SIAM J. NUMER. ANAL.
Vol. 42, No. 5, pp. 1978–1997
© 2005 Society for Industrial and Applied Mathematics

# ON CONVERGENCE OF MINMOD-TYPE SCHEMES*

SERGEI KONYAGIN†, BOJAN POPOV‡, AND OGNIAN TRIFONOV§

**Abstract.** A class of nonoscillatory numerical methods for solving nonlinear scalar conservation laws in one space dimension is considered. This class of methods contains the classical Lax–Friedrichs and the second-order Nessyahu–Tadmor schemes. In the case of linear flux, new $l_2$ stability results and error estimates for the methods are proved. Numerical experiments confirm that these methods are one-sided $l_2$ stable for convex flux instead of the usual Lip+ stability.

**Key words.** conservation laws, minmod-type schemes, convergence

**AMS subject classifications.** 65M15, 65M12

**DOI.** 10.1137/S0036142903423861

## 1. Introduction.

We are interested in the scalar hyperbolic conservation law

$$\begin{cases} u_t + f(u)_x = 0, & (x,t) \in \mathbb{R} \times (0,\infty), \\ u(x,0) = u^0(x), & x \in \mathbb{R}, \end{cases} \tag{1.1}$$

where $f$ is a given flux function. In recent years, there has been enormous activity in the development of the mathematical theory and in the construction of numerical methods for (1.1). Even though the existence-uniqueness theory of weak solutions is complete [12], there are many numerically efficient methods for which the questions of convergence and error estimates are still open. For example, there are many nonoscillatory schemes based on the minmod limiter which are numerically robust, at least in many numerical tests, but theoretical results about convergence and error estimates are still missing [3, 6, 7, 18].

In this paper, we consider a class of the so-called Godunov-type schemes for solving (1.1). There are two main steps in such schemes: evolution and projection. In the original Godunov scheme, the projection is onto piecewise constant functions—the cell averages. In the general Godunov-type method, the projection is onto piecewise polynomials. To determine the properties of these schemes it is necessary to study the properties of the projection operator. We limit our attention to the case of piecewise linear projection based on cell averages using minmod limiters for the slope reconstruction, and we call such a scheme *minmod-type*. For example, the Nessyahu–Tadmor scheme [15] is of minmod-type and is based on staggered evolution; other examples include the second-order nonoscillatory central schemes with nonstaggered grids given in [8, 9], and the UNO and TVD2 schemes in [6]. Theoretical results about convergence of such schemes to the entropy solution, or error estimates, are still missing. In most cases the authors give a variation bound for such a scheme, which is enough to conclude that the method converges to a weak solution; see [10]. The only paper which has a convergence result is the paper of Nessyahu and Tadmor [15] in

*Received by the editors February 26, 2003; accepted for publication (in revised form) April 18, 2004; published electronically January 20, 2005.
http://www.siam.org/journals/sinum/42-5/42386.html
†Department of Mathematics, Moscow State University, Moscow, Russia (konyagin@ok.ru). The research of this author was supported by grant NSh-304.2003.1.
‡Department of Mathematics, Texas A&M University, College Station, TX 77845 (popov@math.tamu.edu). The research of this author was supported by ONR grant N00014-91-J-1076.
§Department of Mathematics, University of South Carolina, Columbia, SC 29208 (trifonov@math.sc.edu). The research of this author was supported by NSF DMS grant 9970455.

which the authors prove a single cell entropy inequality for a minor modification of the original minmod scheme. A single entropy inequality is enough to conclude that the scheme is convergent to the unique entropy solution but does not give any rate of convergence. In order to get a rate, one has to have a family of entropy inequalities (see [1, 2, 11, 14]). Alternatively, for a convex flux, one can impose Lip+ stability on the projection and then prove convergence via Tadmor's Lip$'$ theory [16, 19]. Unfortunately, it is well known that minmod-type schemes are incompatible with the Lip+ condition—the Lip+ seminorm is not preserved by a minmod-type projection. It is easy to think about minmod-type schemes in terms of new/old cell averages. That is, we start with a sequence of cell averages $\{w_j\}$, and after one time step (projection and evolution) we get a new sequence $\{w'_j\}$. A scheme is *total variation diminishing* (TVD) if the variation of the new sequence $\sum_j |w'_j - w'_{j-1}|$ is not bigger than the variation of the old one $\sum_j |w_j - w_{j-1}|$, i.e., the $l_1$ norm of the jumps does not increase in time. In the Lip+ case (for convex flux) the condition on the jumps is that the biggest nonnegative jump does not increase in time:

$$\sup_j (w'_j - w'_{j-1})_+ \le \sup_j (w_j - w_{j-1})_+.$$

In section 3 of this paper, we prove that for linear flux the $l_2$ norm of the jumps for some minmod-type schemes does not increase in time. This class of schemes include the NT scheme and the TVD2 scheme considered in [6]. Based on that, we use the dual approach (see [16, 19]) to derive a new error estimate in $L^2$ in section 4. The rate of convergence that we prove is 1/2 in $L^2$, which improves the known result of 1/4 (see [19]). In section 5, we present numerical examples in the case of linear and convex flux and discuss the nonconvex case. Our numerical tests show that for convex flux the minmod schemes preserve the one-sided analogue

$$\sum_j (w'_j - w'_{j-1})_+^2 \le \sum_j (w_j - w_{j-1})_+^2,$$

which suggests a different approach to proving convergence and error estimates for such schemes in the convex case. The $l_2$ norm of the jumps is a natural candidate norm for the analysis of high-order schemes, such as central or ENO [7] type, due to its numerical viscosity. We view the results of this paper as a step toward obtaining convergence results and estimates for the rate of convergence of minmod-type schemes for solving (1.1) in the case of convex nonlinear flux.

**2. Nonoscillatory central schemes.** In this section, we are concerned with nonoscillatory central differencing approximations to the scalar conservation law

$$u_t + f(u)_x = 0. \tag{2.1}$$

The prototypes of all central schemes are the staggered form of the Lax–Friedrichs (LxF) scheme and its second-order extension, the Nessyahu–Tadmor (NT) scheme [15]. For an introduction to central schemes, see [8, 9, 13, 15]. For simplicity, we limit our attention to the staggered NT scheme described below. Let $v(x,t)$ be an approximate solution to (2.1), and assume that the space mesh $\Delta x$ and the time mesh $\Delta t$ are uniform. Let $x_j := j\Delta x$, $j \in \mathbb{Z}$, $\lambda := \frac{\Delta t}{\Delta x}$, and

$$v_j(t) := \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} v(x,t)\, dx \tag{2.2}$$

be the average of $v$ at time $t$ over $(x_{j-1/2}, x_{j+1/2})$. Let us assume that $v(\cdot,t)$ is a piecewise linear function, and that it is linear on the intervals $(x_{j-1/2}, x_{j+1/2})$, $j \in \mathbb{Z}$, of the form

$$v(x,t) = L_j(x,t) := v_j(t) + (x - x_j)\frac{1}{\Delta x}v_j', \quad x_{j-1/2} < x < x_{j+1/2}, \tag{2.3}$$

where $\frac{1}{\Delta x}v_j'$ is the numerical derivative of $v$, which is yet to be determined. Integration of (2.1) over the staggered space-time cell $(x_j, x_{j+1}) \times (t, t+\Delta t)$ yields

$$\begin{aligned} v_{j+1/2}(t+\Delta t) = {} & \frac{1}{\Delta x}\left(\int_{x_j}^{x_{j+1/2}} L_j(x,t)\,dx + \int_{x_{j+1/2}}^{x_{j+1}} L_{j+1}(x,t)\,dx\right) \\ & - \frac{1}{\Delta x}\left(\int_t^{t+\Delta t} f(v(x_{j+1},\tau))\,d\tau - \int_t^{t+\Delta t} f(v(x_j,\tau))\,d\tau\right). \end{aligned} \tag{2.4}$$

The first two integrals on the right-hand side of (2.4) can be evaluated exactly. Moreover, if the CFL condition

$$\lambda \max_{x_j \le x \le x_{j+1}} |f'(v(x,t))| \le \frac{1}{2}, \quad j \in \mathbb{Z}, \tag{2.5}$$

is met, then the last two integrants on the right of (2.4) are smooth functions of $\tau$. Hence, they can be integrated approximately by the midpoint rule with third-order local truncation error. Note that, in the case of zero slopes $\frac{1}{\Delta x}v_j'$ and $\frac{1}{\Delta x}v_{j+1}'$, the time integration is exact for any flux $f$, and even for nonzero slopes the time integration can be exact for a low degree polynomial flux. Thus, following [15], we arrive at

$$\begin{aligned} v_{j+1/2}(t+\Delta t) = {} & \frac{1}{2}(v_j(t) + v_{j+1}(t)) + \frac{1}{8}(v_j' - v_{j+1}') \\ & - \lambda\left(f\left(v\left(x_{j+1}, t + \frac{\Delta t}{2}\right)\right) - f\left(v\left(x_j, t + \frac{\Delta t}{2}\right)\right)\right). \end{aligned} \tag{2.6}$$

By Taylor expansion and the conservation law (2.1), we obtain

$$v\left(x_j, t + \frac{\Delta t}{2}\right) = v_j(t) - \frac{1}{2}\lambda f_j', \tag{2.7}$$

where $\frac{1}{\Delta x}f_j'$ stand for an approximate numerical derivative of the flux $f(v(x = x_j, t))$. The following choices are widely used as approximations of the numerical derivatives (we drop $t$ to simplify the notation):

$$\begin{aligned} v_j' &= \mathrm{m}(v_{j+1} - v_j, v_j - v_{j-1}), \\ f_j' &= \mathrm{m}(f(v_{j+1}) - f(v_j), f(v_j) - f(v_{j-1})), \end{aligned} \tag{2.8}$$

where $\mathrm{m}(a,b)$ stands for the minmod limiter

$$\mathrm{m}(a,b) \equiv \mathrm{MinMod}(a,b) := \frac{1}{2}(\mathrm{sgn}(a) + \mathrm{sgn}(b)) \cdot \min(|a|, |b|) \tag{2.9}$$

with the usual generalization

$$\mathrm{m}(E) := \begin{cases} \inf(E) & \text{if} \quad E \subset \mathbb{R}_+, \\ \sup(E) & \text{if} \quad E \subset \mathbb{R}_-, \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

A generalization of this numerical approximation is based the so-called minmod-$\theta$ limiters

$$
\begin{aligned}
(2.11)\qquad v_j' &= \mathrm{m}\left(\theta(v_{j+1}-v_j), \frac{1}{2}(v_{j+1}-v_{j-1}), \theta(v_j - v_{j-1})\right),\\
f_j' &= \mathrm{m}\left(\theta(f(v_{j+1})-f(v_j)), \frac{1}{2}(f(v_{j+1})-f(v_{j-1})), \theta(f(v_j)-f(v_{j-1}))\right).
\end{aligned}
$$

Given the approximate slopes and flux derivatives (2.11), we have a family of central schemes in the predictor-corrector form

$$
\begin{aligned}
&v\left(x_j, t+\frac{\Delta t}{2}\right) = v_j(t) - \frac{1}{2}\lambda f_j',\\
(2.12)\qquad &v_{j+1/2}(t+\Delta t) = \frac{1}{2}(v_j(t)+v_{j+1}(t)) + \frac{1}{8}(v_j' - v_{j+1}')\\
&\qquad - \lambda\left(f\left(v\left(x_{j+1}, t+\frac{\Delta t}{2}\right)\right) - f\left(v\left(x_j, t+\frac{\Delta t}{2}\right)\right)\right),
\end{aligned}
$$

where we start with $v_j(0) := \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,dx$. Note that we alternate between two uniform partitions of the real line: all intervals with integer end points for $t = 2k\Delta t$, $k \in \mathbb{Z}$, and half integers for $t = (2k+1)\Delta t$, $k \in \mathbb{Z}$. As a special case, we recover the staggered LxF scheme for $\theta = 0$ and the basic minmod scheme for $\theta = 1$ (the middle slope in the minmod limiter (2.11) drops if $\theta \le 1$).

