k}\right\| \leq 1 + c_{1}\Delta x, \quad \left\|\mathbf{B}_{k}^{k} - \mathbf{B}_{k}\right\| \leq \frac{c_{2}}{2m+1}\Delta x, \tag{2.81}$$

then the scheme is stable. Notice, it is not assumed that $\sum_k \mathbf{B}_k(x) = \mathbf{I}$.

The proof is very little different from the proof when $\sum_k \mathbf{B}_k(x) = \mathbf{I}$. *Actually, in view of (2.79) and the first inequality in (2.81), we obtain*

$$\|\mathbf{y}_{n+1}\|^2 = \sum_k \left(\mathbf{y}_{n+1}, \mathbf{B}_k \cdot \mathbf{y}_n^k\right) \leq \sum_k \left|\left(\mathbf{y}_{n+1}, \mathbf{B}_k \cdot \mathbf{y}_n^k\right)\right| \leq$$

$$0.5 \sum_k \left(\mathbf{B}_k \cdot \mathbf{y}_n^k, \mathbf{y}_n^k\right) + 0.5\left(1 + c_1 \Delta x\right)\|\mathbf{y}_{n+1}\|^2 . \tag{2.82}$$

Since $\mathbf{y}_n(x)$ *and* $\mathbf{B}_k(x)$ *are periodic functions, we obtain from (2.82) that*

$$(1 - \alpha_0 c_1 \Delta t)\|\mathbf{y}_{n+1}\|^2 \leq \sum_k \left(\mathbf{B}_k^k \cdot \mathbf{y}_n^k, \mathbf{y}_n^k\right) - \sum_k \left(\left(\mathbf{B}_k^k - \mathbf{B}_k\right) \cdot \mathbf{y}_n^k, \mathbf{y}_n^k\right) \leq$$

$$\sum_k \left(\mathbf{B}_k \cdot \mathbf{y}_n, \mathbf{y}_n\right) + \sum_k \left|\left(\left(\mathbf{B}_k^k - \mathbf{B}_k\right) \cdot \mathbf{y}_n^k, \mathbf{y}_n^k\right)\right| . \tag{2.83}$$

By virtue of (2.81) and (2.83), we find

$$\|\mathbf{y}_{n+1}\|^2 \leq \frac{1 + \alpha_0 (c_1 + c_2)\Delta t}{1 - \alpha_0 c_1 \Delta t}\|\mathbf{y}_n\|^2 . \tag{2.84}$$

Let $\Delta t_0 = const$ *such that* $1 - \alpha_0 c_1 \Delta t_0 > 0$, *e.g.* $\Delta t_0 = 0.5 / (\alpha_0 c_1)$. *Then, for all* $\Delta t < \Delta t_0$ *the following inequality will be valid*

$$\|\mathbf{y}_{n+1}\|^2 \leq (1 + c_3 \Delta t)\|\mathbf{y}_n\|^2 , \quad c_3 = 2\alpha_0 (2c_1 + c_2) = const. \tag{2.85}$$

The inequality in (2.85) proves Proposition 2.10.

3. Monotone C^1 piecewise cubics in construction of central schemes. In this section we consider some theoretical aspects for high-order interpolation and employment of monotone C^1 piecewise cubics (e.g., [9], [23]) in construction of monotone central schemes. By virtue of the operator-splitting idea (see also LOS in [42]), the following chain of equations corresponds to the problem (1.1)

$$\frac{1}{2}\frac{\partial \mathbf{u}}{\partial t} + \frac{\partial}{\partial x}\mathbf{f}(\mathbf{u}) = 0, \quad t_n < t \leq t_{n+0.5}, \quad \mathbf{u}(x, t_n) = \mathbf{u}^n(x), \tag{3.1}$$

$$\frac{1}{2}\frac{\partial \mathbf{u}}{\partial t} = \frac{1}{\tau}\mathbf{q}(\mathbf{u}), \quad t_{n+0.5} < t \leq t_{n+1}, \quad \mathbf{u}(x, t_{n+0.5}) = \mathbf{u}^{n+0.5}(x). \tag{3.2}$$

