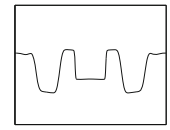


Chapter 3

A Short Course in Difference Methods

Computation will cure what ails you.
— Clifford Truesdell, The Computer, Ruin of Science and Threat to Mankind, 1980/1982

Although front tracking can be thought of as a numerical method, and has indeed been shown to be excellent for one-dimensional conservation laws, it is not part of the standard repertoire of numerical methods for conservation laws. Traditionally, difference methods have been central to the development of the theory of conservation laws, and the study of such methods is very important in applications.

This chapter is intended to give a brief introduction to difference methods for conservation laws. The emphasis throughout will be on methods and general results rather than on particular examples. Although difference methods and the concepts we discuss can be formulated for systems, we will exclusively concentrate on scalar equations. This is partly because we want to keep this chapter introductory, and partly due to the lack of general results for difference methods applied to systems of conservation laws.

3.1 Conservative Methods

We are interested in numerical methods for the scalar conservation law in one dimension. (We will study multidimensional problems in Chapter 4.) Thus we consider

$$u_t + f(u)_x = 0, \qquad u|_{t=0} = u_0. \tag{3.1}$$

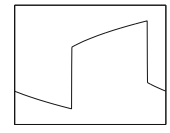
A difference method is created by replacing the derivatives by finite differences, e.g.,

$$\frac{\Delta u}{\Delta t} + \frac{\Delta f(u)}{\Delta x} = 0. \tag{3.2}$$

Here Δt and Δx are small positive numbers. We shall use the notation

$$u_j^n \approx u\,(j\Delta x, n\Delta t) \quad \text{and} \quad u^n = \left(u_{-K}^n, \ldots, u_j^n, \ldots, u_K^n\right),$$

H. Holden, N.H. Risebro, *Front Tracking for Hyperbolic Conservation Laws*,
Applied Mathematical Sciences, DOI 10.1007/978-3-662-47507-2_3



where u_j^n now is our numerical approximation to the solution u of (3.1) at the point $(j\Delta x, n\Delta t)$. Normally, since we are interested in the initial value problem (3.1), we know the initial approximation

$$u_j^0, \quad -K \le j \le K,$$

and we want to use (3.2) to calculate u^n for $n \in \mathbb{N}$. We will not say much about boundary conditions in this book. Often one assumes that the initial data is periodic, i.e.,

$$u_{-K+j}^0 = u_{K+j}^0, \quad \text{for } 0 \le j \le 2K,$$

which gives $u_{-K+j}^n = u_{K+j}^n$. Another commonly used device is to assume that $\partial_x f(u) = 0$ at the boundary of the computational domain. For a numerical scheme this means that

$$f\left(u_{-K-j}^n\right) = f\left(u_{-K}^n\right) \quad \text{and} \quad f\left(u_{K+j}^n\right) = f\left(u_K^n\right) \quad \text{for } j > 0.$$

For nonlinear equations, *explicit* methods are most common. These can be written

$$u^{n+1} = G\left(u^n, \ldots, u^{n-l}\right) \tag{3.3}$$

for some function G. We see that u^{n+1} can depend on the previous $l+1$ approximations $u^n, \ldots, u^{n-l}$. The simplest methods are those with $l = 0$, where $u^{n+1} = G(u^n)$, and we shall restrict ourselves to such methods in this presentation.

◇ Example 3.1 (A nonconservative method)
Consider Burgers's equation written in nonconservative form (writing uu_x instead of $\frac{1}{2}(u^2)_x$)

$$u_t + uu_x = 0.$$

Based on the linear transport equation, if $u_j^n > 0$, a natural discretization of this would be

$$u_j^{n+1} = u_j^n - \lambda u_j^n \left(u_j^n - u_{j-1}^n\right), \tag{3.4}$$

with $\lambda = \Delta t/\Delta x$. Since it is based on the nonconservative formulation, we do not automatically have conservation of u. Indeed,

$$\begin{aligned}
\Delta x \sum_j u_j^{n+1} &= \Delta x \sum_j u_j^n - \lambda \Delta x \sum_j u_j^n \left(u_j^n - u_{j-1}^n\right), \\
&= \Delta x \sum_j u_j^n - \frac{1}{2}\lambda \Delta x \sum_j \left(\left(u_j^n\right)^2 - \left(u_{j-1}^n\right)^2 + \left(u_j^n - u_{j-1}^n\right)^2\right) \\
&= \Delta x \sum_j u_j^n - \frac{1}{2}\lambda \Delta x \sum_j \left(u_j^n - u_{j-1}^n\right)^2.
\end{aligned}$$

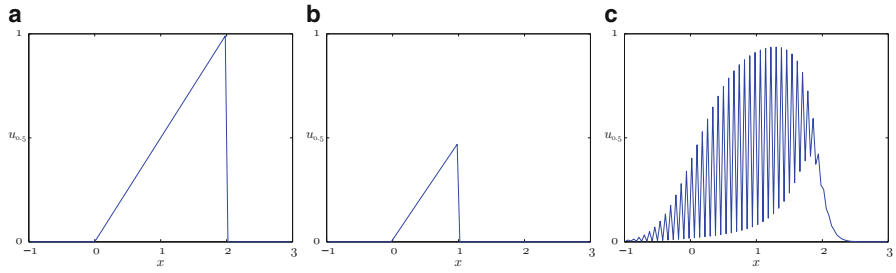


Fig. 3.1 **a** The entropy solution; **b** the scheme (3.4); **c** the scheme (3.5)

This in itself might not seem so bad, since it may happen that $\Delta x \sum_j (u_j^n - u_{j-1}^n)^2$ vanishes as $\Delta x \to 0$. However, let us examine what happens in a specific case. Let the initial data be given by

$$u_0(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The entropy solution to Burgers's equation consists of a rarefaction wave, centered at $x = 0$, and a shock with left value $u = 1$ and right value $u = 0$, starting from $x = 1$ and moving to the right with speed $1/2$. At $t = 2$ the rarefaction wave will catch up with the shock. Thus at $t = 2$ the entropy solution reads

$$u(x, 2) = \begin{cases} \frac{x}{2} & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

We use $u_j^0 = u_0(j\,\Delta x)$ as initial data for the scheme. Then we have that for every j such that $j\,\Delta x > 1$, $u_j^n = 0$ for all $n \ge 0$. So if $N\,\Delta t = 2$, then $u_j^N = 0$, and clearly $u_j^N \not\approx u(j\,\Delta x, 2)$ for $1 \le j\,\Delta x \le 2$. This method simply fails to "detect" the moving shock.

We might think that the situation would be better if we used a (second-order) approximation to u_x instead, resulting in the scheme

$$u_j^{n+1} = \frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right) - \frac{\lambda}{2} u_j^n \left(u_{j+1}^n - u_{j-1}^n\right). \tag{3.5}$$

In practice, this scheme computes something that moves to the right, but the rarefaction part of the solution is not well approximated. In Fig. 3.1 we show how these two nonconservative schemes work on this example. Henceforth, we will not discuss nonconservative schemes. ◇

We call a difference method *conservative* if it can be written in the form

$$u_j^{n+1} = u_j^n - \lambda\left(F\left(u_{j-p}^n, \ldots, u_{j+q}^n\right) - F\left(u_{j-1-p}^n, \ldots, u_{j-1+q}^n\right)\right), \tag{3.6}$$

where

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

The function F is referred to as the *numerical flux*. For brevity, we shall often use the notation

$$G_j(u) = G\left(u_{j-1-p}, \ldots, u_{j+q}\right),$$
$$F_{j+1/2}(u) = F\left(u_{j-p}, \ldots, u_{j+q}\right),$$

so that (3.6) reads

$$u_j^{n+1} = G_j(u^n) = u_j^n - \lambda\left(F_{j+1/2}(u^n) - F_{j-1/2}(u^n)\right). \tag{3.7}$$

The above equation has a nice formal explanation. Set $x_j = j\Delta x$ and $x_{j+1/2} = x_j + \Delta x/2$ for $j \in \mathbb{Z}$. Likewise, set $t_n = n\Delta t$ for $n \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$. Define the interval $I_j = [x_{j-1/2}, x_{j+1/2})$ and the cell $I_j^n = I_j \times [t_n, t_{n+1})$. If we integrate the conservation law

$$u_t + f(u)_x = 0$$

over the cell I_j^n, we obtain

$$\int_{I_j} u(x, t_{n+1})\,dx = \int_{I_j} u(x, t_n)\,dx + \Big(\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2}, t))\,dt - \int_{t_n}^{t_{n+1}} f(u(x_{j-1/2}, t))\,dt\Big).$$

Now defining u_j^n as the average of $u(x, t_n)$ in I_j, i.e.,

$$u_j^n = \frac{1}{\Delta x}\int_{I_j} u(x, t_n)\,dx,$$

we obtain the exact expression

$$u_j^{n+1} = u_j^n - \lambda\Big(\frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2}, t))\,dt - \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(u(x_{j-1/2}, t))\,dt\Big).$$

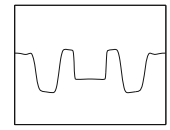
Comparing this with (3.7), we see that it is reasonable that the numerical flux $F_{j+1/2}$ approximates the average flux through the line segment $x_{j+1/2} \times [t_n, t_{n+1}]$. Thus

$$F_{j+1/2}(u^n) \approx \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2}, t))\,dt.$$

With this interpretation of $F_{j+1/2}^n = F_{j+1/2}(u^n)$, equation (3.7) states that the change in the amount of u inside the "volume" I_j equals (approximately) the influx minus the outflux. Methods that can be written on the form (3.7) are often called *finite volume methods*.

If $u(x, t_n)$ is the piecewise constant function

$$u(x, t_n) = u_j^n \text{ for } x \in I_j,$$



we can solve the conservation law exactly for $0 \le t - t_n \le \Delta x/(2\max_u |f'(u)|)$. This is true because the initial data is a series of Riemann problems, whose solutions will not interact in this short time interval. We also see that $f(u(x_{j+1/2}, t))$ is independent of t, and depends only on u_j^n and u_{j+1}^n. So if we set $v = w(x/t)$ to be the entropy solution to

$$v_t + f(v)_x = 0, \quad v(x,0) = \begin{cases} u_j^n & x < 0, \\ u_{j+1}^n & x > 0, \end{cases}$$

then

$$F_{j+1/2}^n = f(w(0)). \tag{3.8}$$

This method is called the *Godunov method*. In general, it is well defined (see Exercise 3.5) for

$$\Delta t \max |f'(u)| \le \Delta x. \tag{3.9}$$

This last condition is called the Courant–Friedrichs–Lewy (CFL) condition.

If $f'(u) \ge 0$ for all u, then $v(0) = u_j^n$, and the Godunov method simplifies to

$$u_j^{n+1} = u_j^n - \lambda \left(f(u_j^n) - f(u_{j-1}^n) \right). \tag{3.10}$$

This is called the *upwind method*.

Conservative methods have the property that $\int u\,dx$ is conserved, since

$$\sum_{j=-K}^{K} u_j^{n+1} \Delta x = \sum_{j=-K}^{K} u_j^n \Delta x - \Delta t \left(F_{K+1/2}^n - F_{-K-1/2}^n \right).$$

If we set u_j^0 equal to the average of u_0 over the jth grid cell, i.e.,

$$u_j^0 = \frac{1}{\Delta x} \int_{I_j} u_0(x)\,dx,$$

and for the moment assume that $F_{-K-1/2}^n = F_{K+1/2}^n$, then

$$\int u^n(x)\,dx = \int u_0(x)\,dx. \tag{3.11}$$

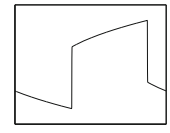
A conservative method is said to be *consistent* if

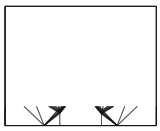
$$F(c, \ldots, c) = f(c), \tag{3.12}$$

and in addition, we demand that F be Lipschitz continuous in all its variables, that is,

$$\left| F(a_{j-p}, \ldots, a_{j+q}) - F(b_{j-p}, \ldots, b_{j+q}) \right| \le L \sum_{i=-p}^{q} \left| a_{j+i} - b_{j+i} \right|, \tag{3.13}$$

for some constant L.





◇ Example 3.2 (Some conservative methods)
We have already seen that the Godunov method (and in particular the upwind method) is an example of a conservative finite volume method.

Another prominent examples is the *Lax–Friedrichs scheme*, usually written

$$u_j^{n+1} = \frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right) - \frac{1}{2}\lambda\left(f\left(u_{j+1}^n\right) - f\left(u_{j-1}^n\right)\right). \tag{3.14}$$

This can be written in conservative form by defining

$$F_{j+1/2}^n = \frac{1}{2\lambda}\left(u_j^n - u_{j+1}^n\right) + \frac{1}{2}\left(f\left(u_j^n\right) + f\left(u_{j+1}^n\right)\right).$$

Some methods, so-called *two-step methods*, use iterates of the flux function. One such method is the *Richtmyer two-step Lax–Wendroff scheme*:

$$F_{j+1/2}^n = f\left(\frac{1}{2}\left(u_{j+1}^n + u_j^n\right) - \frac{\lambda}{2}\left(f\left(u_{j+1}^n\right) - f\left(u_j^n\right)\right)\right). \tag{3.15}$$

Another two-step method is the *MacCormack scheme*:

$$F_{j+1/2}^n = \frac{1}{2}\left(f\left(u_j^n - \lambda(f\left(u_{j+1}^n\right) - f\left(u_j^n\right)\right)\right) + f\left(u_j^n\right)\right). \tag{3.16}$$

The Lax–Friedrichs and Godunov schemes are both of first order in the sense that the local truncation error is of order one. (We shall return to this concept below.) On the other hand, both the Lax–Wendroff and MacCormack methods are of second order. In general, higher-order methods are good for smooth solutions, but they also produce solutions that oscillate in the vicinity of discontinuities. See Sect. 3.2. Lower-order methods have "enough diffusion" to prevent oscillations. Therefore, one often uses *hybrid methods*. These methods usually consist of a linear combination of a lower- and a higher-order method. The numerical flux is then given by

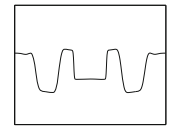
$$F_{j+1/2}^n = \theta_{j+1/2}(u^n) F_{L,j+1/2}^n + \left(1 - \theta_{j+1/2}(u^n)\right) F_{H,j+1/2}^n, \tag{3.17}$$

where F_L denotes a lower-order numerical flux, and F_H a higher-order numerical flux. The function $\theta_{j+1/2}$ is close to zero where u^n is smooth, and close to one near discontinuities. Needless to say, choosing appropriate θ's is a discipline in its own right. We have implemented a method (called *fluxlim* in Fig. 3.2) that is a combination of the (second-order) MacCormack method and the (first-order) Lax–Friedrichs scheme, and this scheme is compared with the "pure" methods in this figure. We somewhat arbitrarily used

$$\theta_{j+1/2} = 1 - \frac{1}{1 + \left|D_+ D_- u_j^n\right|},$$

where $D_\pm$ are the forward and backward divided differences,

$$D_\pm u_j = \pm\frac{u_{j\pm1} - u_j}{\Delta x},$$



so that D_+D_- is an approximation to the second derivative of u with respect to x, namely

$$D_+D_-u_j = \frac{u_{j+1} - 2u_j + u_{j-1}}{\Delta x^2}.$$

Another approach is to try to generalize Godunov's method by replacing the piecewise constant data u^n by a smoother function. The simplest such replacement is by a piecewise linear function. To obtain a proper generalization, one should then solve a generalized "Riemann problem" with linear initial data to the left and right. While this is difficult to do exactly, one can use approximations instead. One such approximation leads to the following method:

$$F_{j+1/2} = \frac{1}{2}\left(g_j + g_{j+1}\right) - \frac{1}{2\lambda}\Delta_+ u_j^n.$$

Here $\Delta_\pm u_j^n = \pm(u_{j\pm1}^n - u_j^n) = \Delta x\, D_\pm u_j^n$, and

$$g_j = f(u_j^{n+1/2}) + \frac{1}{2\lambda}\tilde{u}_j,$$

where

$$\tilde{u}_j = \text{minmod}\left(\Delta_- u_j^n, \Delta_+ u_j^n\right),$$

$$u_j^{n+1/2} = u_j^n - \frac{\lambda}{2} f'\left(u_j^n\right)\tilde{u}_j,$$

and

$$\text{minmod}(a,b) := \frac{1}{2}\left(\text{sign}\,(a) + \text{sign}\,(b)\right)\min\{|a|, |b|\}.$$

This method is labeled *slopelim* in the figures. Now we show how these methods perform on two test examples. In both examples the flux function is given by (see Exercise 2.1)

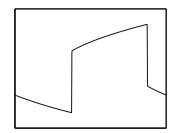
$$f(u) = \frac{u^2}{u^2 + (1-u)^2}. \tag{3.18}$$

The example is motivated by applications in oil recovery, where one often encounters flux functions that have a shape similar to that of f, that is, $f' \geq 0$ and $f''(u) = 0$ at a single point u. The model is called the *Buckley–Leverett* equation. The first example uses initial data

$$u_0(x) = \begin{cases} 1 & \text{for } x \leq 0, \\ 0 & \text{for } x > 0. \end{cases} \tag{3.19}$$

In Fig. 3.2 we show the computed solution at time $t = 1$ for all methods, using 30 grid points in the interval $[-0.1, 1.6]$, and $\Delta x = 1.7/29$, $\Delta t = 0.5\Delta x$. The second example uses initial data

$$u_0(x) = \begin{cases} 1 & \text{for } x \in [0,1], \\ 0 & \text{otherwise}, \end{cases} \tag{3.20}$$



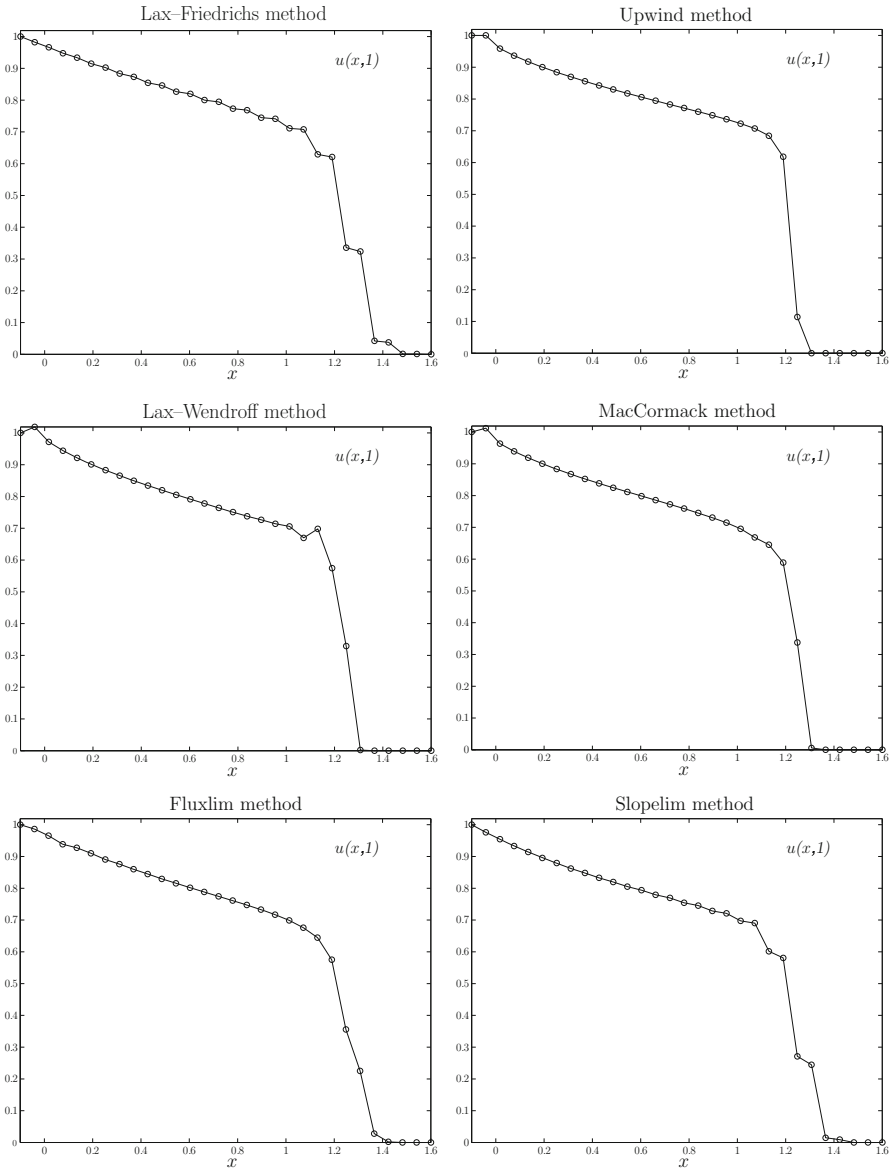
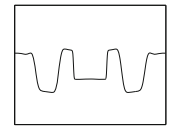