**3. $l_2$ stability for linear flux.** In this section we will prove that the central scheme given in (2.12) is $l_2$ stable for any $\theta$ in the interval $[0,1]$. Based on this stability we will also derive a new error estimate in $L^2$ instead of the usual $L^1$ estimates in the conservation laws. Note that even for linear flux $f$, the minmod-type schemes are not linear and the only global property known is that the total variation does not increase in time under an appropriate CFL condition; see [15]. The class of minmod-type schemes is also not Lip+ stable except for the obvious choice $\theta = 0$. Let us consider a linear flux $f(u) = au$, uniform time steps $t_n = n\Delta t$, and restrict the minmod limiter to $\theta \le 1$. We denote $v_j^n := v_j(t_n)$, $\delta_j^n := v_j^n - v_{j-1}^n$. The minmod scheme (2.12) reduces to

$$
\begin{aligned}
(3.1)\qquad v_{j+1/2}^{n+1} &= \frac{1}{2}(v_j^n + v_{j+1}^n) + \frac{\theta}{8}\left(\mathrm{m}(\delta_j^n, \delta_{j+1}^n) - \mathrm{m}(\delta_{j+1}^n, \delta_{j+2}^n)\right)\\
&\quad - \frac{a\Delta t}{\Delta x}\left(v_{j+1}^n - \frac{a\Delta t}{2\Delta x}\theta\mathrm{m}(\delta_{j+1}^n, \delta_{j+2}^n) - v_j^n + \frac{a\Delta t}{2\Delta x}\theta\mathrm{m}(\delta_j^n, \delta_{j+1}^n)\right).
\end{aligned}
$$

Hence, we have an explicit formula for the new cell averages (at time $t_{n+1}$) on a staggered grid in terms of the old ones (at time $t_n$) on a regular grid. In order to simplify the notation, we drop the time dependence and denote $w_j := v_j^n$, $w_{j+1}' := v_{j+1/2}^{n+1}$, $\delta_j' := w_j' - w_{j-1}'$, $\alpha := \frac{1}{2} + \frac{a\Delta t}{\Delta x}$, and $\beta := \frac{1}{2}\alpha(1-\alpha)$. With this notation, we have the following relation between the sequence of the new averages $\{w_j'\}$ and the old ones $\{w_j\}$:

$$
(3.2)\qquad w_j' = \alpha w_{j-1} + (1-\alpha)w_j + \theta\beta\left(\mathrm{m}(\delta_{j-1}, \delta_j) - \mathrm{m}(\delta_j, \delta_{j+1})\right).
$$

Using that $\delta_j = w_j - w_{j-1}$, we derive the formula for the sequence of new jumps in terms of the old ones:

$$
(3.3)\qquad \delta_j' = \alpha\delta_{j-1} + (1-\alpha)\delta_j - \theta\beta\mathrm{m}(\delta_{j-2}, \delta_{j-1}) + 2\theta\beta\mathrm{m}(\delta_{j-1}, \delta_j) - \theta\beta\mathrm{m}(\delta_j, \delta_{j+1}).
$$

The CFL condition (2.5) reduces to $0 \le \alpha \le 1$ because $\alpha = \frac{1}{2} + \frac{a\Delta t}{\Delta x}$ and $|\frac{a\Delta t}{\Delta x}| \le \frac{1}{2}$. The main result in this section is the following stability result.

THEOREM 3.1. *If the initial condition $u_0 \in L^1_{loc}(\mathbb{R})$, then the $l_2$ norm of the jumps of the approximate solution $v(\cdot, t)$ is nonincreasing in time. That is,*

$$\|\{\delta'_j\}\|_{l_2} \equiv \|\{v_j^{n+1} - v_{j-1}^{n+1}\}\|_{l_2} \le \|\{\delta_j\}\|_{l_2} \equiv \|\{v_j^n - v_{j-1}^n\}\|_{l_2} \tag{3.4}$$

*for all $n \ge 1$.*

*Proof.* It is clear that we have to prove the result for one step, assuming that $\|\{\delta_j\}\|_{l_2} < \infty$. We split the proof into two parts. First, we prove the stability for a monotone sequence $\{w_j\}$. By symmetry, it is sufficient to consider the case $\delta_j \ge 0$ for all $j \in \mathbb{Z}$. Then we apply that result locally to derive the $l_2$ stability for a general sequence.

THEOREM 3.2. *Let us assume that $\delta_j \ge 0$, $j \in \mathbb{Z}$, and $\delta'_j$ are given by* (3.3). *Then*

$$\|\{\delta'_j\}\|_{l_2} \le \|\{\delta_j\}\|_{l_2}. \tag{3.5}$$

*Proof.* Let us recall that $\{\delta_j\}_{-\infty}^{\infty} \in l_2$ and $\delta_j \ge 0$ for all $j$. It is enough to prove Theorem 3.2 only for $0 < \alpha < 1$. Let $\beta_1 := \theta\beta$; then $0 < \beta_1 \le \beta$. We construct the new sequence $\{\delta'_j\}$ by using the rule

$$\delta'_j = (1-\alpha)\delta_j + \alpha\delta_{j-1} - \beta_1 \min(\delta_{j-2}, \delta_{j-1}) + 2\beta_1 \min(\delta_{j-1}, \delta_j) - \beta_1 \min(\delta_j, \delta_{j+1}) \tag{3.6}$$

for each $j$. First we assume that $\{\delta_j\}$ has finite support. It is easy to see how to modify the proof in case the support is not finite. Therefore we assume $\delta_j = 0$ for $j \le 3$ and for $j \ge N-3$ for some integer $N$. Then $\delta'_j = 0$ for $j \le 3$ and $j \ge N-2$. Thus it suffices to prove

$$\sum_{j=1}^{N} \delta_j^2 \ge \sum_{j=1}^{N} (\delta'_j)^2. \tag{3.7}$$

Let us introduce some notation. Let $y_j = \min(\delta_j, \delta_{j+1})$, $\Delta\delta_j = \delta_j - \delta_{j-1}$, $\Delta y_j = y_j - y_{j-1}$, $\Delta^2\delta_j = \delta_j - 2\delta_{j-1} + \delta_{j-2}$, and $\Delta^2 y_j = y_j - 2y_{j-1} + y_{j-2}$. Then (3.6) becomes

$$\delta'_j = ((1-\alpha)\delta_j + \alpha\delta_{j-1}) - \beta_1 \Delta^2 y_j \quad \text{and}$$

$$\sum_{j=1}^{N} (\delta'_j)^2 = \sum_{j=1}^{N} \left( ((1-\alpha)\delta_j + \alpha\delta_{j-1})^2 - 2\beta_1((1-\alpha)\delta_j + \alpha\delta_{j-1})\Delta^2 y_j + \beta_1^2 (\Delta^2 y_j)^2 \right).$$

Note that since $\delta_0 = \delta_1 = 0$ and $\delta_{N-1} = \delta_N = 0$, we have

$$\begin{aligned}\sum_{j=1}^{N} \delta_j^2 - ((1-\alpha)\delta_j + \alpha\delta_{j-1})^2 &= \sum_{j=1}^{N} (1 - (1-\alpha)^2 - \alpha^2)\delta_j^2 - 2\alpha(1-\alpha)\delta_j\delta_{j-1} \\ &= 2\beta \sum_{j=1}^{N} (2\delta_j^2 - 2\delta_j\delta_{j-1}) = 2\beta \sum_{j=1}^{N} (\delta_j - \delta_{j-1})^2 = 2\beta \sum_{j=1}^{N} (\Delta\delta_j)^2.\end{aligned}$$

Thus we get

(3.8)
$$\sum_{j=1}^{N}\delta_j^2-\sum_{j=1}^{N}(\delta_j')^2=\sum_{j=1}^{N}\left(2\beta(\Delta\delta_j)^2+2\beta_1((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta_1^2(\Delta^2y_j)^2\right).$$

To prove Theorem 3.2, we need to prove

$$\sum_{j=1}^{N}\left(2\beta(\Delta\delta_j)^2+2\beta_1((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta_1^2(\Delta^2y_j)^2\right)\geq 0.$$

Note that

$$\begin{aligned}
&\sum_{j=1}^{N}\left(2\beta(\Delta\delta_j)^2+2\beta_1((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta_1^2(\Delta^2y_j)^2\right)\\
&\quad=\beta_1\left(\sum_{j=1}^{N}\left(2\left(\frac{\beta}{\beta_1}\right)(\Delta\delta_j)^2+2((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta_1(\Delta^2y_j)^2\right)\right)\\
&\quad\geq\beta_1\left(\sum_{j=1}^{N}\left(2(\Delta\delta_j)^2+2((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta(\Delta^2y_j)^2\right)\right)\\
&\quad=\left(\frac{\beta_1}{\beta}\right)\sum_{j=1}^{N}\left(2\beta(\Delta\delta_j)^2+2\beta((1-\alpha)\delta_j+\alpha\delta_{j-1})\Delta^2y_j-\beta^2(\Delta^2y_j)^2\right).
\end{aligned}$$