Using the central differencing, we write

$$\left.\frac{\partial \mathbf{u}}{\partial t}\right|_{t=t_{n+0.125},\ x=x_{i+0.5}} = \frac{\mathbf{u}_{i+0.5}^{n+0.25} - \mathbf{u}_{i+0.5}^{n}}{0.25\Delta t} + O\left((\Delta t)^2\right), \tag{3.3}$$

$$\left.\frac{\partial \mathbf{f}}{\partial x}\right|_{t=t_{n+0.125},\ x=x_{i+0.5}} = \frac{\mathbf{f}_{i+1}^{n+0.125} - \mathbf{f}_{i}^{n+0.125}}{\Delta x} + O\left((\Delta x)^2\right). \tag{3.4}$$

By virtue of (3.3)-(3.4) we approximate (3.1) on the cell $[x_i, x_{i+1}] \times [t_n, t_{n+0.25}]$ by the following difference equation

$$\mathbf{v}_{i+0.5}^{n+0.25} = \mathbf{v}_{i+0.5}^{n} - \frac{\Delta t}{2\Delta x}\left(\mathbf{g}_{i+1}^{n+0.125} - \mathbf{g}_{i}^{n+0.125}\right), \tag{3.5}$$

where $\mathbf{v}_{i+\alpha}^{n+\beta}$, $\mathbf{g}_{i+\alpha}^{n+\beta}$ are the grid functions. In perfect analogy, we obtain on the cell $[x_{i-0.5}, x_{i+0.5}] \times [t_{n+0.25}, t_{n+0.5}]$ that

$$\mathbf{v}_{i}^{n+0.5} = \mathbf{v}_{i}^{n+0.25} - \frac{\Delta t}{2\Delta x}\left(\mathbf{g}_{i+0.5}^{n+0.375} - \mathbf{g}_{i-0.5}^{n+0.375}\right). \tag{3.6}$$

As usually, the mathematical treatments for the second step of the staggered scheme (3.5)-(3.6) will, in general, not be included in the text, because (3.6) is quite similar to (3.5).

Considering that (3.5) approximate (3.1) with the accuracy $O\left((\Delta x)^2 + (\Delta t)^2\right)$, the next problem is to approximate $\mathbf{v}_{i+0.5}^{n}$ and $\mathbf{g}_{i}^{n+0.125}$ in such a way as to retain the accuracy of the approximation. For instance, the following approximations

$$\mathbf{v}_{i+0.5}^{n} = 0.5\left(\mathbf{v}_{i}^{n} + \mathbf{v}_{i+1}^{n}\right) + O\left((\Delta x)^2\right), \quad \mathbf{g}_{i}^{n+0.125} = \mathbf{f}\left(\mathbf{v}_{i}^{n}\right) + O\left(\Delta t\right), \tag{3.7}$$

leads to the staggered form of the famed LxF scheme that is of the first-order approximation (see, e.g., [25], [33]). One way to obtain a higher-order scheme is to use a higher order interpolation. At the same time it is required of the interpolant to be monotonicity preserving. Notice, the classic cubic spline does not possess such a property (see Figure 3.1a). Let us consider the problem of high-order interpolation of $\mathbf{v}_{i+0.5}^{n}$ in (3.5) with closer inspection

Let $\mathbf{p} = \mathbf{p}(x) \equiv \left\{p^1(x), \ldots, p^k(x), \ldots, p^m(x)\right\}^T$ be a component-wise monotone C^1 piecewise cubic interpolant (e.g., [9], [23]), and let

$$\mathbf{p}_i = \mathbf{p}(x_i), \quad \mathbf{p}_i' = \mathbf{p}'(x_i), \quad \Delta\mathbf{p}_i = \mathbf{p}_{i+1} - \mathbf{p}_i,$$

$$\mathbf{p}_i' = \mathbb{A}_i \cdot \frac{\Delta\mathbf{p}_i}{\Delta x}, \quad \mathbf{p}_{i+1}' = \mathbb{B}_i \cdot \frac{\Delta\mathbf{p}_i}{\Delta x}, \tag{3.8}$$