Fig. 3.2 Computed solutions at time $t = 1$ for flux function (3.18) and initial data (3.19)

and 30 grid points in the interval $[-0.1, 2.6]$, $\Delta x = 2.7/29$, $\Delta t = 0.5\Delta x$. In Fig. 3.3 we also show a reference solution computed by the upwind method using 500 grid points. The most notable feature of the plots in Fig. 3.3 is the solutions computed by the second-order methods. We shall show that if a sequence of solutions produced by a consistent conservative method converges, then the limit is a weak solution. The exact solution to both these problems can be calculated by the method of characteristics. ◇



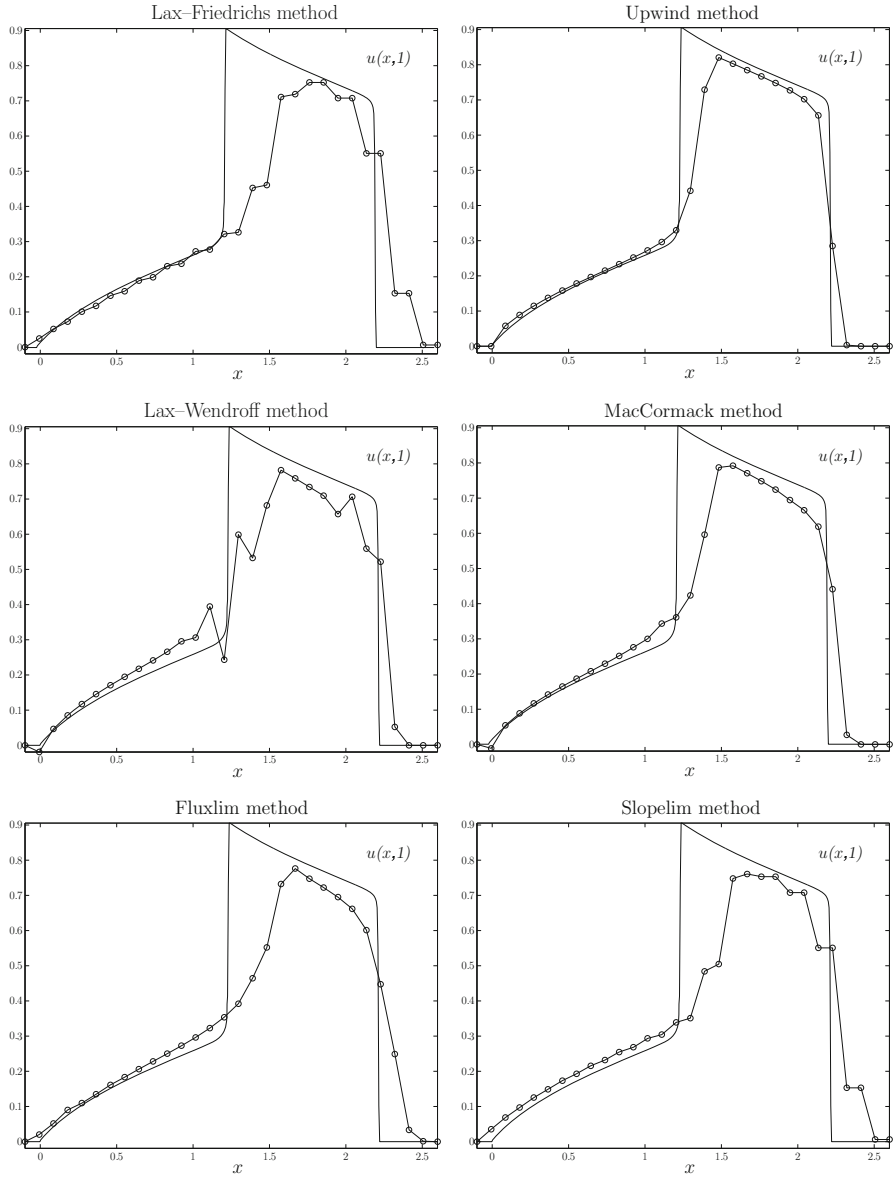
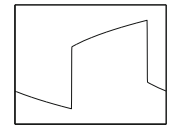


Fig. 3.3 Computed solutions at time $t = 1$ for flux function (3.18) and initial data (3.20)

The *local truncation error* of a numerical method $L_{\Delta t}$ is defined as

$$L_{\Delta t}(x) = \frac{1}{\Delta t}\left(S(\Delta t)u - S_N(\Delta t)u\right)(x), \tag{3.21}$$

where $S(t)$ is the solution operator associated with (3.1), that is, $u = S(t)u_0$ denotes the solution at time t, and $S_N(t)$ is the solution operator associated with the



numerical method, i.e.,

$$S_N(\Delta t)u(x) = u(x) - \lambda \left(F_{j+1/2}(u) - F_{j-1/2}(u)\right).$$

Assuming that we have a *smooth* solution of the conservation law, allowing us to expand all relevant quantities in Taylor series, we say that the method is of kth order if

$$|L_{\Delta t}(x)| = \mathcal{O}\left(\Delta t^k\right)$$

as $\Delta t \to 0$. To compute $L_{\Delta t}(x)$ one uses a Taylor expansion of the exact solution $u(x,t)$ near x. We know that u may have discontinuities, so it does not necessarily have a Taylor expansion. Therefore, the concept of truncation error is formal. However, if $u(x,t)$ is smooth near (x,t), then one would expect that a higher-order method would approximate u better than a lower-order method near (x,t).

◇ Example 3.3 (Local truncation error)
Consider the upwind method. Then

$$S_N(\Delta t)u(x) = u(x) - \frac{\Delta t}{\Delta x}\left(f(u(x)) - f(u(x-\Delta x))\right).$$

We verify that the upwind method is of first order:

$$\begin{aligned}
L_{\Delta t}(x) &= \frac{1}{\Delta t}\Big(u(x,t+\Delta t) - u(x,t) + \frac{\Delta t}{\Delta x}(f(u(x,t)) - f(u(x-\Delta x,t)))\Big) \\
&= \frac{1}{\Delta t}\Big(u + \Delta t\, u_t + \frac{(\Delta t)^2}{2}u_{tt} + \cdots - u \\
&\qquad + \frac{\Delta t}{\Delta x}\big(f(u) - f(u) - (-\Delta x) f(u)_x - \frac{1}{2}(-\Delta x)^2 f(u)_{xx} + \cdots\big)\Big) \\
&= u_t + f(u)_x + \frac{1}{\Delta t}\Big(\frac{(\Delta t)^2}{2}u_{tt} - \frac{\Delta t\,\Delta x}{2}f(u)_{xx} + \cdots\Big) \\
&= \frac{\Delta x}{2}\left(\lambda\, u_{tt} - f(u)_{xx}\right) + \mathcal{O}\left((\Delta t)^2\right).
\end{aligned}$$

Since u is a smooth solution of (3.1), we find that

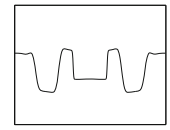
$$u_{tt} = \left((f'(u))^2 u_x\right)_x,$$

and inserting this into the previous equation, we obtain

$$L_{\Delta t} = \frac{\Delta t}{2\lambda}\frac{\partial}{\partial x}\left(f'(u)\left(\lambda f'(u) - 1\right)u_x\right) + \mathcal{O}\left((\Delta t)^2\right). \tag{3.22}$$

Hence, the upwind method is of first order. This means that Godunov's scheme is also of first order. Similarly, computations based on the Lax–Friedrichs scheme yield

$$L_{\Delta t} = \frac{\Delta t}{2\lambda^2}\frac{\partial}{\partial x}\left(\left(\left(\lambda f'(u)\right)^2 - 1\right)u_x\right) + \mathcal{O}\left(\Delta t^2\right). \tag{3.23}$$



Consequently, the Lax–Friedrichs scheme is indeed of first order. From the above computations it also emerges that the Lax–Friedrichs scheme is *second-order* accurate when applied to the equation (see Exercise 3.6)

$$u_t + f(u)_x = \frac{\Delta t}{2\lambda^2}\left(\left(1 - (\lambda f'(u))^2\right) u_x\right)_x . \tag{3.24}$$

This is called the *model equation* for the Lax–Friedrichs scheme. In order for this to be well posed, the coefficient of u_{xx} on the right-hand side must be nonnegative, that is,

$$|\lambda f'(u)| \le 1. \tag{3.25}$$

This is a stability restriction on λ, and it is the Courant–Friedrichs–Lewy (CFL) condition that we encountered in (3.9); see also (1.50).

The model equation for the upwind method is

$$u_t + f(u)_x = \frac{\Delta t}{2\lambda}\left(f'(u)\left(1 - \lambda f'(u)\right) u_x\right)_x . \tag{3.26}$$

In order for this equation to be well posed, we must have $f'(u) \ge 0$ and $\lambda f'(u) \le 1$. ◇

From the above examples, we see that first-order methods have model equations with a diffusive term. Similarly, one finds that second-order methods have model equations with a dispersive right-hand side. Therefore, the oscillations observed in the computations were to be expected.

From now on we let the function $u_{\Delta t}$ be defined by

$$u_{\Delta t}(x,t) = u_j^n, \quad \text{for } (x,t) \in I_j^n. \tag{3.27}$$

Observe that

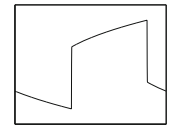
$$\int_{\mathbb{R}} u_{\Delta t}(x,t)\,dx = \Delta x \sum_j u_j^n, \qquad \text{for } t_n \le t < t_{n+1}.$$

We briefly mentioned in Example 3.2 the fact that if $u_{\Delta t}$ converges, then the limit is a weak solution. Precisely, we have the well-known Lax–Wendroff theorem.

Theorem 3.4 (Lax–Wendroff theorem) *Let $u_{\Delta t}$ be computed from a conservative and consistent method. Assume that* $\mathrm{T.V.}_x(u_{\Delta t})$ *is uniformly bounded in Δt. Consider a subsequence $u_{\Delta t_k}$ such that $\Delta t_k \to 0$, and assume that $u_{\Delta t_k}$ converges in L^1_{loc} as $\Delta t_k \to 0$. Then the limit is a weak solution to (3.1).*

Proof The proof uses summation by parts. Let $\varphi(x,t)$ be a test function. For simplicity we write $\varphi_j^n = \varphi(x_j, t_n)$. By the definition of u_j^{n+1},

$$\sum_{n=0}^{N}\sum_{j=-\infty}^{\infty} \varphi_j^n\left(u_j^{n+1} - u_j^n\right) = -\frac{\Delta t}{\Delta x}\sum_{n=0}^{N}\sum_{j=-\infty}^{\infty} \varphi_j^n\left(F_{j+1/2}^n - F_{j-1/2}^n\right),$$



where we choose $T = N\Delta t$ such that $\varphi = 0$ for $t \geq T$. After a summation by parts we get

$$-\sum_{j=-\infty}^{\infty} \varphi_j^0 u_j^0 - \sum_{j=-\infty}^{\infty} \sum_{n=1}^{N} \left(\varphi_j^n - \varphi_j^{n-1}\right) u_j^n - \frac{\Delta t}{\Delta x} \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left(\varphi_{j-1}^n - \varphi_j^n\right) F_{j+1/2}^n = 0.$$

Rearranging, we find that

$$\Delta t \Delta x \sum_{n=1}^{N} \sum_{j=-\infty}^{\infty} \left(\frac{\varphi_j^n - \varphi_j^{n-1}}{\Delta t}\right) u_j^n + \Delta t \Delta x \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left(\frac{\varphi_{j-1}^n - \varphi_j^n}{\Delta x}\right) F_{j+1/2}^n = -\Delta x \sum_{j=-\infty}^{\infty} \varphi\left(x_j, 0\right) u_j^0. \tag{3.28}$$

This almost looks like a Riemann sum for the weak formulation of (3.1). Thus

$$\Delta x \sum_{j=-\infty}^{\infty} \varphi\left(x_j, 0\right) u_j^0 \to \int_0^{\infty} \varphi(x,0) u_0(x)\, dx$$

as $\Delta x \to 0$, and

$$\Delta t \Delta x \sum_{n=1}^{N} \sum_{j=-\infty}^{\infty} \left(\frac{\varphi_j^n - \varphi_j^{n-1}}{\Delta t}\right) u_j^n \to \int_0^T \int_{-\infty}^{\infty} \varphi_t(x,t) u(x,t)\, dx\, dt$$

as $\Delta x, \Delta t \to 0$.

Since

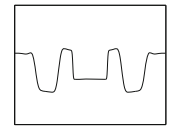
$$\Delta t \Delta x \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left(\frac{\varphi_{j-1}^n - \varphi_j^n}{\Delta x}\right) f(u_j^n) \to \int_0^T \int_{-\infty}^{\infty} \varphi_x(x,t) f(u(x,t))\, dx\, dt \tag{3.29}$$

as $\Delta x, \Delta t \to 0$, it remains to show that

$$\Delta t \Delta x \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left| F_{j+1/2}^n - f(u_j^n) \right| \tag{3.30}$$

tends to zero as $\Delta t \to 0$ in order to conclude that the limit is a weak solution. Using consistency, (3.12), we find that (3.30) equals

$$\Delta t \Delta x \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left| F\left(u_{j-p}^n, \ldots, u_{j+q}^n\right) - F\left(u_j^n, \ldots, u_j^n\right) \right|,$$



which by the Lipschitz continuity of F is less than

$$\begin{aligned}
&\Delta t \Delta x L \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \sum_{k=-p}^{q} \left| u_{j+k}^n - u_j^n \right| \\
&\quad \leq \frac{1}{2}(q(q+1) + p(p+1)) \Delta t \Delta x L \sum_{n=0}^{N} \sum_{j=-\infty}^{\infty} \left| u_{j+1}^n - u_j^n \right| \\
&\quad \leq (q^2 + p^2) \Delta x \, L \, \text{T.V.}\,(u_{\Delta t}) \; T,
\end{aligned}$$

where L is the Lipschitz constant of F. Using the uniform boundedness of the total variation of $u_{\Delta x}$, we infer that (3.30) is small for small Δx, and the limit is a weak solution. □

We proved in Theorem 2.15 that the solution of a scalar conservation law in one dimension possesses several properties. The corresponding properties for conservative and consistent numerical schemes read as follows:

Definition 3.5 Let $u_{\Delta t}$ be computed from a conservative and consistent method.

(i) A method is said to be *total variation bounded (TVB)*, or *total variation stable*,[1] if the total variation of u^n is uniformly bounded, independently of Δx and Δt.

(ii) Assume that u_0 has finite total variation. We say that a numerical method is *total variation diminishing (TVD)* if $\text{T.V.}\left(u^{n+1}\right) \leq \text{T.V.}\,(u^n)$ for all $n \in \mathbb{N}_0$.

(iii) A method is called *monotonicity preserving* if the initial data being monotone implies that u^n is monotone for all $n \in \mathbb{N}$.

(iv) Assume that $u_0 \in L^1(\mathbb{R})$. Let $v_{\Delta t}$ be another solution with initial data $v_0 \in L^1(\mathbb{R})$. A numerical method is called L^1*-contractive* if

$$\|u_{\Delta t}(t) - v_{\Delta t}(t)\|_{L^1} \leq \|u_{\Delta t}(0) - v_{\Delta t}(0)\|_{L^1}$$

for all $t \geq 0$. Alternatively, we can of course write this as

$$\sum_j \left| u_j^{n+1} - v_j^{n+1} \right| \leq \sum_j \left| u_j^n - v_j^n \right|, \quad n \in \mathbb{N}_0.$$

(v) A method is said to be *monotone* if for initial data u^0 and v^0, we have

$$u_j^0 \leq v_j^0, \quad j \in \mathbb{Z} \quad \Rightarrow \quad v_j^n \leq v_j^n, \quad j \in \mathbb{Z}, \, n \in \mathbb{N}.$$

The above notions are strongly interrelated, as the next theorem shows.

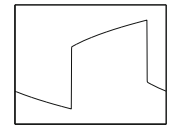
Theorem 3.6 *For conservative and consistent methods the following hold*:

(i) *Assume initial data to be integrable. In that case, every monotone method is* L^1*-contractive.*

(ii) *Every* L^1*-contractive method is TVD.*

(iii) *Every TVD method is monotonicity preserving.*

[1] This definition is slightly different from the standard definition of T.V. stable methods.



Proof (i) We apply the Crandall–Tartar lemma, Lemma 2.13, with $\Omega = \mathbb{R}$, and D equal to the set of all functions in L^1 that are piecewise constant on the grid I_j, $j \in \mathbb{Z}$, and we define $T(u^0) = u^n$. Since the method is conservative (cf. (3.11)), we have that

$$\sum_j u_j^n = \sum_j u_j^0, \quad \text{or} \quad \int T(u^0)\,dx = \int u^n\,dx = \int u^0\,dx.$$

Lemma 2.13 immediately implies that (for $t \in [t_n, t_{n+1})$)

$$\begin{aligned}\|u_{\Delta t}(t) - v_{\Delta t}(t)\|_{L^1} &= \Delta x \sum_j \left|v_j^n - v_j^n\right| \le \Delta x \sum_j \left|u_j^0 - v_j^0\right| \\ &= \|u_{\Delta t}(0) - v_{\Delta t}(0)\|_{L^1}.\end{aligned}$$

(ii) Assume now that the method is L^1-contractive, i.e.,

$$\sum_j \left|u_j^{n+1} - v_j^{n+1}\right| \le \sum_j \left|u_j^n - v_j^n\right|.$$

Let v^n be the numerical solution with initial data

$$v_j^0 = u_{j+1}^0.$$

Then by the translation invariance induced by (3.6), we have $v_i^n = u_{i+1}^n$ for all n. Furthermore,

$$\begin{aligned}\text{T.V.}\left(u_j^{n+1}\right) &= \sum_j \left|u_{j+1}^{n+1} - u_j^{n+1}\right| = \sum_j \left|u_j^{n+1} - v_j^{n+1}\right| \\ &\le \sum_j \left|u_j^n - v_j^n\right| = \text{T.V.}\left(u_j^n\right).\end{aligned}$$

(iii) Consider now a TVD method, and assume that we have monotone initial data. Since T.V. $\left(u^0\right)$ is finite by assumption, the limits

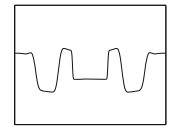
$$u_L = \lim_{j\to-\infty} u_j^0 \text{ and } u_R = \lim_{j\to\infty} u_j^0$$

exist. Then T.V. $\left(u^0\right) = |u_R - u_L|$. If u^1 were not monotone, then T.V. $\left(u^1\right) > |u_R - u_L| =$ T.V. $\left(u^0\right)$, which is a contradiction. □

We can summarize the above theorem as follows:

$$\text{monotone} \Rightarrow L^1\text{-contractive} \Rightarrow \text{TVD} \Rightarrow \text{monotonicity preserving}.$$

Monotonicity is relatively easy to check for explicit methods, e.g., by calculating the partial derivatives $\partial G/\partial u^i$ in (3.3).