Therefore it is sufficient to prove the theorem in the case $\beta_1=\beta$, which is the worst case in a certain sense. Now we use $\Delta^2y_j=\Delta y_j-\Delta y_{j-1}$, $\Delta y_j=0$, $\delta_j=0$ for $j\leq 1$, $j\geq N-1$, and Abel's transform to obtain

$$\sum_{j=1}^{N}\delta_j\Delta^2y_j=\sum_{j=1}^{N}(\delta_j-\delta_{j+1})\Delta y_j\quad\text{and}\quad\sum_{j=1}^{N}\delta_{j-1}\Delta^2y_j=\sum_{j=1}^{N}(\delta_{j-1}-\delta_j)\Delta y_j.$$

Thus, (3.8) becomes

$$\begin{aligned}
\sum_{j=1}^{N}\delta_j^2-\sum_{j=1}^{N}(\delta_j')^2=2\beta\Bigg(&\sum_{j=1}^{N}(\Delta\delta_j)^2-(1-\alpha)\sum_{j=1}^{N}\Delta\delta_{j+1}\Delta y_j\\
&-\alpha\sum_{j=1}^{N}\Delta\delta_j\Delta y_j-\frac{\beta}{2}\sum_{j=1}^{N}(\Delta^2y_j)^2\Bigg).
\end{aligned}$$

Recall that $y_j=\min(\delta_j,\delta_{j+1})$, $\Delta\delta_j=\delta_j-\delta_{j-1}$, $\Delta y_j=y_j-y_{j-1}$, $\Delta^2\delta_j=\delta_j-2\delta_{j-1}+\delta_{j-2}$, and $\Delta^2y_j=y_j-2y_{j-1}+y_{j-2}$. To finish the proof of Theorem 3.2, it is sufficient to prove the following two lemmas.

LEMMA 3.3.

$$\sum_{j=1}^{N}(\Delta\delta_j)^2-(1-\alpha)\sum_{j=1}^{N}\Delta\delta_{j+1}\Delta y_j-\alpha\sum_{j=1}^{N}\Delta\delta_j\Delta y_j-\beta\sum_{j=1}^{N}(\Delta^2\delta_j)^2\geq 0.$$

LEMMA 3.4.

$$2\sum_{j=1}^{N}(\Delta^2\delta_j)^2 \geq \sum_{j=1}^{N}(\Delta^2 y_j)^2.$$

*Proof of Lemma* 3.3. We consider that $\sum_j$ denotes $\sum_{j=1}^{N}$. Define

$$A = \sum_j \Delta\delta_{j+1}\Delta y_j \quad \text{and} \quad B = \sum_j \Delta\delta_j \Delta y_j.$$

Our aim is to prove that

$$\sum_j (\Delta\delta_j)^2 - (1-\alpha)A - \alpha B - \beta\sum_j(\Delta^2\delta_j)^2 \geq 0. \tag{3.9}$$

Let $u_+ = \max(u,0)$, $u_- = \min(u,0)$. It is easy to check that

$$\Delta y_j = (\Delta\delta_j)_+ + (\Delta\delta_{j+1})_-. \tag{3.10}$$

We can transform $A$ as follows:

$$\begin{aligned} A &= \sum_j \Delta\delta_{j+1}((\Delta\delta_j)_+ + (\Delta\delta_{j+1})_-) \\ &= \sum_j \Delta\delta_{j+1}(\Delta\delta_j)_+ + \sum_j \Delta\delta_j(\Delta\delta_j)_- = \sum_{\Delta\delta_j\leq 0}(\Delta\delta_j)^2 + \sum_{\Delta\delta_j\geq 0}\Delta\delta_j\Delta\delta_{j+1} \\ &= \sum_{\Delta\delta_j\leq 0}(\Delta\delta_j)^2 + \sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\leq 0}\Delta\delta_j\Delta\delta_{j+1} + D, \end{aligned} \tag{3.11}$$

where $D = \sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\geq 0}\Delta\delta_j\Delta\delta_{j+1}$. Further,

$$\begin{aligned} D &= \frac{1}{2}\sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\geq 0}\big((\Delta\delta_j)^2 + (\Delta\delta_{j+1})^2 - (\Delta^2\delta_{j+1})^2\big) \\ &= \frac{1}{2}\sum_{\Delta\delta_{j-1}\geq 0,\,\Delta\delta_j\geq 0}\big((\Delta\delta_{j-1})^2 + (\Delta\delta_j)^2 - (\Delta^2\delta_j)^2\big) \\ &= \frac{1}{2}\sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\geq 0}(\Delta\delta_j)^2 + \frac{1}{2}\sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j-1}\geq 0}(\Delta\delta_j)^2 - \frac{1}{2}\sum_{\Delta\delta_{j-1}\geq 0,\,\Delta\delta_j\geq 0}(\Delta^2\delta_j)^2. \end{aligned}$$

By (3.11),

$$\begin{aligned} A = \sum_j(\Delta\delta_j)^2 &- \frac{1}{2}\sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}<0}(\Delta\delta_j)^2 - \frac{1}{2}\sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j-1}<0}(\Delta\delta_j)^2 \\ &- \frac{1}{2}\sum_{\Delta\delta_{j-1}\geq 0,\,\Delta\delta_j\geq 0}(\Delta^2\delta_j)^2 + \sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\leq 0}\Delta\delta_j\Delta\delta_{j+1}. \end{aligned} \tag{3.12}$$

Transform $B$ in the same way as $A$:

$$B = \sum_{\Delta\delta_j\geq 0}(\Delta\delta_j)^2 + \sum_{\Delta\delta_j\geq 0,\,\Delta\delta_{j+1}\leq 0}\Delta\delta_j\Delta\delta_{j+1} + E, \tag{3.13}$$

where $E = \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j+1} \le 0} \Delta\delta_j \Delta\delta_{j+1}$. The quantity $E$ can also be rewritten in the same way as $D$:

$$E = \frac{1}{2} \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j+1} \le 0} (\Delta\delta_j)^2 + \frac{1}{2} \sum_{\Delta\delta_{j-1} \le 0,\, \Delta\delta_j \le 0} (\Delta\delta_j)^2 - \frac{1}{2} \sum_{\Delta\delta_{j-1} \le 0,\, \Delta\delta_j \le 0} (\Delta^2\delta_j)^2.$$

Combining this equality with (3.13), we get

$$\begin{aligned} B = \sum_j (\Delta\delta_j)^2 &- \frac{1}{2} \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j+1} > 0} (\Delta\delta_j)^2 - \frac{1}{2} \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j-1} > 0} (\Delta\delta_j)^2 \\ &- \frac{1}{2} \sum_{\Delta\delta_{j-1} \le 0,\, \Delta\delta_j \le 0} (\Delta^2\delta_j)^2 + \sum_{\Delta\delta_{j+1} \le 0,\, \Delta\delta_j \ge 0} \Delta\delta_j \Delta\delta_{j+1}. \end{aligned} \tag{3.14}$$

By (3.12) and (3.14),

$$\sum_j (\Delta\delta_j)^2 - (1-\alpha)A - \alpha B - \beta \sum_j (\Delta^2\delta_j)^2 = F + G + H + I + J + K + L, \tag{3.15}$$

where

$$F = \frac{1-\alpha}{2} \sum_{\Delta\delta_j \ge 0,\, \Delta\delta_{j+1} < 0} (\Delta\delta_j)^2, \quad G = \frac{1-\alpha}{2} \sum_{\Delta\delta_j \ge 0,\, \Delta\delta_{j-1} < 0} (\Delta\delta_j)^2,$$

$$H = \frac{\alpha}{2} \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j+1} > 0} (\Delta\delta_j)^2, \quad I = \frac{\alpha}{2} \sum_{\Delta\delta_j \le 0,\, \Delta\delta_{j-1} > 0} (\Delta\delta_j)^2,$$

$$\begin{aligned} J = &- \sum_{\Delta\delta_{j+1} \le 0,\, \Delta\delta_j \ge 0} \Delta\delta_j \Delta\delta_{j+1} + \left(\frac{1-\alpha}{2} - \beta\right) \sum_{\Delta\delta_{j-1} \ge 0,\, \Delta\delta_j \ge 0} (\Delta^2\delta_j)^2 \\ &+ \left(\frac{\alpha}{2} - \beta\right) \sum_{\Delta\delta_{j-1} \le 0,\, \Delta\delta_j \le 0} (\Delta^2\delta_j)^2, \end{aligned}$$

$$K = -\beta \sum_{\Delta\delta_{j-1} > 0,\, \Delta\delta_j < 0} (\Delta^2\delta_j)^2, \qquad \text{and} \quad L = -\beta \sum_{\Delta\delta_{j-1} < 0,\, \Delta\delta_j > 0} (\Delta^2\delta_j)^2.$$

We have to prove that $F + G + H + I + J + K + L \ge 0$. Among the sums $F, G, H, I, J, K, L$, only the two last sums might be negative; we will show that and

$$F + I + K \ge 0, \tag{3.16}$$

$$G + H + L \ge 0. \tag{3.17}$$

Indeed,

$$\begin{aligned} \frac{1-\alpha}{2}(\Delta\delta_{j-1})^2 + \frac{\alpha}{2}(\Delta\delta_j)^2 - \beta(\Delta^2\delta_j)^2 &= \frac{1-\alpha}{2}(\Delta\delta_{j-1})^2 + \frac{\alpha}{2}(\Delta\delta_j)^2 \\ &- \frac{(1-\alpha)\alpha}{2}(\Delta\delta_j - \Delta\delta_{j-1})^2 = \frac{1}{2}((1-\alpha)\Delta\delta_{j-1} + \alpha\Delta\delta_j))^2 \ge 0. \end{aligned}$$

Summing the last inequality over all $j$ with $\Delta\delta_{j-1} > 0$, $\Delta\delta_j < 0$, we get (3.16). The inequality (3.17) can be proved similarly.

Additionally, we have

$$J \geq 0. \tag{3.18}$$

Finally, plugging (3.16), (3.17), and (3.18) into (3.15), we obtain the required (3.9). This completes the proof of Lemma 3.3.