where $\mathbf{p}_i'$ denotes the derivative of the interpolant at $x = x_i$. The diagonal matrices $\mathbb{A}_i$ and $\mathbb{B}_i$ in (3.8) are defined as follows

$$\mathbb{A}_i = diag\left\{\alpha_i^1, \alpha_i^2, \ldots, \alpha_i^m\right\}, \ \mathbb{B}_i = diag\left\{\beta_i^1, \beta_i^2, \ldots, \beta_i^m\right\}. \tag{3.9}$$

The cubic interpolant, $\mathbf{p} = \mathbf{p}(x)$, is component-wise monotone on $[x_i, x_{i+1}]$ *iff* one of the following conditions (e.g., [9], [23]) is satisfied:

$$\left(\alpha_i^k - 1\right)^2 + \left(\alpha_i^k - 1\right)\left(\beta_i^k - 1\right) + \left(\beta_i^k - 1\right)^2 - 3\left(\alpha_i^k + \beta_i^k - 2\right) \leq 0, \tag{3.10}$$

$$\alpha_i^k + \beta_i^k \leq 3, \quad \alpha_i^k \geq 0, \ \beta_i^k \geq 0, \quad \forall i, k. \tag{3.11}$$

As reported in [23], the necessary and sufficient conditions for monotonicity of a C^1 piecewise cubic interpolant originally given by Ferguson and Miller (1969), and independently, by Fritsch and Carlson [9]. The region of monotonicity is shown in

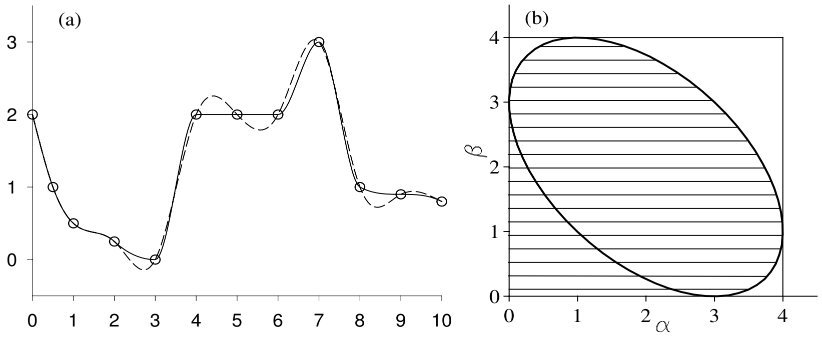

FIG. 3.1. *Monotone piecewise cubic interpolation. (a) Interpolation of a 1-D tabulated function. Circles: prescribed tabulated values; Dashed line: classic cubic spline; Solid line: monotone piecewise cubic. (b) Necessary and sufficient conditions for monotonicity. Horizontal hatching: region of monotonicity; Unshaded: cubic is non-monotone.*

Figure 3.1b. The results of implementing a monotone C^1 piecewise cubic interpolation when compared with the classic cubic spline interpolation, are depicted in Figure 3.1a. We note (Figure 3.1a) that the constructed function produces monotone interpolation and this function coincides with the classic cubic spline at some sections where the classic cubic spline is monotone.

Using the cubic segment of the C^1 piecewise cubic interpolant, $\mathbf{p} = \mathbf{p}\left(x\right)$, (see, e.g., [9], [23]) for $x \in [x_i, x_{i+1}]$, we obtain the following interpolation formula

$$\mathbf{p}_{i+0.5} = 0.5\left(\mathbf{p}_i + \mathbf{p}_{i+1}\right) - \frac{\Delta x}{8}\left(\mathbf{p}'_{i+1} - \mathbf{p}'_i\right) + O\left(\left(\Delta x\right)^r\right). \tag{3.12}$$

If $\mathbf{p}\left(x\right)$ has a continuous fourth derivative, then $r = 4$ in (3.12), see e.g. [21, p. 111]. However, the exact value of $\mathbf{p}'_i$ in (3.12) is, in general, unknown, and hence to construct numerical schemes, employing formulae similar to (3.12), the value of derivatives $\mathbf{p}'_i$ must be estimated.