◇ Example 3.7 (Lax–Friedrichs scheme)
Recall from Example 3.2 that the Lax–Friedrichs scheme is given by

$$u_j^{n+1} = \frac{1}{2}\left(u_{j+1}^n + u_{j-1}^n\right) - \frac{1}{2}\lambda\left(f\left(u_{j+1}^n\right) - f\left(u_{j-1}^n\right)\right).$$

Computing partial derivatives, we obtain, assuming the flux function f to be continuously differentiable,

$$\frac{\partial u_j^{n+1}}{\partial u_k^n} = \begin{cases} (1 - \lambda f'(u_k^n))/2 & \text{for } k = j+1, \\ (1 + \lambda f'(u_k^n))/2 & \text{for } k = j-1, \\ 0 & \text{otherwise,} \end{cases}$$

and hence we see that the Lax–Friedrichs scheme is monotone as long as the CFL condition

$$\lambda \left| f'(u) \right| \le 1$$

is fulfilled. See also Exercise 3.7. ◇

Theorem 3.8 *Fix $T > 0$. Assume that f is Lipschitz continuous. Let $u_0 \in L^1(\mathbb{R})$ have bounded variation. Assume that $u_{\Delta t}$ is computed with a method that is conservative, consistent, total variation bounded, and uniformly bounded, that is,*

$$\text{T.V.}\,(u_{\Delta t}) \le M \text{ and } \|u_{\Delta t}\|_\infty \le M,$$

where M is independent of Δx and Δt.

Then $\{u_{\Delta t}(t)\}$ has a subsequence that converges for all $t \in [0,T]$ to a weak solution $u(t)$ in $L^1_{\text{loc}}(\mathbb{R})$. Furthermore, the limit is in $C\left([0,T]; L^1_{\text{loc}}(\mathbb{R})\right)$.

Proof We intend to apply Theorem A.11. It remains to show that

$$\int_a^b |u_{\Delta t}(x,t) - u_{\Delta t}(x,s)|\, dx \le C\,|t-s| + \nu(\Delta t), \text{ as } \Delta t \to 0, \quad s,t \in [0,T],$$

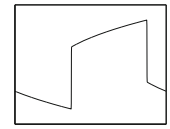
for some nonnegative continuous function ν with $\nu(0) = 0$.

The Lipschitz continuity of the flux function implies, for fixed Δt,

$$\begin{aligned} \left|u_j^{n+1} - u_j^n\right| &= \lambda \left|F_{j+1/2}^n - F_{j-1/2}^n\right| \\ &= \lambda \left|F(u_{j-p}^n, \ldots, u_{j+q}^n) - F(u_{j-p-1}^n, \ldots, u_{j+q-1}^n)\right| \\ &\le \lambda L \left(\left|u_{j-p}^n - u_{j-p-1}^n\right| + \cdots + \left|u_{j+q}^n - u_{j+q-1}^n\right|\right), \end{aligned}$$

from which we conclude that

$$\begin{aligned} \|u_{\Delta t}(\cdot, t_{n+1}) - u_{\Delta t}(\cdot, t_n)\|_{L^1} &= \sum_j \left|u_j^{n+1} - u_j^n\right| \Delta x \\ &\le L(p+q+1)\text{T.V.}\,(u^n)\,\Delta t \\ &\le L(p+q+1) M\,\Delta t, \end{aligned}$$



where L is the Lipschitz constant of F. More generally,

$$\begin{aligned}\|u_{\Delta t}(\cdot, t_m) - u_{\Delta t}(\cdot, t_n)\|_{L^1} &\le L(p+q+1)M\,|n-m|\,\Delta t \\ &= L(p+q+1)M\,|t_n - t_m|\,.\end{aligned}$$

Now let $\tau_1, \tau_2 \in [0,T]$, and choose $\tilde{t}_1, \tilde{t}_2 \in \{n\Delta t \mid 0 \le n \le T/\Delta t\}$ such that

$$0 \le \tau_j - \tilde{t}_j < \Delta t \text{ for } j = 1,2.$$

By construction $u_{\Delta t}(\tau_j) = u_{\Delta t}(\tilde{t}_j)$, and hence

$$\begin{aligned}&\|u_{\Delta t}(\cdot, \tau_{L^1}) - u_{\Delta t}(\cdot, \tau_2)\|_{L^1} \\ &\quad\le \left\|u_{\Delta t}(\cdot, \tau_1) - u_{\Delta t}(\cdot, \tilde{t}_1)\right\|_{L^1} + \left\|u_{\Delta t}(\cdot, \tilde{t}_1) - u_{\Delta t}(\cdot, \tilde{t}_2)\right\|_{L^1} \\ &\qquad + \left\|u_{\Delta t}(\cdot, \tilde{t}_2) - u_{\Delta t}(\cdot, \tau_2)\right\|_{L^1} \\ &\quad\le (p+q+1)L\,M\left|\tilde{t}_1 - \tilde{t}_2\right| \\ &\quad\le (p+q+1)L\,M\,|\tau_1 - \tau_2| + \mathcal{O}(\Delta t)\,.\end{aligned}$$

Observe that this estimate is uniform in $\tau_1, \tau_2 \in [0,T]$. We conclude that

$$u_{\Delta t} \to u \text{ in } C([0,T]; L^1([a,b]))$$

for a sequence $\Delta t \to 0$. The Lax–Wendroff theorem then says that this limit is a weak solution. □

At this point, the reader should review the concept of a Kružkov entropy condition; see Sect. 2.1. A function u is a Kružkov entropy solution of

$$u_t + f(u)_x = 0$$

if it satisfies

$$\eta(u)_t + q(u)_x \le 0 \tag{3.31}$$

in the sense of distributions, where

$$\eta(u) = |u-k|\,, \quad q(u) = \operatorname{sign}(u-k)\,(f(u) - f(k)),$$

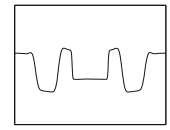
for all $k \in \mathbb{R}$.

The analogue of the Kružkov entropy pair for difference schemes reads as follows. We still employ $\eta(u) = |u-k|$. Write

$$a \vee b = \max\{a,b\} \quad \text{and} \quad a \wedge b = \min\{a,b\},$$

and observe the trivial identity

$$|a-b| = a \vee b - a \wedge b.$$



Then we define the *numerical entropy flux* Q by

$$Q_{j+1/2}(u) = F_{j+1/2}(u \vee k) - F_{j+1/2}(u \wedge k), \tag{3.32}$$

or more explicitly,

$$Q\big(u_{j-p}, \ldots, u_{j+q}\big) = F\big(u_{j-p} \vee k, \ldots, u_{j+q} \vee k\big) - F\big(u_{j-p} \wedge k, \ldots, u_{j+q} \wedge k\big).$$

Note that Q is consistent with the Kružkov entropy flux, i.e.,

$$Q\,(c, \ldots, c) = \operatorname{sign}(c-k)\,(f(c) - f(k)).$$

Returning to monotone difference schemes, we have the following result.

Theorem 3.9 *Fix $T > 0$. Assume that f is Lipschitz continuous. Let $u_0 \in L^1(\mathbb{R})$ have bounded variation. Assume that $u_{\Delta t}$ is computed with a method that is conservative, consistent, and monotone.*

For every sequence $\Delta t_k \to 0$, the family $\{u_{\Delta t_k}(t)\}$ converges in $L^1_{\mathrm{loc}}(\mathbb{R})$ to the Kružkov entropy solution $u(t)$ for all $t \in [0,T]$. Furthermore, the limit is in $C\left([0,T]; L^1_{\mathrm{loc}}(\mathbb{R})\right)$.

Proof Consider a sequence $\Delta t_k \to 0$. Theorem 3.8 allows us to conclude that $u_{\Delta t_k}$ has a subsequence that converges in $C([0,T]; L^1([a,b]))$ to a weak solution. It remains to show that the limit satisfies a discrete Kružkov form. First we find, using (3.7) and (3.32), that

$$G(u^n \vee k) - G(u^n \wedge k) = |u^n - k| - \lambda\big(Q^n_{j+1/2} - Q^n_{j-1/2}\big).$$

Using that $u^{n+1}_j = G_j(u^n)$, cf. (3.3), and the consistency of the scheme, see (3.12), which implies $k = G(k, \ldots, k) = G(k)$, we conclude from the monotonicity of the scheme that

$$\begin{aligned} G_j(u^n \vee k) &\geq G_j(u^n) \vee G(k) = G_j(u^n) \vee k, \\ -G_j(u^n \wedge k) &\geq -\big(G_j(u^n) \wedge G(k)\big) = -\big(G_j(u^n) \wedge k\big). \end{aligned}$$

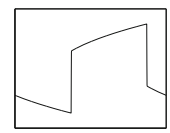
Therefore,

$$\left|u^{n+1}_j - k\right| - \left|u^n_j - k\right| + \lambda\big(Q^n_{j+1/2} - Q^n_{j-1/2}\big) \leq 0. \tag{3.33}$$

Applying the technique used in proving the Lax–Wendroff theorem to (3.33) shows that the limit u satisfies

$$\begin{aligned} &\iint \big(|u-k|\,\varphi_t + \operatorname{sign}(u-k)\,\big(f(u) - f(k)\big)\varphi_x\big)\,dx\,dt \\ &\qquad + \int_{\mathbb{R}} |u_0 - k|\,\varphi(x,0)\,dx - \int_{\mathbb{R}} \left(|u-k|\,\varphi\right)|_{t=T}\,dx \geq 0, \end{aligned}$$

for every nonnegative test function $\varphi \in C_0^\infty(\mathbb{R} \times [0,T])$ and for every $k \in \mathbb{R}$.



Suppose there is another subsequence for which $u_{\Delta t}$ does not converge to the entropy solution. Then by the above argument, this subsequence has another subsequence for which the limit is the unique entropy solution. The uniqueness of the limit gives a contradiction, and we conclude that for all sequences $\Delta t_k \to 0$, the sequence $\{u_{\Delta t_k}(t)\}$ converges to the unique entropy solution $u(t)$. □

Note that the above theorem offers a constructive proof of the existence of weak entropy solutions to scalar conservation laws. The fact that monotone schemes converge to the entropy solution provides an alternative to the front-tracking method discussed in Chapt. 2.

Now we shall examine the local truncation error of a general conservative, consistent, and monotone method. Since this can be written

$$\begin{aligned} u_j^{n+1} &= G_j\left(u^n\right) = G\left(u_{j-p-1}^n, \ldots, u_{j+q}^n\right) \\ &= u_j^n - \lambda\left(F\left(u_{j-p}^n, \ldots, u_{j+q}^n\right) - F\left(u_{j-p-1}^n, \ldots, u_{j+q-1}^n\right)\right), \end{aligned}$$

we write

$$G = G(\alpha_0, \ldots, \alpha_{p+q+1}) \quad \text{and} \quad F = F(\alpha_1, \ldots, \alpha_{p+q+1}).$$

We assume that F, and hence G, is three times continuously differentiable with respect to all arguments, and write the derivatives with respect to the ith argument as

$$\partial_i G(\alpha_0, \ldots, \alpha_{p+q+1}) \quad \text{and} \quad \partial_i F(\alpha_1, \ldots, \alpha_{p+q+1}).$$

We set $\partial_i F = 0$ if $i = 0$. Throughout this calculation, we assume that the jth slot of G contains u_j^n, so that $G(\alpha_0, \ldots, \alpha_{p+q+1}) = u_j - \lambda(\cdots)$. By consistency we have that

$$G(u, \ldots, u) = u \quad \text{and} \quad F(u, \ldots, u) = f(u).$$

Using this, we find that

$$\sum_{i=1}^{p+q+1} \partial_i F(u, \ldots, u) = f'(u), \tag{3.34}$$

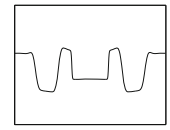
$$\partial_i G = \delta_{i,j} - \lambda\left(\partial_{i-1} F - \partial_i F\right), \tag{3.35}$$

and

$$\partial_{i,k}^2 G = -\lambda\left(\partial_{i-1,k-1}^2 F - \partial_{i,k}^2 F\right). \tag{3.36}$$

Therefore,

$$\sum_{i=0}^{p+q+1} \partial_i G(u, \ldots, u) = \sum_{i=0}^{p+q+1} \delta_{i,j} = 1. \tag{3.37}$$



Furthermore,

$$\begin{aligned}\sum_{i=0}^{p+q+1}(i-j)\partial_i G(u,\dots,u) &= \sum_{i=0}^{p+q+1}\big[(i-j)\delta_{i,j}\\ &\quad -\lambda(i-j)\left(\partial_{i-1}F(u,\dots,u)-\partial_i F(u,\dots,u)\right)\big]\\ &= -\lambda\sum_{i=0}^{p+q+1}((i+1)-i))\partial_i F(u,\dots,u)\\ &= -\lambda f'(u). \end{aligned} \tag{3.38}$$

We also find that

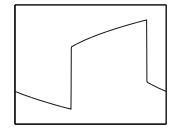
$$\begin{aligned}&\sum_{i,k=0}^{p+q+1}(i-k)^2\partial_{i,k}^2 G(u,\dots,u)\\ &\qquad = -\lambda\sum_{i,k=0}^{p+q+1}(i-k)^2\left(\partial_{i-1,k-1}^2 F(u,\dots,u)-\partial_{i,k}^2 F(u,\dots,u)\right)\\ &\qquad = -\lambda\sum_{i,k=0}^{p+q+1}\left(((i+1)-(k+1))^2-(i-k)^2\right)\partial_{i,k}^2 F(u,\dots,u)\\ &\qquad = 0. \end{aligned} \tag{3.39}$$

Having established this, we now let $u = u(x,t)$ be a smooth solution of the conservation law (3.1). We are interested in applying G to $u(x,t)$, i.e., in calculating

$$G(u(x-(p+1)\Delta x,t)\dots,u(x,t),\dots,u(x+q\Delta x,t)).$$

Set $u_i = u(x+(i-(p+1))\Delta x,t)$ for $i = 0,\dots,p+q+1$. Then we find that

$$\begin{aligned}&G(u_0,\dots,u_{p+q+1})\\ &\quad = G(u_j,\dots,u_j)+\sum_{i=0}^{p+q+1}\partial_i G(u_j,\dots,u_j)\left(u_i-u_j\right)\\ &\qquad +\frac{1}{2}\sum_{i,k=0}^{p+q+1}\partial_{i,k}^2 G(u_j,\dots,u_j)\left(u_i-u_j\right)\left(u_k-u_j\right)+\mathcal{O}\left(\Delta x^3\right)\\ &\quad = u(x,t)+u_x(x,t)\Delta x\sum_{i=0}^{p+q+1}(i-j)\partial_i G(u_j,\dots,u_j)\\ &\qquad +\frac{1}{2}u_{xx}(x,t)\Delta x^2\sum_{i=0}^{p+q+1}(i-j)^2\partial_i G(u_j,\dots,u_j)\\ &\qquad +\frac{1}{2}u_x^2(x,t)\Delta x^2\sum_{i,k=0}^{p+q+1}(i-j)(k-j)\partial_{i,k}^2 G(u_j,\dots,u_j)+\mathcal{O}\left(\Delta x^3\right)\end{aligned}$$



$$
\begin{aligned}
&= u(x,t) + u_x(x,t)\Delta x \sum_{i=0}^{p+q+1} (i-j)\partial_i G(u_j,\dots,u_j) \\
&+ \frac{1}{2}\Delta x^2 \sum_{i=0}^{p+q+1} (i-j)^2 \left[\partial_i G(u_j,\dots,u_j)u_x(x,t)\right]_x \\
&- \frac{1}{2}\Delta x^2 u_x^2(x,t) \sum_{i,k}^{p+q+1} \left((i-j)^2 - (i-j)(k-j)\right)\partial_{i,k}^2 G(u_j,\dots,u_j) \\
&+ \mathcal{O}\left(\Delta x^3\right).
\end{aligned}
$$

Next we observe, since $\partial_{i,k}^2 G = \partial_{k,i}^2 G$ and using (3.39), that

$$
\begin{aligned}
0 &= \sum_{i,k}(i-k)^2\partial_{i,k}^2 G = \sum_{i,k}((i-j)-(k-j))^2\partial_{i,k}^2 G \\
&= \sum_{i,k}((i-j)^2 - 2(i-j)(k-j))\partial_{i,k}^2 G + \sum_{i,k}(k-j)^2\partial_{k,i}^2 G \\
&= 2\sum_{i,k}((i-j)^2 - (i-j)(k-j))\partial_{i,k}^2 G.
\end{aligned}
$$

Consequently, the penultimate term in the Taylor expansion of G above is zero, and we have that

$$
\begin{aligned}
G(u(x-(p+1)\Delta x,t),\dots,u(x+q\Delta x,t)) &= u(x,t) - \Delta t\, f(u(x,t))_x \\
&+ \frac{\Delta x^2}{2}\sum_i (i-j)^2 \left[\partial_i G(u(x,t),\dots,u(x,t))u_x\right]_x + \mathcal{O}\left(\Delta x^3\right). \qquad (3.40)
\end{aligned}
$$

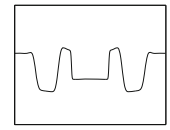
Since u is a smooth solution of (3.1), we have already established that

$$
u(x,t+\Delta t) = u(x,t) - \Delta t\, f(u)_x + \frac{\Delta t^2}{2}\left[\left(f'(u)\right)^2 u_x\right]_x + \mathcal{O}\left(\Delta t^3\right).
$$

Hence, we compute the local truncation error as

$$
\begin{aligned}
L_{\Delta t} &= -\frac{\Delta t}{2\lambda^2}\left[\left(\sum_{i=1}^{p+q+1}(i-j)^2\partial_i G(u,\dots,u) - \lambda^2(f'(u))^2\right)u_x\right]_x \\
&=: -\frac{\Delta t}{2\lambda^2}\left[\beta(u)u_x\right]_x + \mathcal{O}\left(\Delta t^2\right). \qquad (3.41)
\end{aligned}
$$

Thus if $\beta > 0$, then the method is of first order. What we have done so far is valid for every conservative and consistent method where the numerical flux function is three times continuously differentiable. Next, we use that $\partial_i G \geq 0$, so that $\sqrt{\partial_i G}$



is well defined. This means that

$$|-\lambda f'(u)| = \left| \sum_{i=0}^{p+q+1} (i-j)\partial_i G(u,\dots,u) \right| = \sum_{i=0}^{p+q+1} |i-j| \sqrt{\partial_i G(u,\dots,u)} \sqrt{\partial_i G(u,\dots,u)}.$$