*Proof of Lemma* 3.4. First, recall that $\Delta^2 y_j = 0$ for $j \leq 1$ and $j \geq N$. Also, from the proof of Lemma 3.3 we have

$$\Delta y_j = \begin{cases} \Delta\delta_{j+1} & \text{if} \quad \Delta\delta_{j+1} \leq 0, \quad \Delta\delta_j \leq 0, \\ \Delta\delta_{j+1} + \Delta\delta_j & \text{if} \quad \Delta\delta_{j+1} \leq 0, \quad \Delta\delta_j \geq 0, \\ \Delta\delta_j & \text{if} \quad \Delta\delta_{j+1} \geq 0, \quad \Delta\delta_j \geq 0, \\ 0 & \text{if} \quad \Delta\delta_{j+1} \geq 0, \quad \Delta\delta_j \leq 0. \end{cases}$$

Similarly,

$$\Delta y_{j-1} = \begin{cases} \Delta\delta_j & \text{if} \quad \Delta\delta_j \leq 0, \quad \Delta\delta_{j-1} \leq 0, \\ \Delta\delta_j + \Delta\delta_{j-1} & \text{if} \quad \Delta\delta_j \leq 0, \quad \Delta\delta_{j-1} \geq 0, \\ \Delta\delta_{j-1} & \text{if} \quad \Delta\delta_j \geq 0, \quad \Delta\delta_{j-1} \geq 0, \\ 0 & \text{if} \quad \Delta\delta_j \geq 0, \quad \Delta\delta_{j-1} \leq 0. \end{cases}$$

Therefore $\Delta\delta_{j-1}$, $\Delta\delta_j$, $\Delta\delta_{j+1}$ and their signs uniquely determine $\Delta^2 y_j$. We have eight cases depending on what the signs of $\Delta\delta_{j-1}$, $\Delta\delta_j$, $\Delta\delta_{j+1}$ are.

- *Case* I. $(+,+,+)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} \geq 0$. Then, $\Delta y_j = \Delta\delta_j$, $\Delta y_{j-1} = \Delta\delta_{j-1}$, and so $\Delta^2 y_j = \Delta^2\delta_j$; thus, $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_j)^2$ in this case.
- *Case* II. $(+,+,-)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} < 0$. Then, $\Delta y_j = \Delta\delta_{j+1} + \Delta\delta_j$, $\Delta y_{j-1} = \Delta\delta_{j-1}$, and so $\Delta^2 y_j = \Delta\delta_{j+1} + \Delta\delta_j - \Delta\delta_{j-1}$.
- *Case* III. $(+,-,+)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j < 0$, $\Delta\delta_{j+1} \geq 0$. Then, $\Delta y_j = 0$, $\Delta y_{j-1} = \Delta\delta_j + \Delta\delta_{j-1}$, and so $\Delta^2 y_j = -\Delta\delta_j - \Delta\delta_{j-1}$. In this case $(\Delta^2 y_j)^2 - (\Delta^2\delta_j)^2 = 4\Delta\delta_j\Delta\delta_{j-1} \leq 0$ and $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_j)^2$.
- *Case* IV. $(+,-,-)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j < 0$, $\Delta\delta_{j+1} < 0$. Then, $\Delta y_j = \Delta\delta_{j+1}$, $\Delta y_{j-1} = \Delta\delta_j + \Delta\delta_{j-1}$, and so $\Delta^2 y_j = \Delta\delta_{j+1} - \Delta\delta_j - \Delta\delta_{j-1}$.
- *Case* V. $(-,+,+)$, that is, $\Delta\delta_{j-1} < 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} \geq 0$. Then, $\Delta y_j = \Delta\delta_j$, $\Delta y_{j-1} = 0$, and so $\Delta^2 y_j = \Delta\delta_j$. In this case $0 \leq \Delta\delta_j < \Delta\delta_j - \Delta\delta_{j-1}$ and $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_j)^2$.
- *Case* VI. $(-,+,-)$, that is, $\Delta\delta_{j-1} < 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} < 0$. Then, $\Delta y_j = \Delta\delta_{j+1} + \Delta\delta_j$, $\Delta y_{j-1} = 0$, and so $\Delta^2 y_j = \Delta\delta_{j+1} + \Delta\delta_j$. In this case $(\Delta^2 y_j)^2 - (\Delta^2\delta_{j+1})^2 = 4\Delta\delta_{j+1}\Delta\delta_j \leq 0$ and $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_{j+1})^2$.
- *Case* VII. $(-,-,+)$, that is, $\Delta\delta_{j-1} < 0$, $\Delta\delta_j < 0$, $\Delta\delta_{j+1} \geq 0$. Then, $\Delta y_j = 0$, $\Delta y_{j-1} = \Delta\delta_j$, and so $\Delta^2 y_j = -\Delta\delta_j$. In this case $0 < -\Delta\delta_j \leq \Delta\delta_{j+1} - \Delta\delta_j$ and $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_{j+1})^2$.
- *Case* VIII. $(-,-,-)$, that is, $\Delta\delta_{j-1} < 0$, $\Delta\delta_j < 0$, $\Delta\delta_{j+1} < 0$. Then, $\Delta y_j = \Delta\delta_{j+1}$, $\Delta y_{j-1} = \Delta\delta_j$, and so $\Delta^2 y_j = \Delta^2\delta_{j+1}$. In this case $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_{j+1})^2$.

Therefore in Cases I $(+,+,+)$, III $(+,-,+)$, and V $(-,+,+)$, $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_j)^2$, and in Cases VI $(-,+,-)$, VII $(-,-,+)$, and VIII $(-,-,-)$, $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_{j+1})^2$. There are only two "bad" cases: II $(+,+,-)$ and IV $(+,-,-)$, which need a special treatment.

Next, we define a sequence of $+$ and $-$ signs $\{s_j\}$, where $s_j = +$ if $\Delta\delta_j \geq 0$ and $s_j = -$ if $\Delta\delta_j < 0$. Note that $s_j = +$ for $j \leq 3$ and $j \geq N-2$. There are three types of "bad" quadruples:

- Type A quadruple: $(+,+,-,-)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} < 0$, $\Delta\delta_{j+2} < 0$ for some $j$. We claim that in this case the following inequality holds:

$$(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 \leq 2(\Delta^2\delta_j)^2 + 2(\Delta^2\delta_{j+1})^2 + 2(\Delta^2\delta_{j+2})^2. \tag{3.19}$$

In this case $\Delta^2 y_j = \Delta\delta_{j+1} + \Delta\delta_j - \Delta\delta_{j-1}$ and $\Delta^2 y_{j+1} = \Delta\delta_{j+2} - \Delta\delta_{j+1} - \Delta\delta_j$. If we denote $\Delta\delta_{j-1}$ by $a$, $\Delta\delta_j$ by $b$, $\Delta\delta_{j+1}$ by $c$, and $\Delta\delta_{j+2}$ by $d$, the above inequality becomes

$$\begin{aligned}&(c+b-a)^2 + (d-c-b)^2 \leq 2(b-a)^2 + 2(c-b)^2 + 2(d-c)^2\\ &\text{for } a \geq 0,\ b \geq 0,\ c < 0,\ d < 0,\end{aligned}$$

which is equivalent to $a^2 + 2b^2 + 2c^2 + d^2 - 2ab + 2ac - 8bc + 2bd - 2cd \geq 0$, or

$$(a-b+c)^2 + (b-c+d)^2 - 4bc \geq 0,$$

which holds since $b \geq 0$, $c < 0$.

- Type B quadruple: $(+,+,-,+)$, that is, $\Delta\delta_{j-1} \geq 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} < 0$, $\Delta\delta_{j+2} \geq 0$ for some $j$. We claim that in this case the following inequality holds:

$$(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 \leq 2(\Delta^2\delta_j)^2 + 2(\Delta^2\delta_{j+1})^2 + (\Delta^2\delta_{j+2})^2. \tag{3.20}$$

In this case $\Delta^2 y_j = \Delta\delta_{j+1} + \Delta\delta_j - \Delta\delta_{j-1}$ and $\Delta^2 y_{j+1} = -\Delta\delta_{j+1} - \Delta\delta_j$. Using the notation we just introduced, the inequality becomes

$$\begin{aligned}&(c+b-a)^2 + (c+b)^2 \leq 2(b-a)^2 + 2(c-b)^2 + (d-c)^2\\ &\text{for } a \geq 0,\ b \geq 0,\ c < 0,\ d \geq 0.\end{aligned}$$

Since $(d-c)^2 \geq c^2$, it is sufficient to prove

$$(c+b-a)^2 + (c+b)^2 \leq 2(b-a)^2 + 2(c-b)^2 + c^2, \text{ or}$$

$$a^2 + 2b^2 + c^2 - 2ab + 2ac - 8bc \geq 0, \text{ or } (a-b+c)^2 + b^2 - 6bc \geq 0,$$

which holds for $b \geq 0$, $c < 0$.

- Type C quadruple: $(-,+,-,-)$, that is, $\Delta\delta_{j-1} < 0$, $\Delta\delta_j \geq 0$, $\Delta\delta_{j+1} < 0$, $\Delta\delta_{j+2} < 0$ for some $j$. We claim that in this case the following inequality holds:

$$(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 \leq (\Delta^2\delta_j)^2 + 2(\Delta^2\delta_{j+1})^2 + 2(\Delta^2\delta_{j+2})^2. \tag{3.21}$$

In this case $\Delta^2 y_j = \Delta\delta_{j+1} + \Delta\delta_j$ and $\Delta^2 y_{j+1} = \Delta\delta_{j+2} - \Delta\delta_{j+1} - \Delta\delta_j$. Using the notation we just introduced, the inequality becomes

$$\begin{aligned}&(c+b)^2 + (d-c-b)^2 \leq (b-a)^2 + 2(c-b)^2 + 2(d-c)^2\\ &\text{for } a < 0,\ b \geq 0,\ c < 0,\ d < 0.\end{aligned}$$

Since $(b-a)^2 \geq b^2$, it is sufficient to prove

$$(c+b)^2 + (d+c-b)^2 \leq b^2 + 2(c-b)^2 + 2(d-c)^2, \text{ or}$$

$$b^2 + 2c^2 + d^2 - 8bc + 2bd - 2cd \geq 0, \text{ or } (b-c+d)^2 + c^2 - 6bc \geq 0,$$

which holds for $b \geq 0$, $c < 0$.

We will call the Type A (3.19), Type B (3.20), and Type C (3.21) inequalities "long" inequalities; we call the inequalities of type $(\Delta^2 y_j)^2 \le (\Delta^2 \delta_j)^2$ and $(\Delta^2 y_j)^2 \le (\Delta^2 \delta_{j+1})^2$ "short" inequalities.