Using (3.12) and the second formula in (3.7) we obtain from (3.5) the following scheme

$$\mathbf{v}_{i+0.5}^{n+0.25} = 0.5\left(\mathbf{v}_i^n + \mathbf{v}_{i+1}^n\right) - \frac{\Delta x}{8}\left(\mathbf{d}_{i+1}^n - \mathbf{d}_i^n\right) - \frac{\Delta t}{2}\frac{\mathbf{f}\left(\mathbf{v}_{i+1}^n\right) - \mathbf{f}\left(\mathbf{v}_i^n\right)}{\Delta x}, \tag{3.13}$$

where $\mathbf{d}_i^n$ denotes the derivative of the interpolant at $x = x_i$. In view of (3.12) and the second formula in (3.7), the local truncation error [26, p. 142], ψ, on a sufficiently smooth solution $\mathbf{u}(x,t)$ to (3.1) is found to be

$$\psi = O\left(\Delta t\right) + O\left(\frac{\left(\Delta x\right)^r}{\Delta t}\right) + O\left(\left(\Delta t\right)^2 + \left(\Delta x\right)^2\right). \tag{3.14}$$

In view of (3.14) we conclude that the scheme (3.13) generates a conditional approximation, because it approximates (3.1) only if $\left(\Delta x\right)^r \diagup \Delta t \to 0$ as $\Delta x \to 0$ and $\Delta t \to 0$. Let $\mathbf{d}_i^n$ be approximated with the accuracy $O\left(\left(\Delta x\right)^s\right)$, then the value of r in (3.14) can be calculated (see Section 5.1, Proposition 5.1) by the following formula

$$r = \min\left(4, s+1\right). \tag{3.15}$$

Interestingly, since (3.13) provides the conditional approximation, the order of accuracy depends on the pathway taken by Δx and Δt as $\Delta x \to 0$ and $\Delta t \to 0$. Actually, there exists a pathway such that Δt is proportional to $(\Delta x)^{\mu}$ and the CFL condition is fulfilled provided $\mu \geq 1$ and $\Delta x \leq \Delta x_0$, where Δx_0 is a positive value. If we take $\mu = 1$ and $s \geq 1$, then we obtain from (3.14) that the scheme (3.13) is of the first-order. If $\mu = 2$ and $s \geq 3$, then (3.13) is of the second-order. However, if $\mu = 2$ and $s = 2$, then, in view of (3.14) and (3.15), the scheme (3.13) is of the first-order. Moreover, under $\mu = 2$ and $s = 2$, the scheme will be of the first-order even if $\mathbf{g}_i^{n+0.125}$ in (3.5) will be approximated with the accuracy $O((\Delta t)^2)$. It seems likely that Example 6 in [25] can be seen as an illustration of the last assertion. The Nessyahu-Tadmor (NT) scheme with the second-order approximation of $\mathbf{d}_i^n$ is used [25] to solve a Burgers-type equation. Since $\Delta t = O((\Delta x)^2)$ [25], the NT scheme is of the first-order, and hence it can be the main reason for the scheme to exhibit the smeared discontinuity computed in [25, Fig. 6.22].

The approximation of derivatives $\mathbf{p}_i'$ can be done by the following three steps [9]: (i) an initialization of the derivatives $\mathbf{p}_i'$; (ii) the choice of subregion of monotonicity; (iii) modification of the initialized derivatives $\mathbf{p}_i'$ to produce a monotone interpolant.