Using the Cauchy–Schwarz inequality and (3.37), we find that

$$\lambda^2 \left(f'(u)\right)^2 \le \sum_{i=0}^{p+q+1} (i-j)^2 \partial_i G(u,\dots,u) \sum_{i=0}^{p+q+1} \partial_i G(u,\dots,u) = \sum_{i=0}^{p+q+1} (i-j)^2 \partial_i G(u,\dots,u).$$

Thus, $\beta(u) \ge 0$. Furthermore, the inequality is strict if more than one term in the sum on the right-hand side is different from zero. If $\partial_i G(u,\dots,u) = 0$ except for $i = k$ for some k, then $G(u_0,\dots,u_{p+q+1}) = u_k$ by (3.37). Hence the scheme is a linear translation, and by consistency, $f(u) = cu$, where $c = (j-k)\lambda$. Therefore, monotone methods for nonlinear conservation laws are at most first-order accurate. This is indeed their main drawback. To recapitulate, we have proved the following theorem:

Theorem 3.10 *Assume that the numerical flux F is three times continuously differentiable, and that the corresponding scheme is monotone. Then the method is at most first-order accurate.*

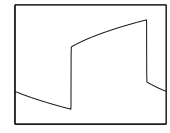
3.2 Higher-Order Schemes

We want to derive a second-order difference approximation to the solution of a conservation law

$$u_t + f(u)_x = 0.$$

In order to derive scheme that is second-order accurate, the local truncation error must be third-order accurate. For a smooth solution we have

$$\begin{aligned} u(x,t+\Delta t) &= u(x,t) + \Delta t u_t(x,t) + \frac{\Delta t^2}{2} u_{tt}(x,t) + \mathcal{O}\left(\Delta t^3\right) \\ &= u(x,t) - \Delta t f(u(x,t))_x - \frac{\Delta t^2}{2} f(u(x,t))_{xt} + \mathcal{O}\left(\Delta t^3\right) \\ &= u - \Delta t f(u)_x + \frac{\Delta t^2}{2} \left(f'(u) f(u)_x\right)_x + \mathcal{O}\left(\Delta t^3\right). \end{aligned}$$



For a difference scheme we have $\Delta x = \mathcal{O}(\Delta t)$, so if the resulting scheme is of second order, the difference approximation to $f(u)_x$ must be second-order accurate, and the approximation to $(f' f_x)_x$ can be first-order accurate. We can use the following (where we write $D_0(g(x)) = (g(x+\Delta x) - g(x-\Delta x))/(2\Delta x)$) relations:

$$\begin{aligned} f(u(x,t))_x &= D_0 f(u(x,t)) + \mathcal{O}\left(\Delta x^2\right) \\ &= \frac{f(u(x+\Delta x,t)) - f(u(x-\Delta x,t))}{2\Delta x} + \mathcal{O}\left(\Delta x^2\right), \\ (f'(u(x,t))f(u(x,t))_x)_x &= \frac{1}{\Delta x}\left(f'\left(u\left(x+\frac{\Delta x}{2},t\right)\right)\frac{f(u(x+\Delta x,t))-f(u(x,t))}{\Delta x}\right. \\ &\quad \left. - f'\left(u\left(x-\frac{\Delta x}{2},t\right)\right)\frac{f(u(x,t))-f(u(x-\Delta x,t))}{\Delta x}\right) \\ &\quad + \mathcal{O}\left(\Delta x^2\right), \\ f'\left(u\left(x \pm \frac{\Delta x}{2},t\right)\right) &= \frac{f(u(x\pm\Delta x,t)) - f(u(x,t))}{u(x\pm\Delta x,t) - u(x,t)} + \mathcal{O}\left(\Delta u^2\right). \end{aligned}$$

This leads to the scheme

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}\left(f_{j+1}^n - f_{j-1}^n\right) + \frac{\lambda^2}{2}\left(v_{j+1/2}^2 \Delta_+ u_j^n - v_{j-1/2}^2 \Delta_- u_j^n\right), \tag{3.42}$$

where

$$\lambda = \frac{\Delta t}{\Delta x}, \quad f_j^n = f(u_j^n), \quad \Delta_\pm v_j = \pm(v_{j\pm1} - v_j), \quad v_{j+1/2} = \frac{\Delta_+ f_j^n}{\Delta_+ u_j^n}.$$

The scheme (3.42) is called the Lax–Wendroff scheme, and by construction it is of second order. We can see that it is conservative with a two-point numerical flux function given by $F_{j+1/2} = F(u_j, u_{j+1})$, where

$$F(u,v) = \frac{1}{2}\left(f(v) + f(u) - \lambda v^2(u,v)(v-u)\right), \quad v(u,v) = \frac{f(v)-f(u)}{v-u}.$$

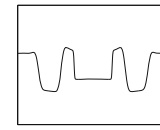
◇ Example 3.11
We test this second-order scheme on the equation

$$u_t + u_x = 0$$

with two sets of periodic initial data

$$u^1(x,0) = \sin^2(\pi x), \quad u^2(x,0) = \begin{cases} 1 & x \in [0.3,\ 0.7], \\ 0 & x \in [0,1] \setminus [0.3,\ 0.7], \end{cases}$$

and u^2 extended periodically. By periodicity, we know that $u^i(x,k) = u^i(x,0)$ for $k \in \mathbb{N}$. In Fig. 3.4 we have plotted the numerical solution at $t = 3$ with initial data u^1 and u^2 and $\Delta x = 1/30$. Note that for the smooth solution the method gives very



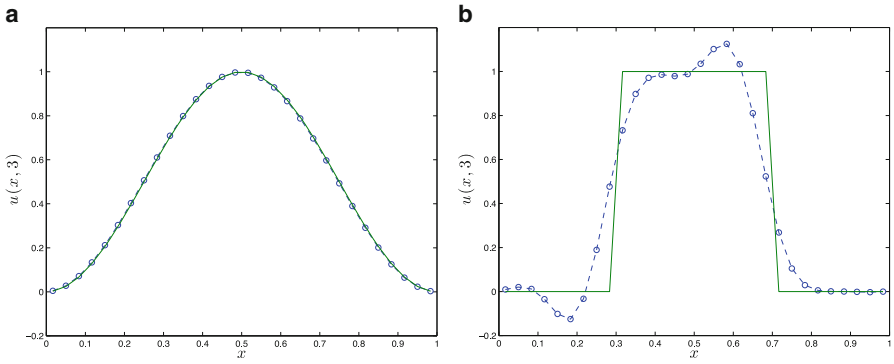


Fig. 3.4 **a** The numerical solution with initial values u^1. **b** The numerical solution with initial value u^2. We use $\Delta x = 1/30$

accurate results, and the errors are indeed of second order. For the discontinuous solution, the errors seem large, and we also see the prominent oscillations trailing the discontinuity. ◇

For simplicity we will for the moment assume that $f' \geq 0$, so that the upwind method is monotone (and hence TVD). If f is not monotone, then the upwind flux below should be replaced by a numerical flux giving a monotone method.

The Lax–Wendroff numerical flux function can be rearranged to read

$$\begin{aligned} F^n_{j+1/2} &= f(u^n_j) - \frac{1}{2} v_{j+1/2} \left(\lambda v_{j+1/2} - 1\right) \Delta_+ u^n_j \\ &= \text{upwind} + \text{second-order correction.} \end{aligned}$$

We would like to modify the Lax–Wendroff method so that it is locally of second order where the solution is smooth, and first order and monotone near discontinuities. Hence, we would like to turn off the second-order correction near discontinuities. One way of doing this is to observe that the oscillations occur near discontinuities (this is the Gibbs phenomenon), and use oscillations as an indicator of when the second-order term should be turned off. As an important side effect, this is likely to make the resulting method TVD.

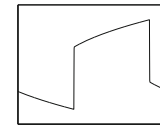
To this end let r_j (whose exact form will be specified later) be some "indicator of oscillations" near x_j. We assume that if there are oscillations, then $r_j < 0$. Let $\varphi(r)$ be a continuous function that is zero if $r < 0$.

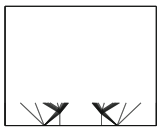
Now we modify the numerical flux for the Lax–Wendroff method to read

$$F^n_{j+1/2} = f^n_j - \frac{1}{2} \varphi(r_j) v_{j+1/2} \left(\lambda v_{j+1/2} - 1\right) \Delta_+ u^n_j. \tag{3.43}$$

If we set

$$\alpha_{j+1/2} = \frac{1}{2} v_{j+1/2} (1 - \lambda v_{j+1/2}), \tag{3.44}$$





the modified scheme reads

$$\begin{aligned} u_j^{n+1} &= u_j^n - \lambda\Delta_- f_j^n - \lambda\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right) \\ &= u_j^n - \lambda v_{j-1/2}\Delta_- u_j^n - \lambda\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right) \\ &= u_j^n - \lambda\Big(v_{j-1/2} + \lambda\frac{\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right)}{\Delta_- u_j^n}\Big)\Delta_- u_j^n \\ &= u_j^n - A_{j-1/2}\Delta_- u_j^n, \end{aligned}$$

where we have defined

$$A_{j-1/2} = v_{j-1/2} + \lambda\frac{\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right)}{\Delta_- u_j^n}.$$

At this point the following lemma is convenient.

Lemma 3.12 (Harten's lemma) *Let v_j be given by*

$$v_j = u_j - A_{j-1/2}\Delta_- u_j + B_{j+1/2}\Delta_+ u_j,$$

where $\Delta_\pm u_j = \pm(u_{j\pm1} - u_j)$.

(i) *If $A_{j+1/2}$ and $B_{j+1/2}$ are nonnegative for all j, and $A_{j+1/2} + B_{j+1/2} \le 1$ for all j, then*

$$\text{T.V.}(v) \le \text{T.V.}(u).$$

(ii) *If $A_{j+1/2}$ and $B_{j+1/2}$ are nonnegative for all j, and $A_{j-1/2} + B_{j+1/2} \le 1$ for all j, then*

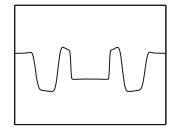
$$\min_k u_k \le v_j \le \max_k u_k, \quad j \in \mathbb{Z}.$$

Proof (i) We have

$$\begin{aligned} \Delta_+ v_j &= u_{j+1} - u_j - A_{j+1/2}\Delta_+ u_j + B_{j+3/2}\Delta_+ u_{j+1} \\ &\quad + A_{j-1/2}\Delta_- u_j - B_{j+1/2}\Delta_+ u_j \\ &= \left(1 - A_{j+1/2} - B_{j+1/2}\right)\Delta_+ u_j + A_{j-1/2}\Delta_- u_j + B_{j+3/2}\Delta_+ u_{j+3/2}. \end{aligned}$$

Hence

$$\begin{aligned} \sum_j \left|\Delta_+ v_j\right| &\le \sum_j \left(1 - A_{j+1/2} - B_{j+1/2}\right)\left|\Delta_+ u_j\right| \\ &\quad + \sum_j A_{j-1/2}\left|\Delta_- u_j\right| + \sum_j B_{j+3/2}\left|\Delta_+ u_{j+3/2}\right| \\ &= \sum_j \left|\Delta_+ u_j\right|. \end{aligned}$$



(ii) We may write

$$v_j = A_{j-1/2}u_{j-1/2} + (1 - A_{j-1/2} - B_{j+1/2})u_j + B_{j+1/2}u_{j+1},$$

from which the statement follows. □

Returning to the scheme (3.43), we introduce

$$\alpha_{j+1/2} = \frac{1}{2}v_{j+1/2}\left(1 - \lambda v_{j+1/2}\right).$$

Hence, we get the scheme

$$\begin{aligned} u_j^{n+1} &= u_j^n - \lambda\left(f_j^n - f_{j-1}^n\right) \\ &\quad - \lambda\Big(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n - \varphi(r_{j-1})\alpha_{j-1/2}\Delta_- u_j^n\Big) \\ &= u_j^n - \lambda v_{j-1/2}\Delta_- u_j^n - \lambda\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right) \\ &= u_j^n - \lambda\left[v_{j-1/2} + \frac{\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right)}{\Delta_- u_j^n}\right]\Delta_- u_j^n \\ &= u_j^n - A_{j-1/2}\Delta_- u_j^n. \end{aligned}$$

We want to choose φ and r such that we can use the above lemma, with $B_{j+1/2} = 0$, to conclude that the scheme is TVD. Note that $\lambda \max_u f'(u) \le 1$ by the CFL condition and thus $\alpha_{j+1/2} \ge 0$ and $\lambda\alpha_{j+1/2} \le 1$.

We define

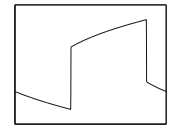
$$r_j = \frac{\alpha_{j-1/2}\Delta_- u_j}{\alpha_{j+1/2}\Delta_+ u_j}. \tag{3.45}$$

To see that this can be used as an "indicator of oscillations," note that since we have assumed that $f' \ge 0$, we have $v_{j+1/2} \ge 0$ for all j, and by the CFL condition, $\lambda v_{j+1/2} \le 1$ for all j. Hence $\alpha_{j+1/2} = \frac{1}{2}v_{j+1/2}(1 - \lambda v_{j+1/2}) \ge 0$ for all j. We say that "oscillations" are present at x_j if u_j is a local maximum or minimum. If so, then sign $\left(\Delta_- u_j\right) \ne$ sign $\left(\Delta_+ u_j\right)$, and consequently, $r_j \le 0$. We also calculate

$$\begin{aligned} \frac{\Delta_-\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n\right)}{\Delta_- u_j^n} &= \frac{1}{\Delta_- u_j^n}\left(\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n - \varphi(r_{j-1})\alpha_{j-1/2}\Delta_- u_j^n\right) \\ &= \alpha_{j-1/2}\left(\frac{\varphi(r_j)}{r_j} - \varphi(r_{j-1})\right). \end{aligned}$$

Hence

$$A_{j+1/2} = \lambda\Big(v_{j+1/2} + \alpha_{j+1/2}\Big(\frac{\varphi(r_{j+1})}{r_{j+1}} - \varphi(r_j)\Big)\Big).$$



Let us assume that

$$\max\left\{\frac{\varphi(r)}{r}, \varphi(r)\right\} \le 2, \quad \text{or} \quad 0 \le \varphi(r) \le \max\{0, \min\{2r, 2\}\}. \tag{3.46}$$

If this assumption holds, then

$$\left|\frac{\varphi(r)}{r} - \varphi(s)\right| \le 2 \quad \text{for all } r \text{ and } s.$$

This means that

$$\begin{aligned} A_{j+1/2} &\le \lambda\left(\nu_{j+1/2} + 2\alpha_{j+1/2}\right) \\ &= \lambda\left(\nu_{j+1/2} + \nu_{j+1/2}\left(1 - \lambda\nu_{j+1/2}\right)\right) \\ &= \lambda\left(2\nu_{j+1/2} - \lambda\nu_{j+1/2}^2\right) \\ &= 1 - (1 - \lambda\nu_{j+1/2})^2 \\ &\le 1. \end{aligned}$$

For the other bound,

$$\begin{aligned} A_{j+1/2} &\ge \lambda\left(\nu_{j+1/2} - 2\alpha_{j+1/2}\right) \\ &= \lambda\left(\nu_{j+1/2} - \nu_{j+1/2}(1 - \lambda\nu_{j+1/2})\right) \\ &= \left(\lambda\nu_{j+1/2}\right)^2 \ge 0. \end{aligned}$$

Summing up, we have proved the following result.

Lemma 3.13 *Assume $f' \ge 0$. Let r_j be defined by (3.45), and assume $\lambda > 0$ is such that the CFL condition $\lambda \max_u f'(u) \le 1$ holds. Assume further that the function φ is such that $\varphi(r)$ vanishes for $r \le 0$ and satisfies (3.46). Then the finite volume scheme with numerical flux function (3.43) is TVD.*

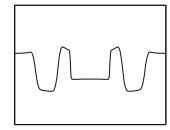
If we choose $\varphi(r) = r$, we get another scheme, called the Beam–Warming (BW) scheme. The Beam–Warming scheme is also of second order, but not TVD. The Lax–Wendroff (LW) scheme is obtained by choosing $\varphi(r) = 1$.

If (for the moment) we do not care about TVD, we can define a family of second-order schemes by linear interpolation between the Beam–Warming and the Lax–Wendroff schemes. This interpolation can be done locally, meaning that we choose φ as

$$\varphi(r) = (1 - \theta(r))\varphi_{\mathrm{LW}}(r) + \theta(r)\varphi_{\mathrm{BW}}(r).$$

The scheme reads

$$u_j^{n+1} = u_j^n - \lambda\Delta_- f_j^n + \lambda\Delta_-\varphi(r_j)\alpha_{j+1/2}\Delta_+ u_j^n. \tag{3.47}$$



If now $u_j^n = u(x_j, t_n)$ is the exact solution, then we can calculate

$$\begin{aligned} u_j^{n+1} &= u_j - \lambda \Delta f_j + \lambda \Delta_- \left(\left((1 - \theta(r_j)) + \theta(r_j) r_j \right) \alpha_{j+1/2} \Delta_+ u_j \right) \\ &= \left(1 - \theta(r_j)\right) \left(u_j - \lambda \Delta_- f_j + \lambda \Delta_- \left(\alpha_{j+1/2} \Delta_+ u_j \right) \right) \\ &\quad + \theta(r_j) \left(u_j - \lambda \Delta_- f_j + \lambda \Delta_- \left(r_j \alpha_{j+1/2} \Delta_+ u_j \right) \right) \\ &\quad + \lambda \alpha_{j-1/2} \left(r_{j-1} - 1 \right) \Delta_- u_j \Delta_- \theta(r_j). \end{aligned}$$

This means that

$$\begin{aligned} u(x_j, t + \Delta t) - u_j^{n+1} = (1 - \theta(r_j)) \quad & (\text{“LW truncation error”}) \\ + \theta(r_j) \quad & (\text{“BW truncation error”}) \\ + \lambda \underbrace{\alpha_{j-1/2} \left(r_{j-1} - 1 \right) \Delta_- u_j \Delta_- \theta(r_j)}_{I} & . \end{aligned}$$

If $I = \mathcal{O}\left(\Delta t^3\right)$, then the combination of the LW and the BW schemes is of second order. By the CFL condition, $0 \le \lambda \alpha_{j-1/2} \le 1$. Furthermore, since u is an exact smooth solution, $\alpha_{j+1/2} \Delta_+ u \approx \Delta x f'(u)(1 - \lambda f'(u)) u_x$, or more precisely

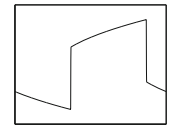
$$\alpha_{j+1/2} \frac{\Delta_+ u_j}{\Delta x} = f'(u)(1 - \lambda f'(u)) u_x \Big|_{x = x_{j+1/2}} + \mathcal{O}\left(\Delta x^2\right).$$

Recall the definition of r_j, equation (3.45), and set $h(x) = f'(u(x,t))(1 - \lambda f'(u(x,t))) u_x(x,t)$. With this notation we get

$$\begin{aligned} \left| \alpha_{j-1/2} (r_{j-1} - 1) \Delta_- u_j \right| &= \left| \Delta_- \left(\alpha_{j-1/2} \Delta_+ u_{j-1} \right) \right| \\ &= \Delta x \left| h(x_{j-1/2}) - h(x_{j-3/2}) + \mathcal{O}\left(\Delta x^2\right) \right| \\ &\le \Delta x^2 \max_{(x,t)} |h'(x)| + \mathcal{O}\left(\Delta x^3\right). \end{aligned}$$

Therefore, to show that $I = \mathcal{O}\left(\Delta t^3\right)$, it suffices to show that $\Delta_- \theta_j = \mathcal{O}(\Delta t)$. Since θ is a smooth function with values in $[0, 1]$, we get

$$\begin{aligned} \left| \Delta_- \theta(r_j) \right| &= \left| \theta(r_j) - \theta(r_{j-1}) \right| \\ &\le C \left| r_j - r_{j-1} \right| \\ &\le C \left| \frac{\alpha_{j-1/2} \Delta_- u_j}{\alpha_{j+1/2} \Delta_+ u_j} - \frac{\alpha_{j-3/2} \Delta_- u_{j-1}}{\alpha_{j-1/2} \Delta_- u_j} \right| \\ &= C \left| \frac{h_{j-1/2} + \mathcal{O}\left(\Delta x^2\right)}{h_{j+1/2} + \mathcal{O}\left(\Delta x^2\right)} - \frac{h_{j-3/2} + \mathcal{O}\left(\Delta x^2\right)}{h_{j-1/2} + \mathcal{O}\left(\Delta x^2\right)} \right| \\ &= C \left| \frac{h_{j-1/2}^2 - h_{j+1/2} h_{j-3/2} + \mathcal{O}\left(\Delta x^2\right)}{h_{j+1/2} h_{j-1/2} + \mathcal{O}\left(\Delta x^2\right)} \right| \\ &\le C \frac{\Delta x \max_{(x,t)} |h'(x)| + \mathcal{O}\left(\Delta x^2\right)}{h_{j+1/2} h_{j-3/2} + \mathcal{O}\left(\Delta x^2\right)} \\ &= \mathcal{O}(\Delta t). \end{aligned}$$



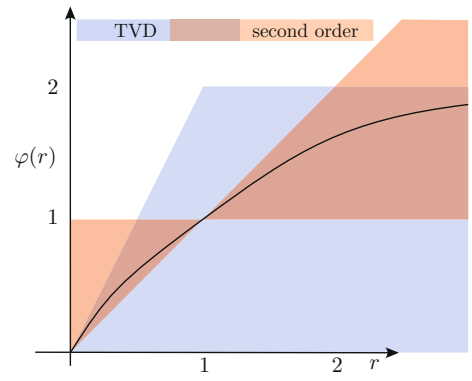


Fig. 3.5 The graph of the limiter must lie in both the TVD region and the second-order region. The graph shown is a possible limiter

Thus we have shown that if θ is a Lipschitz continuous function, the resulting scheme is of second order.