Now, we construct a set of inequalities. We identify all "bad" quadruples and include the corresponding inequality (Type A, B, or C) in the set. Next, for all $j \in [1, N]$ such that $s_j$ is not a middle element of a "bad" quadruple, and such that $j$ does not belong to the "bad" Cases II and IV, we include the corresponding "short" inequality in the set. Finally, we add all inequalities in the set. Taking into account that $\Delta^2 \delta_j = 0$ and $\Delta^2 y_j = 0$ for $j > N$, the resulting inequality is

$$\sum_{j=1}^{N} a_j (\Delta^2 y_j)^2 \le \sum_{j=1}^{N} b_j (\Delta^2 \delta_j)^2, \tag{3.22}$$

where the $a_j$'s and $b_j$'s are nonnegative integers. To finish the proof of the lemma we need to show $a_j \ge 1$ and $b_j \le 2$ for all $j \in [1, N]$.

Note that all "long" inequalities have the form $(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 \le \cdots$, where $s_j$ and $s_{j+1}$ are the middle elements of a "bad" quadruple. Then $a_j \ge 1$ if $s_j$ is a middle element of a "bad" quadruple. (By middle element of a quadruple we mean second or third element of the quadruple.)

Now, suppose that $s_j$ is not a middle element of a "bad" quadruple. Then $j$ does not belong to the "bad" Cases II and IV. Indeed if $j$ is in Case II: $(s_{j-1}, s_j, s_{j+1}) = (+, +, -)$, then $s_j$ is a middle element of Type B quadruple if $s_{j+2} = +$ and a middle element of Type A quadruple if $s_{j+2} = -$. Similarly, if $j$ is in Case IV: $(s_{j-1}, s_j, s_{j+1}) = (+, -, -)$, then $s_j$ is a middle element of Type A quadruple if $s_{j-1} = +$ and a middle element of Type C quadruple if $s_{j-1} = -$. Therefore a "short" inequality for $(\Delta^2 y_j)^2$ has been included in the set of inequalities. Thus $a_j \ge 1$ in this case as well. We have proved $a_j \ge 1$ for all $j \in [1, N]$.

Now, we prove $b_j \le 2$ for all $j \in [1, N]$. Note that $(\Delta^2 \delta_j)^2$ can appear in only two "short" inequalities: $(\Delta^2 y_j)^2 \le (\Delta^2 \delta_j)^2$ and $(\Delta^2 y_{j-1})^2 \le (\Delta^2 \delta_j)^2$. Therefore $b_j \le 2$ if $(\Delta^2 \delta_j)^2$ does not appear in any "long" inequalities, that is, if $s_j$ is not a second, third, or fourth element of a "bad" quadruple.

The case when $s_j$ is a second, third, or fourth element of a "bad" quadruple requires more work. First, note that two distinct "bad" quadruples have at most two common elements. Indeed all "bad" quadruples are of the form $(*, +, -, *)$, where $*$ denotes $+$ or $-$, and no "bad" quadruple has $(+, -)$ as its first two or last two elements. Next, the only case when two "bad" quadruples have two common elements is the following configuration:

$$(s_{j-1}, s_j, s_{j+1}, s_{j+2}, s_{j+3}, s_{j+4}) = (+, +, -, +, -, -). \tag{3.23}$$

Indeed, a Type A quadruple cannot share exactly two elements with another "bad" quadruple because no "bad" quadruple has $(-, -)$ as its first two elements, or $(+, +)$ as its last two elements. Similar analysis shows that the only way a Type B or Type C quadruple can share exactly two elements with another Type B or Type C quadruple is when the configuration (3.23) occurs.

Let us analyze the configuration (3.23). The "long" inequalities which correspond to the two "bad" quadruples in this configuration are

$$(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 \le 2(\Delta^2 \delta_j)^2 + 2(\Delta^2 \delta_{j+1})^2 + (\Delta^2 \delta_{j+2})^2, \tag{3.24}$$

$$(\Delta^2 y_{j+2})^2 + (\Delta^2 y_{j+3})^2 \le (\Delta^2 \delta_{j+2})^2 + 2(\Delta^2 \delta_{j+3})^2 + 2(\Delta^2 \delta_{j+4})^2. \tag{3.25}$$

Their sum is

$$(\Delta^2 y_j)^2 + (\Delta^2 y_{j+1})^2 + (\Delta^2 y_{j+2})^2 + (\Delta^2 y_{j+3})^2 \\ \leq 2(\Delta^2\delta_j)^2 + 2(\Delta^2\delta_{j+1})^2 + 2(\Delta^2\delta_{j+2})^2 + 2(\Delta^2\delta_{j+3})^2 + 2(\Delta^2\delta_{j+4})^2.$$

In this case $s_j$, $s_{j+1}$, $s_{j+2}$, $s_{j+3}$, and $s_{j+4}$ appear as second, third, or fourth elements of a "bad" quadruple. Since the configuration (3.23) starts with $(+,+)$ and ends with $(-,-)$, it cannot share two elements with a "bad" quadruple outside the configuration. This means that none of $s_j$, $s_{j+1}$, $s_{j+2}$, $s_{j+3}$, and $s_{j+4}$ can be a second, third, or fourth element of a "bad" quadruple outside the configuration. In other words, none of $(\Delta^2\delta_j)^2$, $(\Delta^2\delta_{j+1})^2$, $(\Delta^2\delta_{j+2})^2$, $(\Delta^2\delta_{j+3})^2$, and $(\Delta^2\delta_{j+4})^2$ can appear in a "long" inequality other than (3.24) and (3.25). Since, $s_j$, $s_{j+1}$, $s_{j+2}$, and $s_{j+3}$ are middle elements of "bad" quadruples, $(\Delta^2\delta_{j+1})^2$, $(\Delta^2\delta_{j+2})^2$, and $(\Delta^2\delta_{j+3})^2$ cannot appear in "short" inequalities either. Thus $b_{j+1} = b_{j+2} = b_{j+3} = 2$. Also, $(\Delta^2\delta_j)^2$ cannot appear in a "short" inequality. The only way this could happen is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_j)^2$, which is impossible since $j-1$ is either in Case I $(+,+,+)$ or Case V $(-,+,+)$, depending on what $s_{j-2}$ is, and in both cases the short inequality is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_{j-1})^2$. Thus $b_j = 2$. Similarly, $(\Delta^2\delta_{j+4})^2$ cannot appear in a "short" inequality. The only way this could happen is $(\Delta^2 y_{j+4})^2 \leq (\Delta^2\delta_{j+4})^2$, which is impossible since $j+4$ is either in Case VII $(-,-,+)$ or Case VIII $(-,-,-)$, depending on what $s_{j+5}$ is, and in both cases the short inequality is $(\Delta^2 y_{j+4})^2 \leq (\Delta^2\delta_{j+5})^2$. Thus $b_{j+4} = 2$. This concludes the analysis of the configuration (3.23).

Now, let $s_j$ be a second, third, or fourth element of a "bad" quadruple but not an element of a configuration (3.23). This means that $(\Delta^2\delta_j)^2$ appears in exactly one "long" inequality (it cannot be a second, third, or fourth element of two distinct "bad" quadruples). If $s_j$ is a third element of a "bad" quadruple, then $s_{j-1}$ and $s_j$ are the middle elements of the quadruple and $(\Delta^2\delta_j)^2$ does not appear in a short inequality. Thus, $b_j \leq 2$ in this case. The cases when $s_j$ is a second or fourth element of a "bad" quadruple need separate consideration.

1. $s_j$ is a second element of a Type A quadruple $(+,+,-,-)$. The only way $(\Delta^2\delta_j)^2$ could appear in a "short" inequality is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_j)^2$, which is impossible since $j-1$ is either in Case I $(+,+,+)$ or Case V $(-,+,+)$, depending on what $s_{j-2}$ is, and in both cases the short inequality is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_{j-1})^2$. Thus $b_j = 2$.
2. $s_j$ is a fourth element of a Type A quadruple $(+,+,-,-)$. The only way $(\Delta^2\delta_j)^2$ could appear in a "short" inequality is $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_j)^2$, which is impossible since $j$ is either in Case VII $(-,-,+)$ or Case VIII $(-,-,-)$, depending on what $s_{j+1}$ is, and in both cases the short inequality is $(\Delta^2 y_j)^2 \leq (\Delta^2\delta_{j+1})^2$. Thus $b_j = 2$.
3. $s_j$ is a second element of a Type B quadruple $(+,+,-,+)$. Here the argument is word-by-word like in part 1. The only way $(\Delta^2\delta_j)^2$ could appear in a "short" inequality is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_j)^2$, which is impossible since $j-1$ is either in Case I $(+,+,+)$ or Case V $(-,+,+)$, depending on what $s_{j-2}$ is, and in both cases the short inequality is $(\Delta^2 y_{j-1})^2 \leq (\Delta^2\delta_{j-1})^2$. Thus $b_j = 2$.
4. $s_j$ is a fourth element of a Type B quadruple $(+,+,-,+)$. Since the coefficient of $(\Delta^2\delta_j)^2$ in the corresponding "long" inequality (3.20) is 1 and $(\Delta^2\delta_j)^2$ could appear in at most one "short" inequality, we conclude $b_j \leq 2$.
5. $s_j$ is a second element of a Type C quadruple $(-,+,-,-)$. Since the coefficient of $(\Delta^2\delta_j)^2$ in the corresponding "long" inequality (3.21) is 1 and $(\Delta^2\delta_j)^2$ could appear in at most one "short" inequality, we conclude $b_j \leq 2$.

6. $s_j$ is a fourth element of a Type C quadruple $(-,+,-,-)$. Here the argument is word-by-word like in Case 2. The only way $(\Delta^2\delta_j)^2$ could appear in a "short" inequality is $(\Delta^2 y_j)^2 \le (\Delta^2\delta_j)^2$, which is impossible since $j$ is either in Case VII $(-,-,+)$ or Case VIII $(-,-,-)$, depending on what $s_{j+1}$ is, and in both cases the short inequality is $(\Delta^2 y_j)^2 \le (\Delta^2\delta_{j+1})^2$. Thus $b_j = 2$.