The matter of initialization of the derivatives is the most subtle issue of this algorithm. Actually, the approximation of $\mathbf{p}_i'$ must, in general, be done with accuracy $O((\Delta x)^3)$ to obtain the second-order scheme when Δt is proportional to $(\Delta x)^2$, inasmuch as central schemes generate a conditional approximation. Thus, using the two-point or the three-point (centered) difference formula (e.g. [23], [36]) we obtain, in general, the first-order scheme. The so called limiter functions [23] lead, in general, to a low-order scheme as these limiters are often $O(\Delta x)$ or $O((\Delta x)^2)$ accurate. Performing the initialization of the derivatives $\mathbf{p}_i'$ in the interpolation formula (3.12) by the classic cubic spline interpolation [39], we obtain the approximation, which is $O((\Delta x)^3)$ accurate (e.g., [21], [23]), and hence, in general, the second-order scheme. The same accuracy, $O((\Delta x)^3)$, can be achieved by using the four-point approximation [23].

Obviously, for each interval $[x_i, x_{i+1}]$ in which the initialized derivatives $\mathbf{p}_i'$, $\mathbf{p}_{i+1}'$ such that at least one point (α_i^k, β_i^k) does not belong to the region of monotonicity (3.10)-(3.11), the derivatives $\mathbf{p}_i'$, $\mathbf{p}_{i+1}'$ must be modified to $\widetilde{\mathbf{p}}_i'$, $\widetilde{\mathbf{p}}_{i+1}'$ such that the point $(\widetilde{\alpha}_i^k, \widetilde{\beta}_i^k)$ will be in the region of monotonicity. The modification of the initialized derivatives, would be much simplified if we take a square as a subregion of monotonicity. Specifically, Fritsch and Carlson [9] gave an effective algorithm for constructing the monotone piecewise cubic for several subregions including the square: $0 \leq \alpha_i, \beta_i \leq 3$. In connection with this, we will make use the subregions of monotonicity represented in the following form:

$$0 \leq \alpha_i^k \leq 4\aleph, \quad 0 \leq \beta_i^k \leq 4\aleph, \quad \forall i, k, \tag{3.16}$$

where $\aleph$ is a monotonicity parameter. Obviously, the condition (3.16) is sufficient for the monotonicity (see Figure 3.1b) provided that $0 \leq \aleph \leq 0.75$.

Let us now find necessary and sufficient conditions for (3.12) to be G-monotone. By virtue of (3.8), the interpolation formula (3.12) can be rewritten to read

$$\mathbf{p}_{i+0.5} = \left(0.5\mathbf{I} + \frac{\mathbb{B}_i - \mathbb{A}_i}{8}\right) \cdot \mathbf{p}_i + \left(0.5\mathbf{I} - \frac{\mathbb{B}_i - \mathbb{A}_i}{8}\right) \cdot \mathbf{p}_{i+1}. \tag{3.17}$$

The coefficients of (3.17) will be non-negative *iff* $|\beta_i - \alpha_i| \leq 4$. Hence (3.12) will be G-monotone *iff* (3.16) will be valid provided $0 \leq \aleph \leq 1$. Notice, there is no any contradiction between the sufficient conditions, (3.16) provided $0 \leq \aleph \leq 0.75$, for the interpolant, $\mathbf{p} = \mathbf{p}(x)$, to be monotone through the interval $[x_i, x_{i+1}]$, and the necessary and sufficient conditions, (3.16) provided $0 \leq \aleph \leq 1$, for the scheme (3.17) to be G-monotone. In the latter case the interpolant, $\mathbf{p} = \mathbf{p}(x)$, may, in general, be non-monotone, however at the point $i + 0.5$ the value of an arbitrary component of $\mathbf{p}_{i+0.5}$ will be between the corresponding components of $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$.

To fulfill the conditions of monotonicity (3.16), the modification of derivatives $\mathbf{p}'_i = \left\{p'^1_i, p'^2_i, \ldots, p'^m_i\right\}$ can be done by the following algorithm suggested, in fact, by Fritsch and Carlson [9] (see also [23]):

$$S^k_i := 4\aleph \min\mathrm{mod}(\Delta^k_{i-1}, \Delta^k_i), \quad \widetilde{p'}^k_i := \min\mathrm{mod}(p'^k_i, S^k_i), \quad \aleph = const, \tag{3.18}$$

where $\Delta^k_i = \left(p^k_{i+1} - p^k_i\right) \big/ \Delta x$, the function $\min\mathrm{mod}(x, y)$ is defined (e.g., [23], [25], [31], [36], [44]) as follows

$$\min\mathrm{mod}(x, y) \equiv \frac{1}{2}\left[sgn(x) + sgn(y)\right] \min\left(|x|, |y|\right). \tag{3.19}$$