Returning to φ, we have shown that the scheme (3.47) is of second order if φ is Lipschitz continuous and

$$\min\{1, r\} \leq \varphi(r) \leq \max\{1, r\}. \tag{3.48}$$

If φ satisfies both (3.46) and (3.48), then the resulting scheme (3.47) is TVD, and second-order accurate away from local extrema. The scheme also produces a convergent sequence of approximations, and the limit is a weak solution (prove this!).

The function φ is called a limiter; a list of popular limiters follows. It is clear that the graph of a limiter must lie in the shaded region in Fig. 3.5.

$$\begin{aligned}
&\varphi(r) = \max\{0, \min\{r, 1\}\}, && \text{minmod}\\
&\varphi(r) = \max\{0, \min\{2r, 1\}, \min\{r, 2\}\}, && \text{superbee,}\\
&\varphi(r) = \frac{|r| + r}{1 + r}, && \text{van Leer}\\
&\varphi(r) = \frac{r^2 + r}{1 + r^2}, && \text{van Albada}\\
&\varphi(r) = \max\{0, \min\{r, \beta\}\}, \quad 1 \leq \beta \leq 2, && \text{Chakarvarthy \& Osher}
\end{aligned}$$

In Fig. 3.6 we show the approximate solutions to

$$u_t + u_x = 0, \quad u(x, 0) = \begin{cases} 1 & x \in [0, 3,\ 0.7], \\ 0 & x \in [0, 1] \setminus [0, 3,\ 0.7], \end{cases}$$

and for $x \notin [0, 1]$ we extend $u(x, 0)$ periodically. The figure shows approximate solutions at $t = 0$ as well as the exact solution. To the left we see that both the Lax–Wendroff and the Beam–Warming schemes have pronounced oscillations, but the linear combination of the two schemes, in this case using the van Leer limiter, does not. This solution is also superior to the solution found by the upwind method. Since these methods limit the contribution of the higher-order numerical flux function, they are often called *flux-limiter methods*.

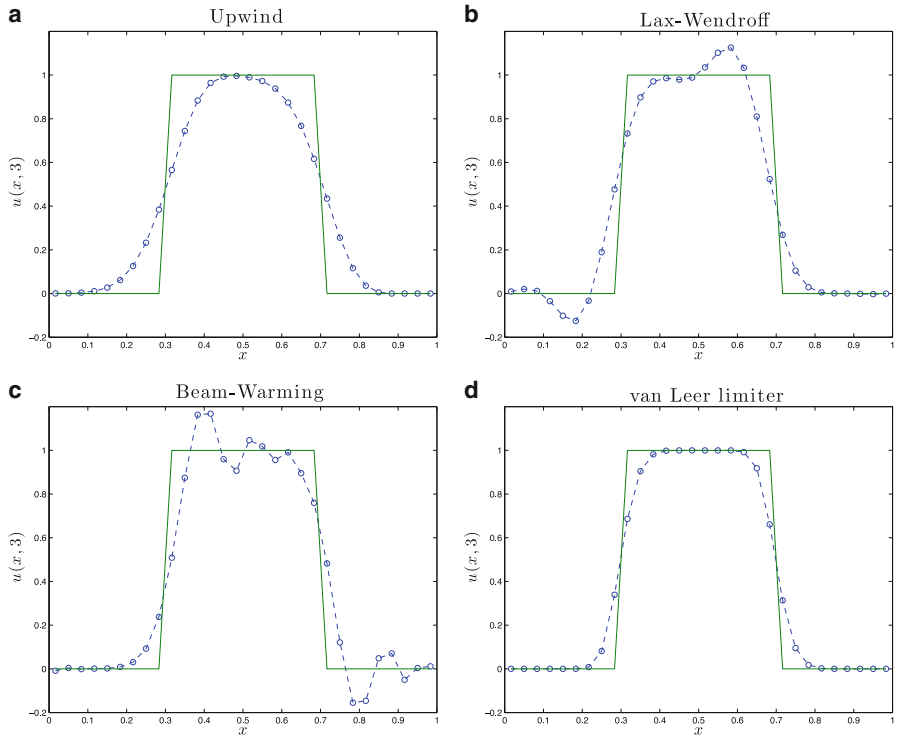


Fig. 3.6 The approximate solutions found by the upwind method (**a**), the Lax–Wendroff method (**b**), the Beam–Warming method (**c**), and the TVD method using the van Leer limiter (**d**). All computations used $\Delta x = 1/30$

Semidiscrete Higher-Order Methods

Let us now consider semidiscrete higher-order methods, where we do not (initially) discretize time, only space. Based on the finite volume approach, such methods can be written

$$u_j'(t) = -\frac{1}{\Delta x}\left(F_{j+1/2} - F_{j-1/2}\right), \tag{3.49}$$

where $u_j(t)$ is some approximation to the average of u in the cell $(x_{j-1/2}, x_{j+1/2}]$. If the right-hand side of the above is a second-order approximation to $-f(u)_x$ for smooth functions $u(x)$, then the method is said to be second-order accurate. To get second-order accuracy in time as well, one could use a second-order Runge–Kutta method to integrate (3.49) numerically. One such example is Heun's method:

$$\begin{aligned}
\tilde{u}_j^n &= u_j^n - \lambda\left(F_{j+1/2} - F_{j-1/2}\right), \\
u_j^{n+1} &= u_j^n - \frac{\lambda}{2}\left(\tilde{F}_{j+1/2} - \tilde{F}_{j-1/2}\right) - \frac{\lambda}{2}\left(F_{j+1/2} - F_{j-1/2}\right).
\end{aligned}$$

The simplest way of achieving second-order accuracy is by choosing

$$F_{j+1/2} = f\left(\frac{u_{j+1} + u_j}{2}\right). \tag{3.50}$$

This, however, gives a nonviable method if we combine it with a first-order Euler method in time. This combination is not stable[2]. To see this, set $f(u) = u$. With the Euler method it gives

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}\left(u_{j+1} - u_{j-1}\right).$$

Making the *ansatz* $u_j^n = \mu_n e^{ij\Delta x}$ (here $i = \sqrt{-1}$) yields

$$\mu_{n+1} = \mu_n\left(1 + i\lambda \sin(\Delta x)\right).$$

Therefore, $|\mu_{n+1}| = |\mu_n|\sqrt{1 + \lambda^2 \sin^2(\Delta x)}$, or

$$|\mu_n| = |\mu_0|\left(1 + \lambda^2 \sin^2(\Delta x)\right)^{n/2}.$$

This is unconditionally unstable. Also using the second-order Heun's method with (3.50) gives an unstable method (see Exercise 3.8). Thus the choice (3.50) is of second order, but useless.

In order to overcome this, we define values to the left and right of a cell edge $u^L_{j+1/2}$ and $u^R_{j-1/2}$ by

$$\begin{aligned} u^L_{j+1/2} &= u_j + \frac{1}{2}\Delta_- u_j, \\ u^R_{j-1/2} &= u_j - \frac{1}{2}\Delta_+ u_j. \end{aligned} \tag{3.51}$$

Then we can use any two-point monotone first-order numerical flux $F(u, v)$ to define a second-order approximation

$$f(u(x))_x = \frac{1}{\Delta x}\left(F\left(u^L_{j+1/2}, u^R_{j+1/2}\right) - F\left(u^L_{j-1/2}, u^R_{j-1/2}\right)\right) + \mathcal{O}\left(\Delta x^2\right). \tag{3.52}$$

Even if we use Heun's method for time integration, the extrapolation values (3.51) do not give a TVD method. This is to be expected, since the method is formally second-order accurate. We illustrate this in Fig. 3.7 for the linear equation $u_t + u_x = 0$ with smooth and discontinuous initial values. We used the upwind first-order numerical flux $F(u, v) = f(u) = u$. From Fig. 3.7 we see that for smooth initial data, the approximation is "reasonably close" to the correct function, whereas for discontinuous initial data, the approximation bears little relation to the exact solution.

[2] Often called von Neumann stability.

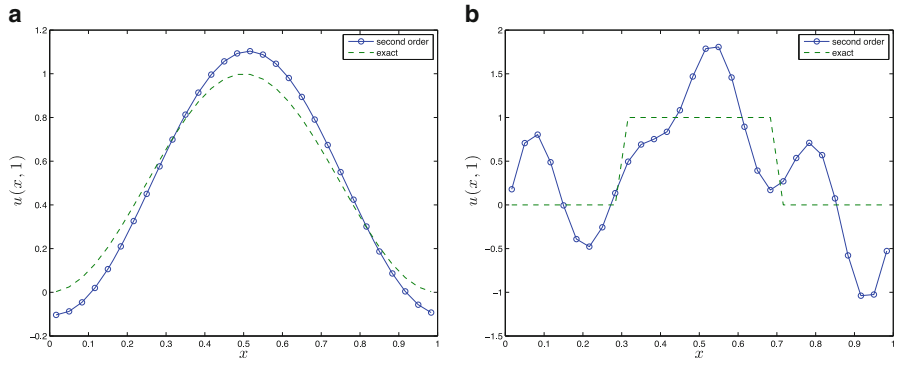


Fig. 3.7 Using the extrapolation (3.51). **a** $u(x,1)$ with smooth initial data. **b** $u(x,1)$ with discontinuous intitial data

These results suggest that the method will be improved if we use some kind of limiter to define the extrapolated values $u^{L,R}_{j+1/2}$. To this end, set $\varphi_j = \varphi(r_j)$, where r_j is to be defined, and redefine the extrapolations as

$$\begin{aligned} u^L_{j+1/2} &= u_j + \frac{1}{2}\varphi_j \Delta_- u_j, \\ u^R_{j-1/2} &= u_j - \frac{1}{2}\varphi_j \Delta_+ u_j. \end{aligned} \tag{3.53}$$

For simplicity, we now assume that $f' \geq 0$, and that the numerical flux function is the upwind flux, i.e., $F(u,v) = f(u)$. In this case the resulting scheme is

$$u^{n+1}_j = u^n_j - \lambda \left(f(u^L_{j+1/2}) - f(u^L_{j-1/2}) \right).$$

We aim to define r_j and find conditions on φ such that the above scheme is TVD but retains the formal second order away from oscillations. In order to use Lemma 3.12, we rewrite the scheme as

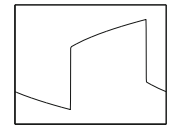
$$u^{n+1}_j = u^n_j - \lambda \frac{\Delta_- f(u^L_{j+1/2})}{\Delta_- u^n_j} \Delta_- u^n_j,$$

where we have used a first-order Euler method for the integration in time. This will of course destroy the formal second-order accuracy, but it is convenient for analysis. With

$$A_{j-1/2} = \lambda \frac{\Delta_- f(u^L_{j+1/2})}{\Delta_- u^n_j}$$

the scheme will be TVD if $0 \leq A_{j-1/2} \leq 1$. Dropping the superscript n, we calculate

$$\begin{aligned} A_{j-1/2} &= \lambda f'\left(\bar{u}_j\right) \frac{u_j + \frac{1}{2}\varphi_j \Delta_- u_j - u_{j-1} - \frac{1}{2}\varphi_{j-1}\Delta_- u_{j-1}}{\Delta_- u_j} \\ &= \lambda f'\left(\bar{u}_j\right) \left(\left(1 + \frac{1}{2}\varphi_j\right) - \frac{1}{2}\varphi_{j-1} \frac{\Delta_- u_{j-1}}{\Delta_- u_j} \right), \end{aligned}$$



where $\bar{u}_j$ is some value between $u^L_{j-1/2}$ and $u^L_{j+1/2}$. If we now choose

$$r_j = \frac{\Delta_+ u_j}{\Delta_- u_j},$$

this can be rewritten as

$$A_{j-1/2} = \lambda f'(\bar{u}_j)\left(1 - \frac{1}{2}\left(\frac{\varphi(r_{j-1})}{r_{j-1}} - \varphi(r_j)\right)\right).$$

We now demand that the scheme satisfy the CFL condition

$$\lambda \max_u f'(u) \le \frac{1}{2}.$$

In this case $0 \le A_{j-1/2} \le 1$ if

$$0 \le \left(1 - \frac{1}{2}\left(\frac{\varphi(r_{j-1})}{r_{j-1}} - \varphi(r_j)\right)\right) \le 2,$$

which can be rewritten

$$-2 \le \frac{\varphi_{j-1}}{r_{j-1}} - \varphi_j \le 2.$$

This is the case if

$$0 \le \varphi(r) \le \min\{2r, 2\},$$

which gives the same TVD-region as for the flux-limiter schemes; see Fig. 3.5.

The scheme with $\phi(r) \equiv 1$ is not TVD, but of second order, and the choice $\phi(r) = r$ gives the (useless) second-order scheme with numerical flux (3.50). It follows as before that every smooth (in r) convex combination of these two schemes will also be of second order. Therefore, we get the same second-order region as in Fig. 3.5. Hence we have the same choice of limiter functions as before. Each choice will give a formally second-order scheme away from local extrema. This method is called *MUSCL* (monotone upstream centered scheme for conservation laws).

If Fig. 3.8 we show how the above schemes perform on the model equation $u_t + u_x = 0$ with smooth and discontinuous initial data. The MUSCL method does not perform as well as the flux limiter method, but a clear difference can be seen between the first-order upstream method and the high-resolution methods (MUSCL and flux limiter). For both the high-resolution methods, the computations in Fig. 3.8 use the van Leer limiter. The perceptive reader may have noticed that the flux-limiter method is further from the exact solution than the methods shown in Fig. 3.6. This is because we choose to use the same timestep for all the methods, this being limited by the MUSCL method. Thus, the upwind and flux limiter methods will also have a time step $\Delta t \le \lambda \Delta x$, with $\lambda = 0.49$.

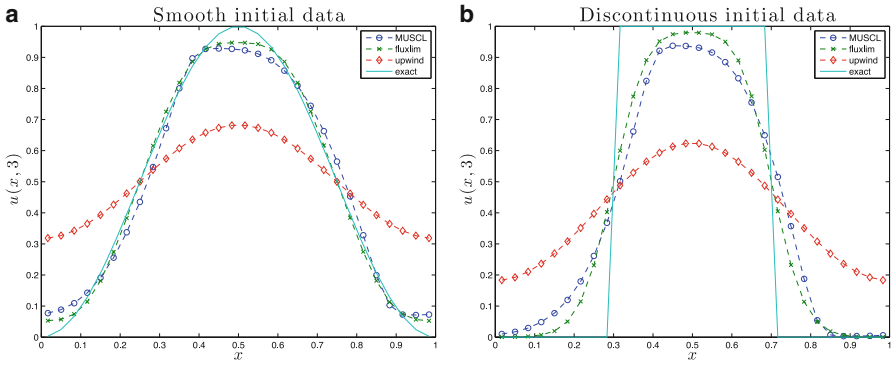


Fig. 3.8 A comparison of the first-order monotone upstream method and high-resolution methods for smooth (**a**) and discontinuous initial data (**b**)

3.3 Error Estimates

Let others bring order to chaos. I would bring chaos to order instead.
— Kurt Vonnegut, Breakfast of Champions (1973)

The concept of local error estimates is based on formal computations, and such estimates indicate how the method performs in regions where the solution is smooth. Since the convergence of the methods discussed was in L^1, it is reasonable to ask how far the approximated solution is from the true solution in this space.

In this section we will consider functions u that are maps $t \mapsto u(t)$ from $[0, \infty)$ to $L^1_{\text{loc}} \cap BV(\mathbb{R})$ such that the one-sided limits $u(t\pm)$ exist in L^1_{loc}, and for definiteness we assume that this map is right continuous. Furthermore, we assume that

$$\|u(t)\|_\infty \leq \|u(0)\|_\infty, \quad \text{T.V.}\,(u(t)) \leq \text{T.V.}\,(u(0))\,.$$

We denote this class of functions by $\mathcal{K}$. From Theorem 2.15 we know that solutions of scalar conservation laws are in the class $\mathcal{K}$.

It is convenient to introduce *moduli of continuity in time* (see Appendix A)

$$\begin{aligned} \nu_t(u,\sigma) &= \sup_{|\tau|\leq\sigma} \|u(t+\tau) - u(t)\|_{L^1}, \quad \sigma > 0, \\ \nu(u,\sigma) &= \sup_{0\leq t\leq T} \nu_t(u,\sigma). \end{aligned} \tag{3.54}$$

From Theorem 2.15 we have that

$$\nu(u,\sigma) \leq |\sigma|\,\|f\|_{\text{Lip}}\text{T.V.}\,(u_0) \tag{3.55}$$

for weak solutions of conservation laws.