We have shown that in all six cases $b_j \le 2$ for $j \in [1, N]$, which completes the proof of Lemma 3.4 and Theorem 3.2. $\square$

Now, we continue with the general case, the proof of Theorem 3.1. That is, we want to show that the $l_2$ norms inequality

$$\|\{\delta'_j\}\|_{l_2} \le \|\{\delta_j\}\|_{l_2} \tag{3.26}$$

holds for *any* initial sequence $\{\delta_j\}$ with finite $l_2$ norm. We consider the sequence $\{w_j\}$ and restrict the index $j$ to a maximal subset $\Lambda_m$ on which the piecewise constant function $w$ is monotone, recalling that $\delta_j = w_j - w_{j-1}$. Given a sequence $\{w_j\}$, we can decompose it into monotone subsequences. This decomposition also gives a decomposition of the sequence $\{\delta_j\}$ into subsequences such that in each subsequence all jumps have the same sign (nonnegative or nonpositive). Without any limitations, we assume that the jumps $\{\delta_j\}$ are nonnegative for all $l \le j \le r$, $\delta_{l-1} < 0$, and $\delta_{r+1} < 0$. That is, $w_{l-1}$ is a local minimum and $w_r$ is a local maximum of the piecewise constant function $w$. Let $w^m$ be the following piecewise constant correction of $w$:

$$w^m_j := \begin{cases} w_j & \text{if} \quad l \le j \le r, \\ w_{l-1} & \text{if} \quad j < l, \\ w_r & \text{if} \quad j > r. \end{cases} \tag{3.27}$$

Note that $\Lambda_m = \{j : l \le j \le r+1\}$, and the jump sequence $\delta^m := \{\delta^m_j\}$ of $w^m$ is given by

$$\delta^m_j := \begin{cases} w_j - w_{j-1} & \text{if} \quad l \le j \le r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.28}$$

Hence, we have a sequence of monotone functions $\{w^m\}$ and the corresponding jump sequences $\{\delta^m_j\}_j$ such that

$$\sum_m \sum_{j\in\Lambda_m} \|\{\delta^m_j\}\|^2_{l_2} = \sum_{m,j\in\mathbb{Z}} \|\{\delta^m_j\}\|^2_{l_2} = \|\{\delta_j\}\|^2_{l_2},$$

because the sequence of the jumps of $\{\delta_j\}$ is decomposed into disjoint jump subsequences $\{\delta^m_j\}$. There are two types of jumps $\delta'_j$. A jump $\delta'_j$ is of *type* 1 if it is equal to the jump $\delta'_j(\delta^m)$, that is, the jump generates with the starting sequence $\{\delta^m_j\}$, where the index $m$ such that $j \in \Lambda_m$. A jump is of *type* 2 if it is not of *type* 1. Note that a *type* 2 jump $\delta'_{j^*}$ occurs only inside an interval which contains a strict local extremum. Near a local extremum we have two nonzero jumps, say $\delta^l_{j^*}$ and $\delta^r_{j^*}$, generated by the two monotone $w^m$'s with index sets finishing/starting with $j^*$. It is easy to verify that

$$|\delta'_{j^*}| = \big||\delta^l_{j^*}| - |\delta^r_{j^*}|\big|.$$

Hence, we have that

$$(\delta'_{j^*})^2 < (\delta^l_{j^*})^2 + (\delta^r_{j^*})^2,$$

and we conclude that

$$\sum_j (\delta'_j)^2 \le \sum_m \sum_{j\in\Lambda_m} (\delta'_j(\delta^m))^2 \le \sum_m \sum_{j\in\Lambda_m} (\delta^m_j)^2 = \sum_n (\delta_j)^2,$$

where we use the notation $\delta'_j(\delta^m)$ for the new jumps generated by $\delta^m$. It is also easy to prove a local inequality but with index set for $\delta'_j$ starting from an interval right after an extremum and finishing right before one. ∎

**4. Error estimates for linear flux.** Recall that $u$ is the entropy solution to the conservation law $u_t + f(u)_x = 0$ with initial condition $u^0$, and $v$ is the numerical solution described in (2.12). In the case of linear flux and $0 \le \theta \le 1$, the formula for the new averages of the minmod scheme is given in (3.1), and the conservation law (2.1) reduces to

$$u_t + au_x = 0. \tag{4.1}$$

Let $S_\tau$ be the shift operator defined by $S_\tau g(\cdot) := g(\cdot - \tau)$. Then the exact solution of (4.1) at time $t$ for any initial data $u^0$ is $u(\cdot, t) = S_{at}u^0$. Let $A_h$ be the averaging operator defined on a uniform partition by $A_h g|_I := \frac{1}{h}\int_I g(s)ds$, where $|I| = h$. It will be useful to define a global approximate solution $v$. We first define the approximate solution at discrete times by $v^n := v(\cdot, n\Delta t)$, $n = 0, 1, \ldots, N$, in the following way: (i) $v^0 := u^0$; (ii) $v^n := S_{a\Delta t}P_h v^{n-1}$, where for odd $n$, $1 \le n \le N$, $P_h v$ is the linear function on $I_j := (x_{j-1/2}, x_{j+1/2})$ defined in (2.3) with the minmod slopes (2.8), and for even $n$ we have the analogous definition of $P_h$ on the shifted partition $\{I_{j+1/2} | \, j \in \mathbb{Z}\}$. Note that $P_h v^n = P_h A_h v^n$ because the piecewise linear projection $P_h$, defined in (2.3) and (2.8), is based only on the averages of $v^n$ on the corresponding partition. The formula (3.1) for the new cell averages can be written as

$$v^n_{j+1/2} = A_h(v^n)|_{I_{j+1/2}} = A_h(S_{a\Delta t}P_h v^{n-1})|_{I_{j+1/2}}$$

for odd $n$, with $A_h$ based on the staggered partition $\{I_{j+1/2} | \, j \in \mathbb{Z}\}$ and $P_h$ based on regular partition $\{I_j | \, j \in \mathbb{Z}\}$. For even $n$, we have the same sequence of operators but on the reversed partitions. The global approximate solution $v$ is defined by $v(\cdot, n\Delta t) = v^n$ and $v(\cdot, t) = S_{a(t-n\Delta t)}(P_h v^n)$ for $n\Delta t < t \le (n+1)\Delta t$ and $n = 0, 1, \ldots, N-1$. That is, $v$ solves (4.1) exactly for $n\Delta t < t \le (n+1)\Delta t$ with initial data $P_h v^n$, $n = 0, 1, \ldots, N-1$.

In order to describe the next result, we need to introduce some notation. A function $g$ is of bounded variation, i.e., $g \in \mathrm{BV}(\mathbb{R})$, if

$$|g|_{\mathrm{BV}(\mathbb{R})} := \sup \sum_{i=1}^{n} |g(x_{i+1}) - g(x_i)| \; < \; \infty,$$

where the supremum is taken over all finite sequences $x_1 < \cdots < x_n$ in $\mathbb{R}$. Functions of bounded variation have at most countably many discontinuities, and their left and right limits $g(x^-)$ and $g(x^+)$ exist at each point $x \in \mathbb{R}$. Since the values of the initial condition $u^0$ on a set of measure zero have no influence on the numerical solution $v$ and the entropy solution $u$, it is desirable to replace the seminorm $|\cdot|_{\mathrm{BV}(\mathbb{R})}$ by a similar quantity independent of the function values on sets of measure zero. The standard approach in conservation laws is to consider the space $\mathrm{Lip}(1, \mathrm{L}^1(\mathbb{R}))$ of all functions $g \in L^1(\mathbb{R})$ such that the seminorm

$$|g|_{\mathrm{Lip}(1,\mathrm{L}^1(\mathbb{R}))} := \limsup_{s>0} \frac{1}{s} \int_{\mathbb{R}} |g(x+s) - g(x)| \, dx \tag{4.2}$$

is finite. It is clear that $|g|_{\mathrm{Lip}(1,\mathrm{L}^1(\mathbb{R}))}$ will not change if $g$ is modified on a set of measure zero. At the same time the above two seminorms are equal for functions $g \in \mathrm{BV}(\mathbb{R})$ such that the value of $g$ at a point of discontinuity lies between $g(x^-)$ and $g(x^+)$ (see Theorem 9.3 in [5]). Similarly, we define the space $\mathrm{Lip}(1, \mathrm{L}^p(\mathbb{R}))$, $1 \le p \le \infty$, which is the set of all functions $g \in L^p(\mathbb{R})$ for which

$$\|g(\cdot - s) - g(\cdot)\|_{L^p(\mathbb{R})} \le Ms, \quad s > 0. \tag{4.3}$$

The smallest $M \ge 0$ for which (4.3) holds is $|g|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}$. It is easy to see that in the case $p = 1$ the seminorm given in (4.3) is the same as the one in (4.2). In the case $p > 1$, the space $\mathrm{Lip}(1, \mathrm{L}^p(\mathbb{R}))$ is essentially the same as $W^1(L^p(\mathbb{R}))$; see [5] for details. With this notation, we have the following result.

THEOREM 4.1. *Let $u(x,t) = u(x - at, 0)$ be the solution to* (2.1) *with linear flux $f(z) = az$, and let $v$ be the numerical solution described in* (3.1) *with $0 \le \theta \le 1$. If the CFL condition* (2.5) *is satisfied, $t_n = n\Delta t$, $0 \le n \le N$, and $T = N\Delta t$, we have*

$$\|u(\cdot, T) - v(\cdot, T)\|_{L^p(\mathbb{R})} \le C(Nh)^{1/2} h^{1/2} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))} \tag{4.4}$$

*for $p = 1, 2$, where $C$ is an absolute constant.*

*Proof.* The $L_1$ estimate is based on the TVD property of the numerical solution $v$, and the $L_2$ estimate is based on the $l_2$ stability of the jumps proved in Theorem 3.1. Both estimates use a dual argument similar to the one in [19], and in the proof we use an index $p$, where $p \in \{1, 2\}$. Note that we consider the case of linear flux, and the usual Lip+ stability requirement is not needed in the dual approach because the negative norm stability (4.6) holds for any initial data (not just Lip+). In the proof, $C$ will be an absolute constant that can be different at different places.