4. Concluding remarks. The advantage of Theorem 2.2 (and Theorem 2.1) is that the scheme in variations corresponding to a non-linear scheme will always be linear, and hence the stability theory (e.g., [11], [40], [43]) for linear equation systems can be applied to establish stability of the non-linear scheme. Investigation of the stability and monotonicity for non-linear schemes is reduced by Theorems 2.2, 2.1, and 2.6 to the sensitivity analysis for the eigenvectors of the Jacobian matrix of $\mathbf{f}(\mathbf{u})$ in (1.1). It should be pointed out that Theorem 2.6 may be used to analyze schemes approximating (1.1) only if the operators depending on the eigenvectors of the Jacobian matrix will be Lipschitz-continuous in space and time. Theorem 2.8 is proven without any assumptions on continuity of the coefficients, $\mathbf{B}^n_{ij}$, in scheme (2.42). It is important, e.g., in the case when variational schemes are used in the stability and monotonicity analysis of non-linear difference schemes approximating hyperbolic PDE systems. The generalization of Friedrichs' theorem developed in Proposition 2.10 gives a possibility to investigate the stability of a non-linear scheme via its scheme in variations without the sensitivity analysis for the eigenvectors of the Jacobian matrix.

Central schemes provide a conditional approximation, and hence the order of accuracy depends on the pathway taken by Δx and Δt as $\Delta x \to 0$ and $\Delta t \to 0$. Such a peculiarity of a central scheme may decrease (or increase) the order of its accuracy.

The theoretical investigation of a possibility to use monotone C^1 piecewise cubics in construction of central schemes revealed that the range of values for the monotonicity parameter $\aleph$ is the segment $0 \leq \aleph \leq 1$, i.e. the entire square shown in Figure 3.1b.

5. Appendix. PROPOSITION 5.1. *Let us find the order of accuracy, r, in (3.12) if d_i will be approximated by $\widetilde{d}_i$ with the order of accuracy s, i.e. let*

$$d_i = \widetilde{d}_i + O\left((\Delta x)^s\right). \tag{5.1}$$

Let $U(x)$ be sufficiently smooth, then we can write

$$U_{i+1} = U_{i+05} + U'_{i+05}\frac{\Delta x}{2} + \frac{1}{2}U''_{i+05}\left(\frac{\Delta x}{2}\right)^2 + O\left((\Delta x)^3\right), \tag{5.2}$$

$$U_i = U_{i+05} - U'_{i+05}\frac{\Delta x}{2} + \frac{1}{2}U''_{i+05}\left(\frac{\Delta x}{2}\right)^2 + O\left((\Delta x)^3\right). \tag{5.3}$$

Combining the equalities (5.2) and 5.3 we obtain

$$U_{i+1}+U_i=2U_{i+05}+\left.\frac{\partial^2 U}{\partial x^2}\right|_{i+05}\left(\frac{\Delta x}{2}\right)^2+O\left((\Delta x)^3\right). \tag{5.4}$$

In a similar manner we write:

$$d_{i+1}=U'_{i+05}+U''_{i+05}\frac{\Delta x}{2}+\frac{1}{2}U'''_{i+05}\left(\frac{\Delta x}{2}\right)^2+O\left((\Delta x)^3\right), \tag{5.5}$$

$$d_i=U'_{i+05}-U''_{i+05}\frac{\Delta x}{2}+\frac{1}{2}U'''_{i+05}\left(\frac{\Delta x}{2}\right)^2+O\left((\Delta x)^3\right). \tag{5.6}$$

Subtracting the equations (5.5) and (5.6), we obtain

$$\left.\frac{\partial^2 U}{\partial x^2}\right|_{i+05}=\frac{d_{i+1}-d_i}{\Delta x}+O\left((\Delta x)^2\right). \tag{5.7}$$

In view of (5.7) and (5.1) we obtain from (5.4) the following interpolation formula

$$U_{i+05}=\frac{1}{2}\left(U_{i+1}+U_i\right)-\frac{\Delta x}{8}\left(\widetilde{d}_{i+1}-\widetilde{d}_i\right)+O\left((\Delta x)^4+(\Delta x)^{s+1}\right). \tag{5.8}$$

In view of (5.8) we obtain that $r=\min(4,s+1)$.