Now let $u(x,t)$ be any function in $\mathcal{K}$, not necessarily a solution of (3.1). In order to measure how far u is from being a solution of (3.1) we insert u in the Kružkov

form (cf. (2.23))

$$\Lambda_T(u,\phi,k) = \int_0^T \int \left(|u-k|\,\phi_t + q(u,k)\phi_x\right)\,dx\,ds \tag{3.56}$$
$$- \int |u(x,T)-k|\,\phi(x,T)\,dx + \int |u_0(x)-k|\,\phi(x,0)\,dx.$$

If u is a solution, then $\Lambda_T \geq 0$ for all constants k and all nonnegative test functions ϕ. We shall now use the special test function

$$\Omega(x,x',s,s') = \omega_{\varepsilon_0}(s-s')\omega_\varepsilon(x-x'),$$

where

$$\omega_\varepsilon(x) = \frac{1}{\varepsilon}\omega\left(\frac{x}{\varepsilon}\right)$$

and $\omega(x)$ is an even C^∞ function satisfying

$$0 \leq \omega \leq 1, \quad \omega(x) = 0 \quad \text{for } |x| > 1, \quad \int \omega(x)\,dx = 1.$$

Let $v(x',s')$ be the unique weak solution of (3.1), and define

$$\Lambda_{\varepsilon,\varepsilon_0}(u,v) = \int_0^T \int \Lambda_T\left(u, \Omega(\cdot,x',\cdot,s'), v(x',s')\right)\,dx'ds'.$$

The comparison result reads as follows.

Theorem 3.14 (Kuznetsov's lemma) *Let $u(\cdot,t)$ be a function in $\mathcal{K}$, and v a solution of (3.1). If $0 < \varepsilon_0 < T$ and $\varepsilon > 0$, then*

$$\|u(\cdot,T-) - v(\cdot,T)\|_{L^1(\mathbb{R})} \leq \|u_0 - v_0\|_{L^1(\mathbb{R})} + \text{T.V.}\,(v_0)\left(2\varepsilon + \varepsilon_0\|f\|_{\text{Lip}}\right)$$
$$+ \nu(u,\varepsilon_0) - \Lambda_{\varepsilon,\varepsilon_0}(u,v), \tag{3.57}$$

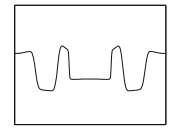
where $u_0 = u(\cdot,0)$ and $v_0 = v(\cdot,0)$.

Proof We use special properties of the test function Ω, namely that

$$\Omega(x,x',s,s') = \Omega(x',x,s,s') = \Omega(x,x',s',s) = \Omega(x',x,s',s) \tag{3.58}$$

and

$$\Omega_x = -\Omega_{x'}, \quad \text{and} \quad \Omega_s = -\Omega_{s'}. \tag{3.59}$$



Using (3.58) and (3.59), we find that

$$\begin{aligned}\Lambda_{\varepsilon,\varepsilon_0}(u,v) &= -\Lambda_{\varepsilon,\varepsilon_0}(v,u) - \int_0^T \iint \Omega(x,x',s,T)\big(\,|u(x,T)-v(x',s)| \\ &\qquad + |v(x',T)-u(x,s)|\,\big)\,dx\,dx'\,ds \\ &\quad + \int_0^T \iint \Omega(x,x',s,0)\big(\,|v_0(x')-u(x,s)| \\ &\qquad + |u_0(x)-v(x',s)|\,\big)\,dx\,dx'\,ds \\ &:= -\Lambda_{\varepsilon,\varepsilon_0}(v,u) - A \;+ B.\end{aligned}$$

Since v is a weak solution, $\Lambda_{\varepsilon,\varepsilon_0}(v,u) \geq 0$, and hence

$$A \leq B - \Lambda_{\varepsilon,\varepsilon_0}(u,v).$$

Therefore, we would like to obtain a lower bound on A and an upper bound on B, the lower bound on A involving $\|u(T)-v(T)\|_{L^1}$ and the upper bound on B involving $\|u_0-v_0\|_{L^1}$. We start with the lower bound on A.

Let ρ_ε be defined by

$$\rho_\varepsilon(u,v) = \iint \omega_\varepsilon(x-x')\,|u(x)-v(x')|\;dx\,dx'. \tag{3.60}$$

Then

$$A = \int_0^T \omega_{\varepsilon_0}(T-s)\,(\rho_\varepsilon(u(T),v(s)) + \rho_\varepsilon(u(s),v(T)))\;ds.$$

Now by a use of the triangle inequality,

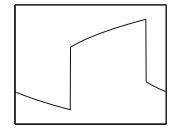
$$\begin{aligned}&\|u(x,T)-v(x',s)\| + |u(x,s)-v(x',T)| \\ &\quad \geq |u(x,T)-v(x,T)| + |u(x,T)-v(x,T)| \\ &\qquad - |v(x,T)-v(x',T)| - |u(x,T)-u(x,s)| \\ &\qquad - |v(x',T)-v(x',s)| - |v(x,T)-v(x',T)|\,.\end{aligned}$$

Hence

$$\begin{aligned}\rho_\varepsilon(u(T),v(s)) + \rho_\varepsilon(u(s),v(T)) &\geq 2\|u(T)-v(T)\|_{L^1} - 2\rho_\varepsilon(v(T),v(T)) \\ &\quad - \|u(T)-u(s)\|_{L^1} - \|v(T)-v(s)\|_{L^1}.\end{aligned}$$

Regarding the upper estimate on B, we similarly have that

$$B = \int_0^T \omega_{\varepsilon_0}(s)\,[\rho_\varepsilon(u_0,v(s)) + \rho_\varepsilon(u(s),v_0)]\;ds,$$



and we also obtain

$$\rho_\varepsilon(u_0, v(s)) + \rho_\varepsilon(u(s), v_0) \le 2\|u_0 - v_0\|_{L^1} + 2\rho_\varepsilon(v_0, v_0) + \|u_0 - u(s)\|_{L^1} + \|v_0 - v(s)\|_{L^1}.$$

Since v is a solution, it satisfies the TVD property, and hence

$$\begin{aligned}\rho_\varepsilon(v(T), v(T)) &= \int \int_{-\varepsilon}^{\varepsilon} \omega_\varepsilon(z)\, |v(x+z,T) - v(x,T)|\, dz\, dx \\ &\le \int_{-\varepsilon}^{\varepsilon} \omega_\varepsilon(z) \sup_{|z|\le\varepsilon} \left(\int |v(x+z,T) - v(x,T)|\, dx \right) dz \\ &= |\varepsilon| \int_{-\varepsilon}^{\varepsilon} \omega_\varepsilon(z) \mathrm{T.V.}\,(v(T))\, dz \le |\varepsilon|\, \mathrm{T.V.}\,(v_0)\,,\end{aligned}$$

using (A.10). By the properties of ω,

$$\int_0^T \omega_\varepsilon(T-s)\, ds = \int_0^T \omega_\varepsilon(s)\, ds = \frac{1}{2}.$$

Applying (3.55), we obtain (recall that $\varepsilon_0 < T$)

$$\begin{aligned}&\int_0^T \omega_{\varepsilon_0}(T-s)\|v(T) - v(s)\|_{L^1}\, ds \\ &\qquad \le \int_0^T \omega_{\varepsilon_0}(T-s)\,(T-s)\|f\|_{\mathrm{Lip}}\mathrm{T.V.}\,(v_0)\, ds \\ &\qquad \le \frac{1}{2}\varepsilon_0\|f\|_{\mathrm{Lip}}\mathrm{T.V.}\,(v_0)\end{aligned}$$

and

$$\int_0^T \omega_{\varepsilon_0}(s)\|v_0 - v(s)\|_{L^1}\, ds \le \frac{1}{2}\varepsilon_0\|f\|_{\mathrm{Lip}}\mathrm{T.V.}\,(v_0)\,.$$

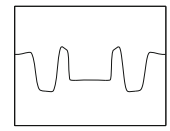
Similarly,

$$\int_0^T \omega_{\varepsilon_0}(T-s)\|u(T) - u(s)\|_{L^1}\, ds \le \frac{1}{2}\nu\,(u, \varepsilon_0)$$

and

$$\int_0^T \omega_{\varepsilon_0}(s)\|u_0 - u(s)\|_{L^1}\, ds \le \frac{1}{2}\nu\,(u, \varepsilon_0)\,.$$

If we collect all the above bounds, we should obtain the statement of the theorem. □



Observe that in the special case that u is a solution of the conservation law (3.1), we know that $\Lambda_{\varepsilon,\varepsilon_0}(u,v) \geq 0$, and hence we obtain, as $\varepsilon, \varepsilon_0 \to 0$, the familiar stability result

$$\|u(\cdot\,,T) - v(\cdot\,,T)\|_{L^1} \leq \|u_0 - v_0\|_{L^1}.$$

We shall now show in three cases how Kuznetsov's lemma can be used to give estimates on how fast a method converges to the entropy solution of (3.1).

◇ Example 3.15 (The smoothing method)
While not a proper numerical method, the smoothing method provides an example of how the result of Kuznetsov may be used. The smoothing method is a (semi)numerical method approximating the solution of (3.1) as follows: Let $\omega_\delta(x)$ be a standard mollifier with support in $[-\delta,\delta]$, and let $t_n = n\Delta t$. Set $u^0 = u_0 * \omega_\delta$. For $0 \leq t < \Delta t$ define u^1 to be the solution of (3.1) with initial data u^0. If Δt is small enough, u^1 remains differentiable for $t < \Delta t$. In the interval $[(n-1)\Delta t, n\Delta t)$, we define u^n to be the solution of (3.1), with $u^n(x,(n-1)\Delta t) = u^{n-1}(\cdot\,,t_n-) * \omega_\delta$. The advantage of doing this is that u^n will remain differentiable in x for all times, and the solution in the strips $[t_n, t_{n+1})$ can be found by, e.g., the method of characteristics. To show that u^n is differentiable, we calculate

$$\begin{aligned} |u^n_x(x,t_{n-1})| &= \left| \int u^{n-1}_x(y,t_{n-1})\omega_\delta(x-y)\,dy \right| \\ &\leq \frac{1}{\delta}\text{T.V.}\left(u^{n-1}(t_{n-1})\right) \leq \frac{\text{T.V.}(u_0)}{\delta}. \end{aligned}$$

Let $\mu(t) = \max_x |u_x(x,t)|$. Using that u is a classical solution of (3.1), we find by differentiating (3.1) with respect to x that

$$u_{xt} + f'(u)u_{xx} + f''(u)u_x^2 = 0.$$

Write

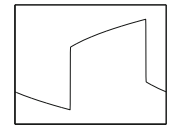
$$\mu(t) = u_x(x_0(t),t),$$

where $x_0(t)$ is the location of the maximum of $|u_x|$. Then

$$\begin{aligned} \mu'(t) &= u_{xx}(x_0(t),t)x_0'(t) + u_{xt}(x_0(t),t) \\ &\leq u_{xt}(x_0(t),t) = -f''(u)\big(u_x(x_0(t),t)\big)^2 \\ &\leq c\mu(t)^2, \end{aligned}$$

since $u_{xx} = 0$ at an extremum of u_x. Thus

$$\mu'(t) \leq c\mu^2(t), \tag{3.61}$$



where $c = \|f''\|_\infty$. The idea is now that (3.61) has a blowup at some finite time, and we choose Δt less than this time. We shall be needing a precise relation between Δt and δ and must therefore investigate (3.61) further. Solving (3.61) we obtain

$$\mu(t) \le \frac{\mu(t_n)}{1 - c\mu(t_n)(t - t_n)} \le \frac{\text{T.V.}(u_0)}{\delta - c\text{T.V.}(u_0)\Delta t}.$$

So if

$$\Delta t < \frac{\delta}{c\text{T.V.}(u_0)}, \tag{3.62}$$

the method is well defined. Choosing $\Delta t = \delta/(2c\text{T.V.}(u_0))$ will do.

Since u is an exact solution in the strips $[t_n, t_{n+1})$, we have

$$\int_{t_n}^{t_{n+1}} \int (|u-k|\,\phi_t + q(u,k)\phi_x)\,dx\,dt + \int \Big(|u(x,t_n+) - k|\,\phi(x,t_n) - |u(x,t_{n+1}-) - k|\,\phi(x,t_{n+1}) \Big)\,dx \ge 0.$$

Summing these inequalities and setting $k = v(y,s)$, where v is an exact solution of (3.1), we obtain

$$\Lambda_T(u,\Omega,v(y,s)) \ge -\sum_{n=0}^{N-1} \int \Omega(x,y,t_n,s) \Big(|u(x,t_n+) - v(y,s)| - |u(x,t_n-) - v(y,s)| \Big)\,dx,$$

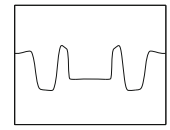
where we use the test function $\Omega(x,y,t,s) = \omega_{\varepsilon_0}(t-s)\omega_\varepsilon(x-y)$. Integrating this over y and s, and letting ε_0 tend to zero, we get

$$\liminf_{\varepsilon_0 \to 0} \Lambda_{\varepsilon,\varepsilon_0}(u,v) \ge -\sum_{n=0}^{N-1} \left(\rho_\varepsilon(u(t_n+), v(t_n)) - \rho_\varepsilon(u(t_n-), v(t_n)) \right).$$

Using this in Kuznetsov's lemma, and letting $\varepsilon_0 \to 0$, we obtain

$$\|u(T) - v(T)\|_1 \le \|u_0 - u^0\|_1 + 2\varepsilon\,\text{T.V.}(u_0) + \sum_{n=0}^{N-1} \left(\rho_\varepsilon(u(t_n+), v(t_n)) - \rho_\varepsilon(u(t_n-), v(t_n)) \right), \tag{3.63}$$

where we have used that $\lim_{\varepsilon_0 \to 0} \nu_t(u, \varepsilon_0) = 0$, which holds because u is a solution of the conservation law in each strip $[t_n, t_{n+1})$.



To obtain a more explicit bound on the difference of u and v, we investigate $\rho_\varepsilon(\omega_\delta * u, v) - \rho_\varepsilon(u, v)$, where ρ_ε is defined by (3.60),

$$\begin{aligned}\rho_\varepsilon(u * \omega_\delta, v) - \rho_\varepsilon(u, v) &\leq \iiint\limits_{|z|\leq 1} \omega_\varepsilon(x-y)\omega(z)\Big(|u(x+\delta z) - v(y)| \\ &\qquad - |u(x) - v(y)| \Big)\, dx\, dy\, dz \\ &= \frac{1}{2} \iiint\limits_{|z|\leq 1} \left(\omega_\varepsilon(x-y) - \omega_\varepsilon(x+\delta z - y)\right) \omega(z) \\ &\quad \times \left(|u(x+\delta z) - v(y)| - |u(x) - v(y)|\right)\, dx\, dy\, dz,\end{aligned}$$

which follows after writing $\iiint = \frac{1}{2}\iiint + \frac{1}{2}\iiint$ and making the substitution $x \mapsto x - \delta z$, $z \mapsto -z$ in one of these integrals. Therefore,

$$\begin{aligned}\rho_\varepsilon(u * \omega_\delta, v) - \rho_\varepsilon(u, v) &\leq \frac{1}{2} \iiint\limits_{|z|\leq 1} |\omega_\varepsilon(y+\delta z) - \omega_\varepsilon(y)| \\ &\qquad \times \omega(z)\, |u(x+\delta z) - u(x)|\, dx\, dy\, dz \\ &\leq \frac{1}{2}\, \text{T.V.}\,(\omega_\varepsilon)\ \text{T.V.}\,(u)\ \delta^2 \\ &\leq \text{T.V.}\,(u)\, \frac{\delta^2}{\varepsilon},\end{aligned}$$

by the triangle inequality and a further substitution $y \mapsto x - y$. Since $N = T/\Delta t$, the last term in (3.63) is less than

$$N\ \text{T.V.}\,(u_0)\, \frac{\delta^2}{\varepsilon} \leq (\text{T.V.}\,(u_0))^2\ 2cT \frac{\delta}{\varepsilon},$$

using (3.62). Furthermore, we have that

$$\left\| u^0 - u_0 \right\|_1 \leq \delta \text{T.V.}\,(u_0)\,.$$

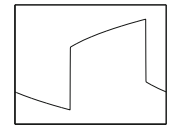
Letting $K = \text{T.V.}\,(u_0)\, c$, we find that

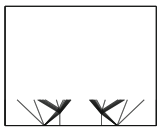
$$\|u(T) - v(T)\|_1 \leq 2\text{T.V.}\,(u_0) \left[\delta + \varepsilon + \frac{KT\delta}{\varepsilon} \right],$$

using (3.63). Minimizing with respect to ε, we find that

$$\|u(T) - v(T)\|_1 \leq 2\text{T.V.}\,(u_0) \left(\delta + 2\sqrt{KT\delta} \right). \tag{3.64}$$

So, we have shown that the smoothing method is of order $\frac{1}{2}$ in the smoothing coefficient δ. ◇





◇ Example 3.16 (The method of vanishing viscosity)
Another (semi)numerical method for (3.1) is the method of vanishing viscosity. Here we approximate the solution of (3.1) by the solution of

$$u_t + f(u)_x = \delta u_{xx}, \quad \delta > 0, \tag{3.65}$$

using the same initial data. Let u^δ denote the solution of (3.65). Due to the dissipative term on the right-hand side, the solution of (3.65) remains a classical (twice differentiable) solution for all $t > 0$. Furthermore, the solution operator for (3.65) is TVD. Hence a numerical method for (3.65) will (presumably) not experience the same difficulties as a numerical method for (3.1). If (η, q) is a convex entropy pair, we have, using the differentiability of the solution, that

$$\eta(u)_t + q(u)_x = \delta \eta'(u) u_{xx} = \delta \left(\eta(u)_{xx} - \eta''(u) u_x^2 \right).$$

Multiplying by a nonnegative test function φ and integrating by parts, we get

$$\iint \left(\eta(u)\varphi_t + q(u)\varphi_x \right) dx\, dt \geq \delta \iint \eta(u)_x \varphi_x \, dx\, dt,$$

where we have used the convexity of η. Applying this with $\eta = \left| u^\delta - u \right|$ and $q = F(u^\delta, u)$, we can bound $\lim_{\varepsilon_0 \to 0} \Lambda_{\varepsilon, \varepsilon_0}(u^\delta, u)$ as follows:

$$\begin{aligned}
-\lim_{\varepsilon_0 \to 0} \Lambda_{\varepsilon,\varepsilon_0}(u^\delta, u) &\leq \delta \int_0^T \iint \left| \frac{\partial \omega_\varepsilon(x-y)}{\partial x} \right| \frac{\partial \left| u^\delta(x,t) - u(y,t) \right|}{\partial x} \, dx\, dy\, dt \\
&\leq \delta \int_0^T \iint \left| \frac{\partial \omega_\varepsilon(x-y)}{\partial x} \right| \left| \frac{\partial u^\delta(x,t)}{\partial x} \right| dx\, dy\, dt \\
&\leq 2\text{T.V.}\left(u^\delta\right) T \frac{\delta}{\varepsilon} \\
&\leq 2T\ \text{T.V.}\left(u_0\right) \frac{\delta}{\varepsilon}.
\end{aligned}$$

Now letting $\varepsilon_0 \to 0$ in (3.57), we obtain

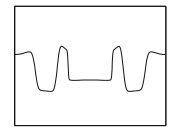
$$\left\| u^\delta(T) - u(T) \right\|_1 \leq \min_\varepsilon \left(2\varepsilon + \frac{2T\delta}{\varepsilon} \right) \text{T.V.}\left(u_0\right) = 2\text{T.V.}\left(u_0\right) \sqrt{T\delta}.$$

So the method of vanishing viscosity also has order $\frac{1}{2}$. ◇

◇ Example 3.17 (Monotone schemes)
We will here show that monotone schemes converge in L^1 to the solution of (3.1) at a rate of $(\Delta t)^{1/2}$. In particular, this applies to the Lax–Friedrichs scheme.