Let $e(x,t) := u(x,t) - v(x,t)$ and $E(x,t) := \int_{-\infty}^{x} e(s,t)ds$, where we assume that $u^0 \in L^1(\mathbb{R})$ to guarantee that $E$ is well defined for all $(x,t) \in \mathbb{R} \times (0,T)$. We have that $E$ also satisfies (4.1) for $n\Delta t < t \le (n+1)\Delta t$ with initial data $\int_{-\infty}^{x} u(s, n\Delta t) - P_h v^n(s)ds$, $n = 0, 1, \ldots, N-1$. For a function $g \in L^1(\mathbb{R})$ and $1 \le p \le \infty$, we define a *minus one* norm in the following way:

$$\|g\|_{-1,p} := \left\| \int_{-\infty}^{\cdot} g(s)\, ds \right\|_{L^p(\mathbb{R})}. \tag{4.5}$$

It is easy to verify that for any $\tau \in \mathbb{R}$

$$\|S_\tau g\|_{-1,p} = \|g\|_{-1,p}. \tag{4.6}$$

Recall that $T = N\Delta t$. Then we have the representations $u(\cdot, T) = (S_{a\Delta t})^N u^0$ and $v(\cdot, T) = (S_{a\Delta t} P_h)^N u^0$. Using (4.6), we have

$$\begin{aligned} \|e(\cdot, T)\|_{-1,p} &= \|(S_{a\Delta t})^N u^0 - (S_{a\Delta t} P_h)^N u^0\|_{-1,p} \\ &= \|(S_{a\Delta t})^{N-1} u^0 - P_h (S_{a\Delta t} P_h)^{N-1} u^0\|_{-1,p}, \end{aligned}$$

and by the triangle inequality we obtain

$$\begin{aligned} \|e(\cdot, T)\|_{-1,p} \le{}& \|(S_{a\Delta t})^{N-1} u^0 - (S_{a\Delta t} P_h)^{N-1} u^0\|_{-1,p} \\ &+ \|P_h (S_{a\Delta t} P_h)^{N-1} u^0 - (S_{a\Delta t} P_h)^{N-1} u^0\|_{-1,p}. \end{aligned} \tag{4.7}$$

Let $e^n = ((S_{a\Delta t})^n - (S_{a\Delta t} P_h)^n)\, u^0$, $n = 0, 1, \ldots, N$. Then (4.7) is equivalent to

$$\|e^N\|_{-1,p} \le \|e^{N-1}\|_{-1,p} + \|P_h v^{N-1} - v^{N-1}\|_{-1,p}, \tag{4.8}$$

and applying (4.8) for $n = N, N-1, \dots, 1$, we get

$$\|e^N\|_{-1,p} \le \sum_{n=1}^{N-1} \|P_h v^n - v^n\|_{-1,p} \tag{4.9}$$

because $e^0 \equiv 0$. To prove the error estimates, we need the following technical lemma.

LEMMA 4.2. *For any $p \in \{1,2\}$ and any $n = 0, 1, \dots, N$, we have*

$$\text{(i)} \quad \|\{\delta_j^n\}\|_{l_p} \le h^{1-\frac{1}{p}} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))},$$

$$\text{(ii)} \quad \|P_h v^n - A_h v^n\|_{-1,p} \le \left(\frac{2}{p+1}\right)^{\frac{1}{p}} h^{1+\frac{1}{p}} \|\{\delta_j\}\|_{l_p},$$

$$\text{(iii)} \quad \|A_h v^n - v^n\|_{-1,p} \le \left(\frac{4}{p+1}\right)^{\frac{1}{p}} h^2 |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}.$$

*Proof.* The inequalities (i) and (ii) follow by standard arguments; therefore, we only prove (i) in the case $p = 2$ and omit the rest because their proofs are similar. Recall that $\delta_j^n = v_j^n - v_{j-1}^n$, and by Theorem 3.1 we have

$$\left(\sum_j (\delta_j^n)^2\right)^{1/2} \le \left(\sum_j (\delta_j^0)^2\right)^{1/2},$$

where $\delta_j^0 = u_j^0 - u_{j-1}^0$, $u_j^0 := \frac{1}{h}\int_{I_j} u^0(s)ds$. Hence, to prove (i) for $p = 2$, we need to prove

$$\sum_j (\delta_j^0)^2 \le h|u^0|^2_{\mathrm{Lip}(1,\mathrm{L}^2(\mathbb{R}))}.$$

Since

$$\sum_j (\delta_j^0)^2 = \sum_j \left(\frac{1}{h}\int_{I_j} (u^0(s+h) - u^0(s))ds\right)^2 \le h^{-2} \sum_j \left(\int_{I_j} |u^0(s+h) - u^0(s)|ds\right)^2,$$

and since by the Cauchy–Schwarz inequality $\left(\int_{I_j} |u^0(s+h) - u^0(s)|ds\right)^2 \le h \int_{I_j} |u^0(s+h) - u^0(s)|^2 ds$, we obtain

$$\sum_j (\delta_j^0)^2 \le h^{-1} \int_{\mathbb{R}} |u^0(s+h) - u^0(s)|^2 ds. \tag{4.10}$$

From (4.3), we have $\int_{\mathbb{R}} |u^0(s+h) - u^0(s)|^2 ds \le h^2 |u^0|^2_{\mathrm{Lip}(1,\mathrm{L}^2(\mathbb{R}))}$, and using that in (4.10), we conclude

$$\sum_j (\delta_j^0)^2 \le h|u^0|^2_{\mathrm{Lip}(1,\mathrm{L}^2(\mathbb{R}))},$$

which proves (i) for $p = 2$. To prove (iii), we note that

$$|A_h v^n - v^n|_{I_j} \le \max_{x \in I_j} v^n(x) - \min_{x \in I_j} v^n(x), \tag{4.11}$$

and because $v^n = S_{a\Delta t} v^{n-1}$, we have that

$$\max_{x\in I_j} v^n(x) - \min_{x\in I_j} v^n(x) \le 2\max(|\delta_{j-1}^{n-1}|, |\delta_j^{n-1}|).$$

The rest of the proof of (iii) is analogous to the proof of (i). ∎

Combining (i)–(iii), we have $\|P_h v^n - v^n\|_{-1,p} \le Ch^2 |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}$, and after applying the above inequality in (4.9), we derive the following estimate:

$$\|e^N\|_{-1,p} \le CNh^2 |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}. \tag{4.12}$$

Because $v^N \notin \mathrm{Lip}(1, \mathrm{L}^2(\mathbb{R}))$, we approximate $v^N$ by

$$\tilde{v} := \frac{1}{h}\int_{x-h/2}^{x+h/2} A_h v^N(s)\, ds.$$

Similar to Lemma 4.2, it is easy to verify that for $p \in \{1, 2\}$ we have

$$\|\tilde{v} - v^N\|_{-1,p} \le Ch^2 |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}, \tag{4.13}$$

$$\|\tilde{v} - v^N\|_{L^p(\mathbb{R})} \le Ch |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}, \tag{4.14}$$

and

$$|\tilde{v}|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))} \le |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}. \tag{4.15}$$

Let $\tilde{e} := u(\cdot, T) - \tilde{v}$. Then $\|\tilde{e}\|_{-1,p} \le \|e^N\|_{-1,p} + \|\tilde{v} - v^N\|_{-1,p}$, and combining the estimates (4.12) and (4.13), we have

$$\|\tilde{e}\|_{-1,p} \le CNh^2 |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}. \tag{4.16}$$

Kolmogorov–Landau inequalities in $L^p(\mathbb{R})$ (p. 156 in [5]) for the functions $\tilde{E}(x) := \int_\infty^x \tilde{e}(s)\, ds$, $\tilde{E}'$, and $\tilde{E}''$ give

$$\|\tilde{e}\|_{L^p} = \|\tilde{E}'\|_{L^p} \le \sqrt{2}\|\tilde{E}\|_{L^p}^{1/2}\|\tilde{E}''\|_{L^p}^{1/2} = \sqrt{2}\|\tilde{e}\|_{-1,p}^{1/2}|\tilde{e}|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}^{1/2}.$$

Using (4.16) and (4.15), we arrive at

$$\|\tilde{e}\|_{L^p} \le C(Nh)^{1/2} h^{1/2} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}. \tag{4.17}$$

Finally, by the triangle inequality,

$$\|e\|_{L^p} \le \|\tilde{e}\|_{L^p} + \|\tilde{v} - v^N\|_{L^p(\mathbb{R})} \le C(Nh)^{1/2} h^{1/2} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))},$$

and we combine (4.14) and (4.17) to conclude

$$\|u(\cdot, T) - v(\cdot, T)\|_{L^p(\mathbb{R})} = \|e\|_{L^p} \le C(Nh)^{1/2} h^{1/2} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}.$$

Note that $C$ can be computed explicitly and is not very big ($C < 20$). In the case $p = 2$ and $u^0 \notin L^1(\mathbb{R})$, we get the same error estimate via an approximation procedure because the estimate is independent of the $L^1$ norm. ∎

COROLLARY 4.3. *In the case of $Nh \le C$, we get the convergence rate*

$$\|u(\cdot, T) - v(\cdot, T)\|_{L^p(\mathbb{R})} \le Ch^{1/2} |u^0|_{\mathrm{Lip}(1,\mathrm{L}^p(\mathbb{R}))}$$

*for $p = 1$ and $p = 2$.*

The $L^1$ estimate is not new—it follows from the arguments in [19]—but the 1/2 rate in $L^2$ is new. Note that, using the $L^1$ estimate, by interpolation arguments we get only a 1/4 rate in $L^2$. The rate 1/2 is optimal for the case $\theta = 0$ because the numerical method in that case reduces to the LxF scheme, a special case of a monotone scheme. In the case $p = 1$, the sharpness of the 1/2 bound is given in [20], with an extension to the nonlinear case in [17]. The sharpness in the case $p = 2$ follows from the more general result for formal first-order linear schemes; see [4]. The case $\theta > 0$ is more complicated because the schemes are nonlinear, and it will be addressed elsewhere.

**5. Numerical examples.** In this section, we present numerical evidence for the new $l_2$ stability result we proved in section 3. Our numerical tests suggest that in the case of linear flux the NT schemes do not increase the $l_2$ norm of the jumps for either $\theta \leq 1$ (as proved in Theorem 3.1) or for $1 < \theta \leq 2$. In the case of convex flux, we numerically observe the one-sided analogue of this property. We now give generic examples for this $l_2$ stability in the linear and convex case.