REFERENCES

[1] R. Abgrall and P. L. Roe, High Order Fluctuation Schemes on Triangular Meshes, Journal of Scientific Computing, Vol. 19, Nos. 1-3 (2003), pp. 3-36.

[2] François Bereux, Lionel Sainsaulieu, A roe-type Riemann solver for hyperbolic systems with relaxation based on time-dependent wave decomposition, Numer. Math. 77: 143-185 (1997).

[3] I. Boglaev, Monotone Iterates for Solving Nonlinear Monotone Difference Schemes, Computing 78 (2006), 17-30.

[4] V. S. Borisov and S. Sorek, - On monotonicity of difference schemes for computational physics, SIAM J. Sci. Comput., Vol. 25, No. 5 (2004), pp. 1557-1584.

[5] V. S. Borisov, On discrete maximum principles for linear equation systems and monotonicity of difference schemes, SIAM J. Matrix Anal. Appl., Vol. 24, No. 4 (2003), pp. 1110-1135.

[6] Russel E. Caflisch, Shi Jin, and Giovanni Russo, Uniformly accurate schemes for hyperbolic systems with relaxation, SIAM J. Numer. Anal., Vol. 34, No. 1 (1997), pp. 246-281.

[7] S. R. Chacravarthy, S. Osher, High resolution applications of the Osher upwind scheme for the Euler equations, AIAA paper #83-1943, Danver, Mass. (19843), pp. 363-372.

[8] Duboshin G. N., *Theoretical foundations of stability of motion*, Moscow University, Moscow, 1952 (in Russian).

[9] F.N. Fritsch and R.E. Carlson, Monotone piecewise cubic interpolation, SIAM J. Numer. Anal. 17, No. 2, 238-246 (1980).

[10] Gil' M. I., *Stability of finite and infinite dimensional systems*, Kluwer Academic Publishers, Boston, 1998.

[11] Gil' M. I., *Difference Equations in Normed Spaces, Stability and Oscillations*, Elsevier, Amsterdam, 2007.

[12] Edwige Godlewski and Pierre-Arnaud Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws,* Springer-Verlag, New York, 1996.

[13] S. K. Godunov, A finite difference method for the numerical computation and discontinuous solutions of fluid dynamics, Mat. Sb., 47 (1959), pp. 271-306 (in Russian).

[14] A. Harten, J. M. Hyman, and P. D. Lax, with appendix by B. Keyfiz, On finite-difference approximations and entropy conditions for shocks, Comm. Pure Appl. Math., 29 (1976), pp. 297-322.

[15] A. Harten, High resolution schemes for hyperbolic conservation laws, J. Comput. Phys., V. 49 (1983), pp. 357-393.

[16] Ami Harten, On a class of high resolution total-variaton-stable finite difference schemes, SIAM J. Numer. Anal., Vol. 21, No. 1 (1984), pp. 1-23.

[17] Ami Harten, Uniformly High Order Accurate Essentially Non-oscillatory Schemes, Ill, J. Comput. Phys., V. 71 (1987), pp. 231-303.

[18] Róbert Horváth, On the monotonicity conservation in numerical solutions of the heat equation, Appl. Numer. Math., 42 (2002), pp. 189-199.

[19] Shi Jin, Runge-Kutta Methods for Hyperbolic Conservation Laws with Stiff Relaxation Terms, J. Comp. Phys. 122 (1995), 51-67.

[20] Shi Jin and C. David Levermore, Numerical Schemes for Hyperbolic Conservation Laws with Stiff Relaxation Terms, J. Comp. Phys. 126, 449-467 (1996).