Let $u_{\Delta t}$ be defined by (3.27), where u_j^n is defined by (3.6), that is,

$$u_j^{n+1} = u_j^n - \lambda \left(F_{j+1/2}^n - F_{j-1/2}^n \right), \tag{3.66}$$



where $F^n_{j+1/2} = F\left(u^n_{j-p}, \dots, u^n_{j+p'}\right)$, for a scheme that is assumed to be monotone; cf. Definition 3.5. In the following we use the notation

$$\eta^n_j = \left|u^n_j - k\right|, \quad q^n_j = f\left(u^n_j \vee k\right) - f\left(u^n_j \wedge k\right).$$

We find that

$$\begin{aligned}
-\Lambda_T(u_{\Delta t}, \phi, k) &= -\sum_j \sum_{n=0}^{N-1} \int_{x_{j-1/2}}^{x_{j+1/2}} \int_{t_n}^{t_{n+1}} \left(\eta^n_j \phi_t(x,s) + q^n_j \phi_x(x,s)\right) ds\, dx \\
&\quad - \sum_j \int_{x_{j-1/2}}^{x_{j+1/2}} \eta^0_j \phi(x,0)\, dx + \sum_j \int_{x_{j-1/2}}^{x_{j+1/2}} \eta^N_j \phi(x,T)\, dx \\
&= -\sum_j \Bigg[\sum_{n=0}^{N-1} \int_{x_{j-1/2}}^{x_{j+1/2}} \eta^n_j \left(\phi(x,t_{n+1}) - \phi(x,t_n)\right) dx \\
&\qquad + \int_{x_{j-1/2}}^{x_{j+1/2}} \eta^0_j \phi(x,0)\, dx - \int_{x_{j-1/2}}^{x_{j+1/2}} \eta^N_j \phi(x,T)\, dx \\
&\qquad + \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} q^n_j \left(\phi(x_{j+1/2},s) - \phi(x_{j-1/2},s)\right) ds \Bigg] \\
&= \sum_j \sum_{n=0}^{N-1} \Bigg((\eta^{n+1}_j - \eta^n_j) \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x,t_{n+1})\, dx \\
&\qquad + (q^n_j - q^n_{j-1}) \int_{t_n}^{t_{n+1}} \phi(x_{j-1/2},s)\, ds \Bigg)
\end{aligned}$$

by a summation by parts. Recall that we define the numerical entropy flux by

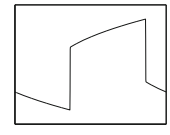
$$Q^n_{j+1/2} = F(u^n_{j-p} \vee k, \dots, u^n_{j+p'} \vee k) - F(u^n_{j-p} \wedge k, \dots, u^n_{j+p'} \wedge k).$$

Monotonicity of the scheme implies, cf. (3.33), that

$$\eta^{n+1}_j - \eta^n_j + \lambda(Q^n_{j+1/2} - Q^n_{j-1/2}) \le 0.$$

For a nonnegative test function ϕ we obtain

$$\begin{aligned}
-\Lambda_T(u_{\Delta t}, \phi, k) \le \sum_j \sum_{n=0}^{N-1} \Bigg(& -\lambda(Q^n_{j+1/2} - Q^n_{j-1/2}) \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x,t_{n+1})\, dx \\
& + (q^n_j - q^n_{j-1}) \int_{t_n}^{t_{n+1}} \phi(x_j,s)\, ds \Bigg)
\end{aligned}$$



$$
\begin{aligned}
&= \sum_j \sum_{n=0}^{N-1} \Big[\lambda \left(\left(q_j^n - Q_{j+1/2}^n \right) - \left(q_{j-1}^n - Q_{j-1/2}^n \right) \right) \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x, t_{n+1})\, dx \\
&\qquad + (q_j^n - q_{j-1}^n) \Big(\int_{t_n}^{t_{n+1}} \phi(x_{j-1/2}, s)\, ds - \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x, t_{n+1})\, dx \Big) \Big] \\
&= \sum_j \sum_{n=0}^{N-1} \Big[\lambda \left(Q_{j+1/2}^n - q_j^n \right) \Big(\int_{x_{j+1/2}}^{x_{j+3/2}} \phi(x, t_{n+1})\, dx - \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x, t_{n+1})\, dx \Big) \\
&\qquad + \left(q_j^n - q_{j-1}^n \right) \Big(\int_{t_n}^{t_{n+1}} \phi(x_{j-1/2}, s)\, ds - \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x, t_{n+1})\, dx \Big) \Big] \\
&= \sum_j \sum_{n=0}^{N-1} \Big[\lambda \left(Q_{j+1/2}^n - q_j^n \right) \Big(\int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x + \Delta x, t_{n+1}) - \phi(x, t_{n+1})\, dx \Big) \\
&\qquad + (q_j^n - q_{j-1}^n) \Big(\int_{t_n}^{t_{n+1}} \phi(x_{j-1/2}, s)\, ds - \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \phi(x, t_{n+1})\, dx \Big) \Big].
\end{aligned}
$$

We also have that

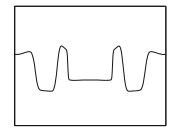
$$
\left| q_j^n - Q_{j+1/2}^n \right| \le \| f \|_{\mathrm{Lip}} \sum_{m=-p}^{p'} \left| u_{j+m}^n - u_j^n \right|
$$

and

$$
\left| q_j^n - q_{j-1}^n \right| \le \| f \|_{\mathrm{Lip}} \left| u_j^n - u_{j-1}^n \right|,
$$

which implies that

$$
\begin{aligned}
-\Lambda_T(u_{\Delta t}, \phi, k) \le \| f \|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \Big[& \Big(\sum_{m=-p}^{p'} \left| u_{j+m}^n - u_j^n \right| \Big) \\
& \times \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \left| \phi(x + \Delta x, t_{n+1}) - \phi(x, t_{n+1}) \right| dx \\
& + \left| u_j^n - u_{j-1}^n \right| \\
& \times \Big| \int_{t_n}^{t_{n+1}} \phi(x_{j-1/2}, s)\, ds - \lambda \int_{x_j}^{x_{j+1}} \phi(x, t_{n+1})\, dx \Big| \Big].
\end{aligned}
$$



Next, we subtract $\phi(x_{j-1/2}, t_{n+1})$ from the integrand in each of the latter two integrals. Since $\Delta t = \lambda \Delta x$, the extra terms cancel, and we obtain

$$
\begin{aligned}
-\Lambda_T(u_{\Delta t}, \phi, k) \le \|f\|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \Bigg[\Big(\sum_{m=-p}^{p'} \left| u_{j+m}^n - u_j^n \right| \Big) \\
\times \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} |\phi(x + \Delta x, t_{n+1}) - \phi(x, t_{n+1})| \, dx \\
+ \left| u_j^n - u_{j-1}^n \right| \Big(\int_{t_n}^{t_{n+1}} \left| \phi(x_{j-1/2}, t) - \phi(x_{j-1/2}, t_{n+1}) \right| \, dt \\
+ \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \left| \phi(x, t_{n+1}) - \phi(x_{j-1/2}, t_{n+1}) \right| \, dx \Big) \Bigg].
\end{aligned} \tag{3.67}
$$

Let $v = v(y,s)$ denote the unique entropy solution of (3.1), and let $k = v(y,s)$. Then

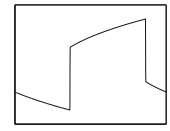
$$
-\Lambda_{\varepsilon_0,\varepsilon}(u,v) = -\int_0^T \int_{\mathbb{R}} \Lambda_T \left(u, v(y,s), \omega_{\varepsilon_0}(\cdot - s) \omega_\varepsilon(\cdot - x) \right) dy \, ds.
$$

Thus to estimate $-\Lambda_{\varepsilon_0,\varepsilon}(u,v)$ we must integrate the terms on the right-hand side of (3.67) in (y,s). To this end,

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}} \int_{x_{j-1/2}}^{x_{j+1/2}} \omega_{\varepsilon_0}(t_{n+1} - s) \left| \omega_\varepsilon(x + \Delta x - y) - \omega_\varepsilon(x - y) \right| \, dx \, dy \, ds \\
&\qquad = \int_{\mathbb{R}} \int_{x_{j-1/2}}^{x_{j+1/2}} \left| \omega_\varepsilon(x + \Delta x - y) - \omega_\varepsilon(x - y) \right| \, dx \, dy \\
&\qquad \le \Delta x^2 \, |\omega_\varepsilon|_{BV} \\
&\qquad \le \frac{2 \Delta x^2}{\varepsilon}.
\end{aligned}
$$

Recalling that $\lambda = \Delta t / \Delta x$, we get

$$
\begin{aligned}
&\int_0^T \int_{\mathbb{R}} \|f\|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \Bigg[\Big(\sum_{m=-p}^{p'} \left| u_{j+m}^n - u_j^n \right| \Big) \\
&\qquad \times \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} |\phi(x + \Delta x, t_{n+1}) - \phi(x, t_{n+1})| \, dx \, dy \, ds \\
&\le \|f\|_{\mathrm{Lip}} \frac{1}{2} (p(p-1) + p'(p'-1)) \sum_{n=0}^{N-1} \sum_j \left| u_j^n - u_{j-1}^n \right| \frac{2 \Delta x^2}{\varepsilon} \lambda \\
&\le C T \frac{\Delta x}{\varepsilon}.
\end{aligned} \tag{3.68}
$$



We also have that

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}} \int_{t_n}^{t_{n+1}} \omega_\varepsilon(x_{j-1/2}-y)\,|\omega_{\varepsilon_0}(t-s)-\omega_{\varepsilon_0}(t_{n+1}-s)|\;dt\,dy\,ds \\
&\qquad = \int_0^T \int_{t_n}^{t_{n+1}} |\omega_{\varepsilon_0}(t-s)-\omega_{\varepsilon_0}(t_{n+1}-s)|\;dt\,ds \\
&\qquad \le \int_{t_n}^{t_{n+1}} \int_t^{t_{n+1}} \int_0^T \left|\omega'_{\varepsilon_0}(\tau-s)\right|\;ds\,d\tau\,dt \\
&\qquad \le \frac{C\,\Delta t^2}{\varepsilon_0}.
\end{aligned}$$

Therefore,

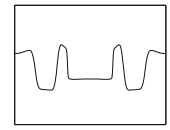
$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}} \|f\|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \left|u_j^n - u_{j-1}^n\right| \int_{t_n}^{t_{n+1}} \left|\phi(x_{j-1/2},t) - \phi(x_{j-1/2},t_{n+1})\right|\;dt\,dy\,ds \\
&\qquad \le \|f\|_{\mathrm{Lip}} \sum_j \left|u_j^0 - u_{j-1}^0\right| \sum_{n=0}^{N-1} \frac{C\,\Delta t^2}{\varepsilon_0} \\
&\qquad \le C\,T\,\frac{\Delta t}{\varepsilon_0}. \qquad (3.69)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}} \int_{t_n}^{t_{n+1}} \omega_{\varepsilon_0}(t_{n+1}-s)\left|\omega_\varepsilon(x-y)-\omega_\varepsilon(x_{j-1/2}-y)\right|\;dx\,dy\,ds \\
&\qquad \le \int_{t_n}^{t_{n+1}} \int_{x_{j-1/2}}^{x} \int_{\mathbb{R}} |\omega'_\varepsilon(z-y)|\;dy\,dz\,dx \\
&\qquad \le \frac{C\,\Delta x \Delta t}{\varepsilon_0},
\end{aligned}$$

and therefore

$$\begin{aligned}
&\|f\|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \int_0^T \int_{\mathbb{R}} \left|u_j^n - u_{j-1}^n\right| \lambda \int_{x_{j-1/2}}^{x_{j+1/2}} \left|\phi(x,t_{n+1}) - \phi(x_{j-1/2},t_{n+1})\right|\;dx\,dy\,ds \\
&\qquad \le \|f\|_{\mathrm{Lip}} \sum_j \sum_{n=0}^{N-1} \left|u_j^0 - u_{j-1}^0\right| \lambda \frac{C\,\Delta x \Delta t}{\varepsilon_0} \\
&\qquad \le C\,T\,\frac{\Delta t}{\varepsilon_0}. \qquad (3.70)
\end{aligned}$$



Collecting the estimates (3.68)–(3.70), we obtain

$$-\Lambda_{\varepsilon_0,\varepsilon}(u,v) \le CT\left(\frac{\Delta x}{\varepsilon}+\frac{\Delta t}{\varepsilon_0}\right), \tag{3.71}$$

where the constant C depends only on f, F, and $|u_0|_{BV}$. Regarding the term $\nu(u,\varepsilon_0)$, we have that $t \mapsto u_{\Delta t}(x,\cdot)$ is "almost" L^1 Lipschitz continuous, so

$$\nu(u_{\Delta t},\varepsilon_0) \le C\left(\max\{\varepsilon_0,\Delta t\}+\Delta t\right).$$

The entropy solution v is of uniformly bounded variation in x for each t. Therefore, we conclude that

$$\begin{aligned}\|u_{\Delta t}(\cdot,T)-v(\cdot,T)\|_{L^1} \le{}& \|u_{\Delta t}(\cdot,0)-v_0\|_1\\ &+CT\left(\max\{\varepsilon_0,\Delta t\}+\varepsilon_0+\varepsilon+\frac{\Delta t}{\varepsilon_0}+\frac{\Delta x}{\varepsilon}\right).\end{aligned}$$

Choosing

$$u_j^0=\frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} v_0(y)\,dy,$$

we have that $\|u_{\Delta t}(\cdot,0)-v_0\|_1 \le \Delta x\,|v_0|_{BV}$. Then we can choose $\varepsilon=\sqrt{\Delta x}$ and $\varepsilon_0=\sqrt{\Delta t}$ to find that

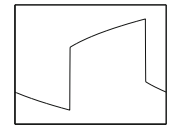
$$\|u_{\Delta t}(\cdot,T)-v(\cdot,T)\|_1 \le C\sqrt{\Delta t}, \tag{3.72}$$

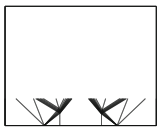
where C depends on T, $|v_0|_{BV}$, f, and F. ◇

If one uses Kuznetsov's lemma to estimate the error of a scheme, one must estimate the modulus of continuity $\tilde{\nu}_t(u,\varepsilon_0)$ and the term $\Lambda_{\varepsilon,\varepsilon_0}(u,v)$. In other words, one must obtain regularity estimates on the *approximation* u. Therefore, this approach gives a posteriori error estimates, and perhaps the proper use for this approach should be in adaptive methods, in which it would provide error control and govern mesh refinement. However, despite this weakness, Kuznetsov's theory is still actively used.

3.4 A Priori Error Estimates

We shall now describe an application of a variation of Kuznetsov's approach in which we obtain an error estimate for the method of vanishing viscosity without using the regularity properties of the viscous approximation. Of course, this application only motivates the approach, since regularity of the solutions of parabolic equations is not difficult to obtain elsewhere. Nevertheless, it is interesting in its own right, since many difference methods have (3.73) as their model equation. We first state the result.





Theorem 3.18 *Let $v(x,t)$ be a solution of (3.1) with initial value v_0, and let u solve the equation*

$$u_t + f(u)_x = (\delta(u)u_x)_x\,, \qquad u(x,0) = u_0(x), \tag{3.73}$$

in the classical sense, with $\delta(u) > 0$. Then

$$\|u(T) - v(T)\|_{L^1(\mathbb{R})} \le 2\|u_0 - v_0\|_{L^1(\mathbb{R})} + 4\text{T.V.}\,(v_0)\,\sqrt{8T\|\delta\|_v},$$

where

$$\|\delta\|_v = \sup_{\substack{t\in[0,T]\\ x\in\mathbb{R}}} \tilde{\delta}\,(v(x-,t), v(x+,t))$$

and

$$\tilde{\delta}(a,b) = \frac{1}{b-a}\int_a^b \delta(c)\,dc.$$

This result is not surprising, and in some sense is weaker than the corresponding result found using Kuznetsov's lemma. The new element here is that the proof does *not* rely on any smoothness properties of the function u, and is therefore also considerably more complicated than the proof using Kuznetsov's lemma.

Proof The proof consists in choosing new Λ's, and using a special form of the test function φ. Let ω^∞ be defined as

$$\omega^\infty(x) = \begin{cases} \frac{1}{2} & \text{for } |x| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We will consider a family of smooth functions ω such that $\omega \to \omega^\infty$. To keep the notation simple we will not add another parameter to the functions ω, but rather write $\omega \to \omega^\infty$ when we approach the limit. Let

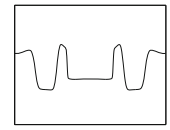
$$\varphi(x,y,t,s) = \omega_\varepsilon(x-y)\omega_{\varepsilon_0}(t-s)$$

with $\omega_\alpha(x) = (1/\alpha)\,\omega(x/\alpha)$ as usual. In this notation,

$$\omega_\varepsilon^\infty(x) = \begin{cases} 1/(2\varepsilon) & \text{for } |x| \le \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

In the following we will use the entropy pair

$$\eta(u,k) = |u-k| \quad \text{and} \quad q(u,k) = \text{sign}\,(u-k)\,(f(u) - f(k)),$$



and except where explicitly stated, we always let $u = u(y,s)$ and $v = v(x,t)$. Let $\eta_\sigma(u,k)$ and $q_\sigma(u,k)$ be smooth approximations to η and q such that

$$\eta_\sigma(u) \to \eta(u) \quad \text{as } \sigma \to 0, \qquad q_\sigma(u,k) = \int \eta'_\sigma(z-k)(f(z)-f(k))\,dz.$$

For a test function φ define

$$\Lambda_T^\sigma(u,k) = \int_0^T \int \eta'_\sigma(u-k)\left(u_s + f(u)_y - \left(\delta(u)u_y\right)_y\right)\varphi\,dy\,ds$$

(which is clearly zero because of (3.73)) and

$$\Lambda_{\varepsilon,\varepsilon_0}^\sigma(u,v) = \int_0^T \int \Lambda_T^\sigma(u,v(x,t))\,dx\,dt.$$

Note that since u satisfies (3.73), $\Lambda_{\varepsilon,\varepsilon_0}^\sigma = 0$ for every v. We now split $\Lambda_{\varepsilon,\varepsilon_0}^\sigma$ into two parts. Writing (cf. (2.15))

$$\begin{aligned}
&\left(u_s + f(u)_x - (\delta(u)u_y)_y\right)\eta'_\sigma(u-k)\\
&\quad = \eta(u-k)_s + ((f(u)-f(k))'\eta'_\sigma(u-k)u_y - (\delta(u)u_y)_y\eta'_\sigma(u-k)\\
&\quad = \eta_\sigma(u-k)_s + q_\sigma(u,k)_u u_y - (\delta(u)u_y)_y\eta'_\sigma(u-k)\\
&\quad = \eta_\sigma(u-k)_s + q_\sigma(u,k)_y - (\delta(u)\eta_\sigma(u-k)_y)_y + \eta''_\sigma(u-k)\delta(u)(u_y)^2\\
&\quad = \eta_\sigma(u-k)_s + (q_\sigma(u,k) - \delta(u)\eta_\sigma(u-k)_y)_y + \eta''(u-k)\delta(u)(u_y)^2,
\end{aligned}$$