*Example* 1. Linear equation. We take a piecewise linear initial condition $u^0$ (see top left of Figure 5.1) and compare three different approximate solutions. It is easy to see that for a bigger value of $\theta$ we get smaller numerical diffusion (see top right panel of Figure 5.1). The other two plots on Figure 5.1 give the behaviors of the $l_2$ and the $l_\infty$ norms of the jumps in time where the time is rescaled from $[0, 0.15]$ to $[0, 1]$ and the $l_2$ norm is also rescaled. Note the oscillatory behavior of the $l_\infty$ norm

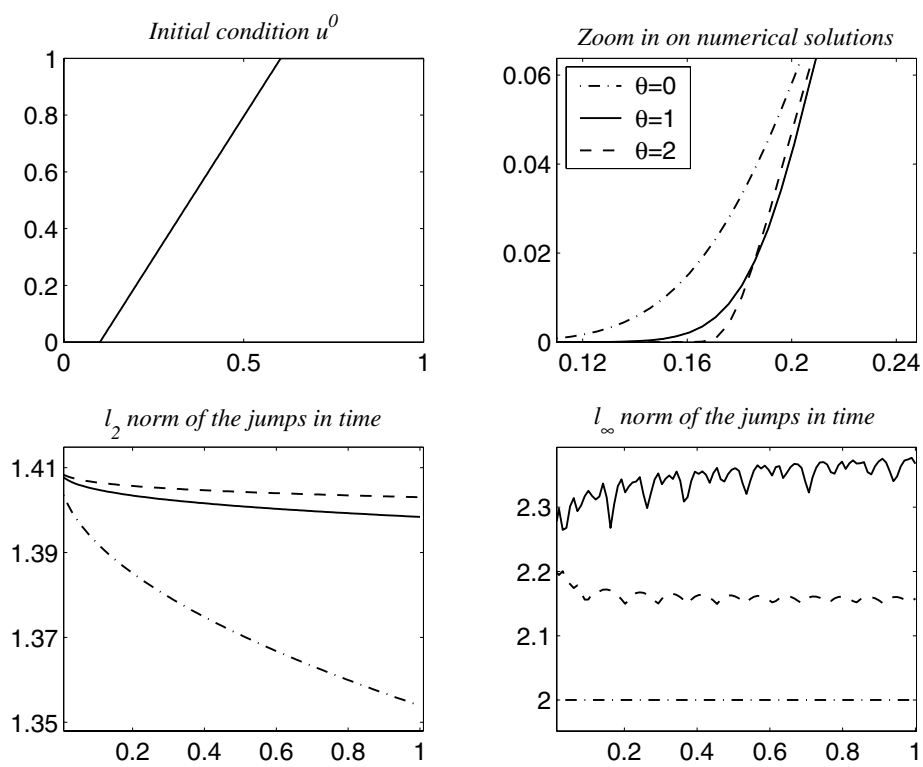


FIG. 5.1. $u_t + 0.5\, u_x = 0$. *The solid line represents* $\theta = 1$, *the dashed line represents* $\theta = 2$, *and the dash-dotted line stands for* $\theta = 0$, *the staggered LxF scheme. The values we used are* $\Delta x = 0.005$, $\lambda = 0.15$, *final time* $T = 0.15$, *and the flux is* $f(u) = 0.5u$.

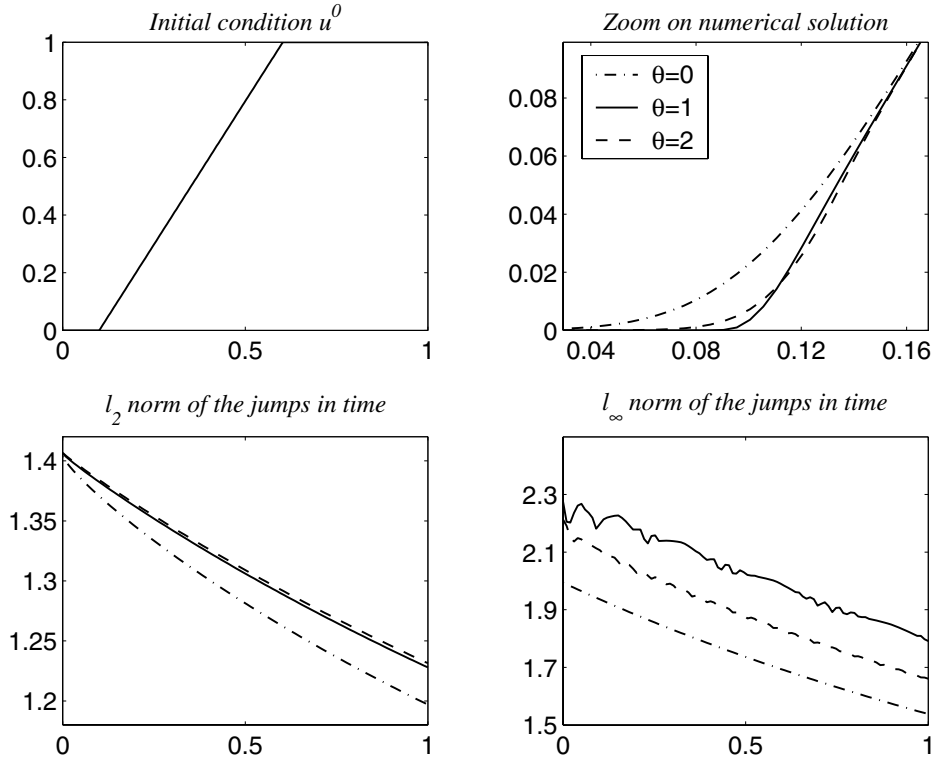


FIG. 5.2. $u_t + (0.5\,u^2)_x = 0$.

and the monotonicity of the $l_2$ norm for $\theta = 1, 2$.

The presence of shocks or local extrema in the initial data will only make the decrease of the $l_2$ norm of the jumps faster in the beginning, and then for large time the $l_2$ norm will decrease very slowly again. In some sense, the total amount of numerical diffusion is given in the decrease of the $l_2$ norm. In the so-called second-order methods (like $\theta = 1, 2$), the amount of diffusion is much smaller than in first-order methods represented here by the LxF scheme ($\theta = 0$). We will address this issue in a different paper and use it to improve the error estimate for $\theta = 1$.

*Example* 2. Burgers' equation (see Figure 5.2). We consider the same initial data, numerical schemes, $\Delta x$, $\lambda$, and $T$, as in the first example. Note again the oscillatory behavior of the $l_\infty$ norm and the monotonicity of the $l_2$ norm for $\theta = 1, 2$. This is a generic case of nondecreasing initial data which corresponds to a region of spreading of the characteristics.

The nonlinearity of the flux in such regions helps to decrease overall any norm of the jumps. In the case of a general initial condition, the $l_2$ norm of the jumps decreases in every region of rarefaction. That is, for convex flux numerical schemes decrease the *one-sided* $l_2$ norm of the jumps,

$$\sum_j (v_j^{n+1} - v_{j-1}^{n+1})_+^2 \le \sum_j (v_j^n - v_{j-1}^n)_+^2.$$

It is important to note that in the case of convex/concave flux extreme values separate the regions of rarefactions from the regions of shocks, and we observe numerically that

the $l_2$ norm of the jumps decreases in every interval where the numerical solution is nondecreasing/nonincreasing.

In the nonconvex case (at least one inflection point), the situation is quite different. In one interval of monotonicity we can have both shocks and rarefaction waves. In that case, the NT scheme with $\theta = 2$ converges to a wrong weak solution even for the Buckley–Leverett problem; see Example 3 in [8]. Our numerical tests show that the NT scheme gives a wrong solution to that problem for any value of $\theta \geq 1.2$ and in general it looks like the biggest reliable value of $\theta$ for a nonconvex flux is $\theta = 1$.

**Acknowledgments.** The authors are grateful to Ronald DeVore for his inspiring discussions and constant support. We also thank the anonymous referees. Their comments and suggestions helped improve the paper.

## REFERENCES

[1] F. Bouchut, Ch. Bourdarias, and B. Perthame, *A MUSCL method satisfying all entropy inequalities*, Math. Comp., 65 (1996), pp. 1439–1461.

[2] F. Bouchut and B. Perthame, *Kružkov's estimates for scalar conservation laws revisited*, Trans. Amer. Math. Soc., 350 (1998), pp. 2847–2870.

[3] Y. Brenier and S. Osher, *The discrete one-sided Lipschitz condition for convex scalar conservation laws*, SIAM J. Numer. Anal., 25 (1988), pp. 8–23.

[4] P. Brenner, V. Thomée, and L. B. Wahlbin, *Besov Spaces and Applications to Difference Methods for Initial Value Problems*, Lecture Notes in Math. 434, A. Dold and B. Eckmann, eds., Springer-Verlag, Berlin, New York, 1975.

[5] R. A. DeVore and G. G. Lorentz, *Constructive Approximation*, Springer-Verlag, Berlin, 1993.

[6] A. Harten and S. Osher, *Uniformly high order accurate non-oscillatory schemes,* I, J. Appl. Numer. Math., 71 (1987), pp. 279–309.

[7] A. Harten, B. Enquist, S. Osher, and S. R. Chakravarthy, *Uniformly high order accurate essentially non-oscillatory schemes,* III, J. Comput. Phys., 71 (1987), pp. 231–303.

[8] G.-S. Jiang, D. Levy, C.-T. Lin, S. Osher, and E. Tadmor, *High-resolution nonoscillatory central schemes with nonstaggered grids for hyperbolic conservation laws*, SIAM J. Numer. Anal., 35 (1998), pp. 2147–2168.

[9] G.-S. Jiang and E. Tadmor, *Nonoscillatory central schemes for multidimensional hyperbolic conservation laws*, SIAM J. Sci. Comput., 19 (1998), pp. 1892–1917.

[10] P. Lax and B. Wendroff, *Systems of conservation laws*, Comm. Pure Appl. Math., 13 (1960), pp. 217–237.

[11] K. Kopotun, M. Neamtu, and B. Popov, *Weakly non-oscillatory schemes for scalar conservation laws*, Math. Comp., 72 (2003), pp. 1747–1767.

[12] S. N. Kruzhkov, *First order quasi-linear equations in several independent variables*, Math. USSR Sbornik, 10 (1970), pp. 217–243.

[13] A. Kurganov and E. Tadmor, *New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations*, J. Comput. Phys., 160 (2000), pp. 241–282.

[14] N. N. Kuznetsov, *Accuracy of some approximate methods for computing the weak solutions of a first order quasi-linear equation*, USSR Comput. Math. Math. Phys., 16 (1976), pp. 105–119.

[15] H. Nessyahu and E. Tadmor, *Non-oscillatory central differencing for hyperbolic conservation laws*, J. Comput. Phys., 87 (1990), pp. 408–463.

[16] H. Nessyahu and E. Tadmor, *The convergence rate of approximate solutions for nonlinear scalar conservation laws*, SIAM J. Numer. Anal., 29 (1992), pp. 1505–1519.

[17] F. Şabac, *The optimal convergence rate of monotone finite difference methods for hyperbolic conservation laws*, SIAM J. Numer. Anal., 34 (1997), pp. 2306–2318.

[18] C.-W. Shu, *Numerical experiments on the accuracy of ENO and modified ENO schemes*, J. Comput. Phys., 5 (1990), pp. 127–149.

[19] E. Tadmor, *Local error estimates for discontinuous solutions of nonlinear hyperbolic equations*, SIAM J. Numer. Anal., 28 (1991), pp. 891–906.

[20] T. Tang and Z.-H. Teng, *The sharpness of Kuznetsov's $O(\sqrt{\Delta x})$ $L^1$-error estimate for monotone difference scheme*, Math. Comp., 64 (1995), pp. 581–589.