[21] David Kahaner, Cleve Moler, and Stephen Nash, *Numerical methods and software*, Prentice-Hall, New Jersey, 1989.

[22] N. N. Kalitkin, *Numerical Methods*, Nauka, Moscow, 1978 (in Russian).

[23] L.M. Kocić and G.V. Milovanović, Shape Preserving Approximations by Polynomials and Splines, Computers Math. Applic. Vol. 33, No. 11, pp. 59-97, 1997.

[24] A. G. Kulikovskii , N. V. Pogorelov, A. Yu. Semenov, *Mathematical problems of numerical solution of hyperbolic systems*, Nauka, Moscow, 2001 (in Russian).

[25] Alexander Kurganov and Eitan Tadmor, New High-Resolution Central Schemes for Nonlinear Conservation Laws and Convection-Diffusion Equations, Journal of Computational Physics 160, 241-282 (2000)

[26] Randall J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge University Press, Cambridge, 2002.

[27] Xu-Dong Liu and Eitan Tadmor, Third order nonoscillatory central scheme for hyperbolic conservation laws, Numer. Math. (1998) 79: 397-425.

[28] L. A. Monthé, A study of splitting scheme for hyperbolic conservation laws, Journal of Computational and Applied Mathematics 137 (2001) 1-12.

[29] L. Mirsky, *An Introduction to Linear Algebra*, Dover Publications, New York, 1990.

[30] K. W. Morton, *Numerical Solution of Convection-Diffusion Problems*, Chapman & Hall, London, 1996.

[31] K. W. Morton, Discretization of unsteady hyperbolic conservation laws, SIAM J. Numer. Anal., Vol. 39, No. 5 (2001), pp. 1556-1597.

[32] Giovanni Naldi and Lorenzo Pareschi, Numerical schemes for hyperbolic systems of conservation laws with stiff diffusive relaxation, SIAM J. Numer Anal., Vol. 37, No. 4 (2000), pp. 1246-1270.

[33] Haim Nessyahu and Eitan Tadmor, Non-oscillatory Central Differencing for Hyperbolic Conservation Laws, Journal of Computational Physics, Vol. 87, No 2., April 1990, pp. 408-463.

[34] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, London, 1970.

[35] V. V. Ostapenko, On the strong monotonicity of nonlinear difference schemes, Comput. Math. Math. Phys., 38 (1998), pp. 1119-1133.

[36] Lorenzo Pareschi, Central differencing based numerical schemes for hyperbolic conservation laws with relaxation terms, SIAM J. Numer. Anal., Vol. 39, No. 4 (2001), pp. 1395-1417.

[37] Lorenzo Pareschi and Giovanni Russo, Implicit-Explicit Runge-Kutta Schemes and Applications to Hyperbolic Systems with Relaxation, Journal of Scientific Computing, Vol. 25, Nos. 1/2, November 2005, pp. 129-155.

[38] Lorenzo Pareschi, Gabriella Puppo, and Giovanni Russo, Central Runge-Kutta Schemes for conservation laws, SIAM J. Sci. Comput., Vol. 26, No. 3 (2005), pp. 979-999.

[39] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling, *Numerical Recipes in C*, The Art of Scientific Computing, Cambridge University Press, New York, 1988.

[40] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd edn, Wiley-Interscience, New York, 1967.

[41] A. A. Samarskiy, On the monotone difference schemes for elliptical and parabolic equations in the case of not self-conjugate elliptical operator, Zh. Vychisl. Mat. Mat. Fiz., 5 (1965), pp. 548-551 (in Russian).

[42] A. A. Samarskii, *The theory of difference schemes*, Marcel Dekker, New York, 2001.

[43] A. A. Samarskiy and A.V. Gulin, *Stability of Finite Difference Schemes*, Nauka, Moscow, 1973 (in Russian).

[44] Susana Serna and Antonio Marquina, Capturing shock waves in inelastic granular gases, Journal of Computational Physics 209 (2004) 787-795.