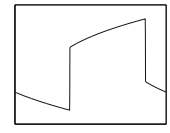
we may introduce

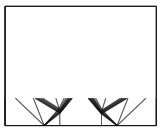
$$\Lambda_1^\sigma(u,v) = \int_0^T\int_0^T\int\int \eta''_\sigma(u-v)\delta(u)\left(u_y\right)^2\varphi\,dy\,ds\,dx\,dt,$$

$$\Lambda_2^\sigma(u,v) = \int_0^T\int_0^T\int\int \left(\eta_\sigma(u-v)_s + \left(q_\sigma(u,v) - \delta(u)\eta_\sigma(u-v)_y\right)_y\right)\varphi\,dy\,ds\,dx\,dt,$$

such that $\Lambda_{\varepsilon,\varepsilon_0}^\sigma = \Lambda_1^\sigma + \Lambda_2^\sigma$. Note that if $\delta(u) > 0$, we always have $\Lambda_1^\sigma \geq 0$, and hence $\Lambda_2^\sigma \leq 0$. Then we have that

$$\Lambda_2 := \limsup_{\sigma\to 0} \Lambda_2^\sigma \leq 0.$$





To estimate Λ_2, we integrate by parts:

$$\begin{aligned}\Lambda_2(u,v) &= \int_0^T\iint_0^T\iint \left(-\eta(u-v)\varphi_s - q(u,v)\varphi_y + V(u,v)\varphi_{yy}\right)\,dy\,ds\,dx\,dt \\ &\quad + \int_0^T\iint \eta(u(T)-v)\varphi|_{s=T}\,dy\,dx\,dt - \int_0^T\iint \eta(u_0-v)\varphi|_{s=0}\,dy\,dx\,dt \\ &= \int_0^T\iint_0^T\iint \left(\eta(u-v)\varphi_t + F(u,v)\varphi_x - V(u,v)\varphi_{xy}\right)\,dy\,ds\,dx\,dt \\ &\quad + \int_0^T\iint \eta(u(T)-v)\varphi|_{s=T}\,dy\,dx\,dt - \int_0^T\iint \eta(u_0-v)\varphi|_{s=0}\,dy\,dx\,dt,\end{aligned}$$

where

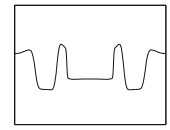
$$V(u,v) = \int_u^v \delta(s)\eta'(s-v)\,ds.$$

Now define (the "dual of Λ_2")

$$\begin{aligned}\Lambda_2^* &:= -\int_0^T\iint_0^T\iint \left(\eta(u-v)\varphi_t + q(u,v)\varphi_x - V(u,v)\varphi_{xy}\right)\,dy\,ds\,dx\,dt \\ &\quad - \int_0^T\iint \eta(u-v(T))\varphi\,\Big|_{t=0}^{t=T}\,dx\,dy\,ds.\end{aligned}$$

Then we can write

$$\begin{aligned}\Lambda_2 = -\Lambda_2^* &+ \underbrace{\int_0^T\iint \left(\eta(u(T)-v)\varphi\right)|_{s=T}\,dy\,dx\,dt}_{\Phi_1} \\ &\underbrace{- \int_0^T\iint \left(\eta(u_0-v)\varphi\right)|_{s=0}\,dy\,dx\,dt}_{\Phi_2} \\ &\underbrace{+ \int_0^T\iint \left(\eta(u-v(T))\varphi\right)|_{t=T}\,dx\,dy\,ds}_{\Phi_3} \\ &\underbrace{- \int_0^T\iint \left(\eta(u_0-v_0)\varphi\right)|_{t=0}\,dx\,dy\,ds}_{\Phi_4} \\ =: -\Lambda_2^* &+ \Phi.\end{aligned}$$



We will need later that

$$\Phi = \Lambda_2^* + \Lambda_2 \le \Lambda_2^*. \tag{3.74}$$

Let

$$\Omega_{\varepsilon_0}(t) = \int_0^t \omega_{\varepsilon_0}(s)\, ds$$

and

$$e(t) = \|u(t) - v(t)\|_{L^1} = \int \eta(u(x,t) - v(x,t))\, dx.$$

To continue estimating, we need the following proposition.

Proposition 3.19

$$\Phi \ge \Omega_{\varepsilon_0}(T)e(T) - \Omega_{\varepsilon_0}(T)e(0) + \int_0^T \omega_{\varepsilon_0}(T-t)e(t)\, dt - \int_0^T \omega_{\varepsilon_0}(t)e(t)\, dt$$
$$- 4\Omega_{\varepsilon_0}(T)\left(\varepsilon_0 \|f\|_{\mathrm{Lip}} + \varepsilon\right) \mathrm{T.V.}\,(v_0)\,.$$

Proof (of Proposition 3.19) We start by estimating Φ_1. First note that

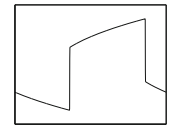
$$\begin{aligned}
\eta(u(y,T) - v(x,t)) &= |u(y,T) - v(x,t)| \\
&\ge |u(y,T) - v(y,T)| \\
&\quad - |v(y,T) - v(y,t)| - |v(y,t) - v(x,t)| \\
&= \eta(u(y,T) - v(y,T)) \\
&\quad - |v(y,T) - v(y,t)| - |v(y,t) - v(x,t)|\,.
\end{aligned}$$

Thus

$$\begin{aligned}
\Phi_1 &\ge \int_0^T \iint \eta(u(y,T) - v(y,T))\varphi|_{s=T}\, dy\, dx\, dt \\
&\quad - \int_0^T \iint |v(y,T) - v(y,t)|\, \varphi|_{s=T}\, dy\, dx\, dt \\
&\quad - \int_0^T \iint |v(y,t) - v(x,t)|\, \varphi|_{s=T}\, dy\, dx\, dt \\
&\ge \Omega_{\varepsilon_0}(T)e(T) - \Omega_{\varepsilon_0}(T)\left(\varepsilon_0 \|f\|_{\mathrm{Lip}} + \varepsilon\right) \mathrm{T.V.}\,(v_0)\,.
\end{aligned}$$

Here we have used that v is an exact solution. The estimate for Φ_2 is similar, yielding

$$\Phi_2 \ge -\Omega_{\varepsilon_0}(T)e(0) - \Omega_{\varepsilon_0}(T)\left(\varepsilon_0 \|f\|_{\mathrm{Lip}} + \varepsilon\right) \mathrm{T.V.}\,(v_0)\,.$$



To estimate Φ_3 we proceed in the same manner:

$$\eta(u(y,s)-v(x,T)) \geq \eta(u(y,s)-v(y,s)) - |v(y,s)-v(x,s)| \\ - |v(x,s)-v(x,T)|\,.$$

This gives

$$\Phi_3 \geq \int_0^T \omega_{\varepsilon_0}(T-t)e(t)\,dt - \Omega_{\varepsilon_0}(T)\left(\varepsilon_0\|f\|_{\mathrm{Lip}}+\varepsilon\right)\mathrm{T.V.}\,(v_0)\,,$$

while by the same reasoning, the estimate for Φ_4 reads

$$\Phi_4 \geq -\int_0^T \omega_{\varepsilon_0}(t)e(t)\,dt - \Omega_{\varepsilon_0}(T)\left(\|f\|_{\mathrm{Lip}}\varepsilon_0+\varepsilon\right)\mathrm{T.V.}\,(v_0)\,.$$

The proof of Proposition 3.19 is complete. □

To proceed further, we shall need the following Gronwall-type lemma:

Lemma 3.20 *Let θ be a nonnegative function that satisfies*

$$\Omega^\infty_{\varepsilon_0}(\tau)\theta(\tau) + \int_0^\tau \omega^\infty_{\varepsilon_0}(\tau-t)\theta(t)\,dt \leq C\,\Omega^\infty_{\varepsilon_0}(\tau) + \int_0^\tau \omega^\infty_{\varepsilon_0}(t)\theta(t)\,dt, \qquad (3.75)$$

for all $\tau \in [0,T]$ and some constant C. Then

$$\theta(\tau) \leq 2C.$$

Proof (of Lemma 3.20) If $\tau \leq \varepsilon_0$, then for $t \in [0,\tau]$, $\omega^\infty_{\varepsilon_0}(t) = \omega^\infty_{\varepsilon_0}(\tau-t) = 1/(2\varepsilon_0)$. In this case (3.75) immediately simplifies to $\theta(t) \leq C$.

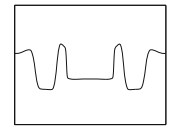
For $\tau > \varepsilon_0$, we can write (3.75) as

$$\theta(\tau) \leq C + \frac{1}{\Omega^\infty_{\varepsilon_0}(\tau)}\int_0^{\varepsilon_0}\left(\omega^\infty_{\varepsilon_0}(t)-\omega^\infty_{\varepsilon_0}(\tau-t)\right)\theta(t)\,dt.$$

For $t \in [0,\varepsilon_0]$ we have $\theta(t) \leq C$, and this implies

$$\theta(\tau) \leq C\left(1+\frac{1}{\Omega^\infty_{\varepsilon_0}(\tau)}\int_0^{\varepsilon_0}\left(\omega^\infty_{\varepsilon_0}(t)-\omega^\infty_{\varepsilon_0}(\tau-t)\right)\,dt\right) \leq 2C.$$

This concludes the proof of the lemma. □



Now we can continue the estimate of $e(T)$.

Proposition 3.21 *We have that*

$$e(T) \leq 2e(0) + 8\left(\varepsilon + \varepsilon_0 \| f \|_{\mathrm{Lip}}\right) \mathrm{T.V.}\,(v_0) + 2 \lim_{\omega \to \omega^\infty} \sup_{t \in [0,T]} \frac{\Lambda_2^*(u,v)}{\Omega_{\varepsilon_0}^\infty(t)}.$$

Proof (of Proposition 3.21) Starting with the inequality (3.74), using the estimate for Φ from Proposition 3.19, we have, after passing to the limit $\omega \to \omega^\infty$, that

$$\begin{aligned}\Omega_{\varepsilon_0}^\infty(T)e(T) + \int_0^T \omega_{\varepsilon_0}^\infty(T-t)e(t)\,dt \leq{}& \Omega_{\varepsilon_0}^\infty(t)e(0) + \int_0^T \omega_{\varepsilon_0}^\infty(t)e(t)\,dt \\ &+ 4\Omega_{\varepsilon_0}^\infty(t)\left(\varepsilon + \varepsilon_0 \| f \|_{\mathrm{Lip}}\right) \mathrm{T.V.}\,(v_0) \\ &+ \Omega_{\varepsilon_0}^\infty(T) \lim_{\omega \to \omega^\infty} \sup_{t \in [0,T]} \frac{\Lambda_2^*(u,v)}{\Omega_{\varepsilon_0}^\infty(t)}.\end{aligned}$$

We apply Lemma 3.20 with

$$C = 4\left(\varepsilon + \varepsilon_0 \| f \|_{\mathrm{Lip}}\right) \mathrm{T.V.}\,(v_0) + \lim_{\omega \to \omega^\infty} \sup_{t \in [0,T]} \frac{\Lambda_2^*(u,v)}{\Omega_{\varepsilon_0}^\infty(t)} + e(0)$$

to complete the proof. □

To finish the proof of the theorem, it remains only to estimate

$$\lim_{\omega \to \omega^\infty} \sup_{t \in [0,T]} \frac{\Lambda_2^*(u,v)}{\Omega(t)}.$$

We will use the following inequality:

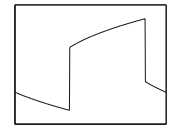
$$\left| \frac{V\left(u, v^+\right) - V\left(u, v^-\right)}{v^+ - v^-} \right| \leq \frac{1}{v^+ - v^-} \int_{v^-}^{v^+} \delta(s)\,ds. \tag{3.76}$$

Since v is an entropy solution to (3.1), we have that

$$\Lambda_2^* \leq -\int_0^T \iint_0^T V(u,v)\varphi_{xy}\,dy\,ds\,dx\,dt. \tag{3.77}$$

Since v is of bounded variation, it suffices to study the case that v is differentiable except on a countable number of curves $x = x(t)$. We shall bound Λ_2^* in the case that we have one such curve; the generalization to more than one is straightforward. Integrating (3.77) by parts, we obtain

$$\Lambda_2^* \leq \int_0^T \int \Psi(y,s)\,dy\,ds, \tag{3.78}$$



where Ψ is given by

$$\Psi(y,s) = \int_0^T \Big(\int_{-\infty}^{x(t)} V(u,v)_v \, v_x \varphi_y \, dx + \frac{[\![V]\!]}{[\![v]\!]} [\![v]\!] \, \varphi_y|_{x=x(t)} + \int_{x(t)}^{\infty} V(u,v)_v \, v_x \varphi_y \, dx \Big) \, dt.$$

As before, $[\![a]\!]$ denotes the jump in a, i.e., $[\![a]\!] = a(x(t)+,t) - a(x(t)-,t)$. Using (3.76), we obtain

$$\begin{aligned} |\Psi(y,s)| \le & \|\delta\|_v \int_0^T \Big(\int_{-\infty}^{x(t)} |v_x| \, |\varphi_y| \, dx \\ & + |[\![v]\!]| \, |\varphi_y|_{x=x(t)}| + \int_{x(t)}^{\infty} |v_x| \, |\varphi_y| \, dx \Big) \, dt. \end{aligned} \tag{3.79}$$

Let D be given by

$$D(x,t) = \int_0^T \int |\varphi_y| \, dy \, ds.$$

A simple calculation shows that

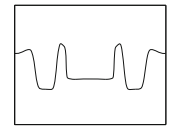
$$D(x,t) = \frac{1}{\varepsilon} \int_0^T \omega_{\varepsilon_0}(t-s) \, ds \int |\omega'(y)| \, dy \le \frac{1}{\varepsilon} \int_0^T \omega_{\varepsilon_0}(t-s) \, ds.$$

Consequently,

$$\begin{aligned} \int_0^T \sup_x D(x,t) \, dt &\le \frac{1}{\varepsilon} \int_0^T \int_0^T \omega_{\varepsilon_0}(t-s) \, ds \, dt \\ &= \frac{2}{\varepsilon} \int_0^T (T-t) \omega_{\varepsilon_0}(t) \, dt \\ &\le \frac{2T\Omega(T)}{\varepsilon}. \end{aligned}$$

Inserting this in (3.79), and the result in (3.78), we find that

$$\Lambda_2^*(u,v,T) \le \frac{2}{\varepsilon} T \, \text{T.V.}(v_0) \, \|\delta\|_v \Omega(T).$$



Summing up, we have now shown that

$$e(T) \leq 2e(0) + 8\left(\varepsilon + \varepsilon_0 \| f \|_{\text{Lip}}\right) \text{T.V.}\,(v_0) + \frac{4}{\varepsilon} T\, \text{T.V.}\,(v_0)\, \|\delta\|_v.$$

We can set ε_0 to zero, and minimize over ε, obtaining

$$\|u(T) - v(T)\|_{L^1} \leq 2\|u_0 - v_0\|_{L^1} + 4\text{T.V.}\,(v_0)\, \sqrt{8T\|\delta\|_v}.$$

The theorem is proved. □

The main idea behind this approach to getting a priori error estimates is to choose the "Kuznetsov-type" form $\Lambda_{\varepsilon,\varepsilon_0}$ such that

$$\Lambda_{\varepsilon,\varepsilon_0}(u, v) = 0$$

for every function v, and then write $\Lambda_{\varepsilon,\varepsilon_0}$ as the sum of a nonnegative and a nonpositive part. Given a numerical scheme, the task is then to prove a discrete analogue of the previous theorem.

3.5 Measure-Valued Solutions

You try so hard, but you don't understand ...
— Bob Dylan, Ballad of a Thin Man (1965)

Monotone methods are at most first-order accurate. Consequently, one must work harder to show that higher-order methods converge to the entropy solution. While this is possible in one space dimension, i.e., in the above setting, it is much more difficult in several space dimensions. One useful tool to aid the analysis of higher-order methods is the concept of *measure-valued solutions*. This is a rather complicated concept, which requires a solid background in analysis beyond this book. Therefore, the presentation in this section is brief, and is intended to give the reader a first flavor, and an idea of what this method can accomplish.

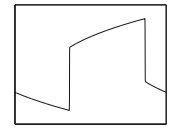
The Young Measure

Consider a sequence $\{u_n\}_{n\in\mathbb{N}}$ that is uniformly bounded in $L^\infty(\mathbb{R} \times [0, \infty))$. This is typically the result of a numerical method, where one has L^∞ bounds, but no uniform bounds on the total variation. Passing to a subsequence, we can still infer that the weak-star limit

$$u_n \overset{*}{\rightharpoonup} u,$$

exists, which means that for all $\varphi \in L^1(\mathbb{R} \times [0, \infty))$,

$$\iint_\Omega u_n \varphi \, dx\, dt \to \iint_\Omega u \varphi \, dx\, dt,$$



with $\Omega = \mathbb{R} \times [0, \infty)$. In order to show that the limit u is a weak solution to the conservation law, we must study

$$\iint_{\Omega} \big(u_n \varphi_t + f(u_n)\varphi_x\big)\, dx\, dt.$$

The first term in this equation has a limit $\iint u\varphi_t\, dx\, dt$, but the second term is more complicated, as the next example shows.

◇ Example 3.22
Let $u_n = \sin(nx)$ and $f(u) = u^2$, and φ a smooth function in $L^1(\mathbb{R})$. Then

$$\left| \int \sin(nx)\varphi(x)\, dx \right| \le \frac{1}{n} \left| \int \cos(nx)\varphi'(x)\, dx \right| \le \frac{C}{n} \to 0 \text{ as } n \to \infty.$$

On the other hand, $f(u_n) = \sin^2(nx) = (1 - \cos(2nx))/2$, and hence a similar estimate shows that

$$\left| \int (f(u_n) - \frac{1}{2})\varphi(x)\, dx \right| \le \frac{C}{n} \to 0 \text{ as } n \to \infty.$$

Thus we conclude that

$$u_n \overset{*}{\rightharpoonup} 0, \quad f(u_n) \overset{*}{\rightharpoonup} \frac{1}{2} \neq 0 = f(0). \qquad \diamond$$

The Young measure is one method for studying the weak limits of nonlinear functions of a weak-star convergent sequence.

In order to define it, we first define the function

$$\chi(\lambda, u) = \begin{cases} 1 & 0 \le \lambda \le u, \\ -1 & u \le \lambda \le 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.80}$$

It is easily verified that for every differentiable function f,

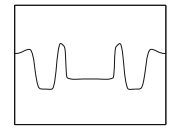
$$\int_{-\infty}^{\infty} f'(\lambda)\chi(\lambda, u)\, d\lambda = f(u) - f(0). \tag{3.81}$$

Furthermore, let $g(\lambda)$ be a function such that

$$u = \int_{\mathbb{R}} g(\lambda)\, d\lambda, \qquad \operatorname{sign}(\lambda)\, g(\lambda) = |g(\lambda)| \le 1. \tag{3.82}$$

Define $m(\lambda)$ by

$$m(\lambda) = \int_{-\infty}^{\lambda} \big(\chi(\xi, u) - g(\xi)\big)\, d\xi.$$



Then $\lim_{\lambda\to-\infty} m(\lambda) = 0$, and

$$\lim_{\lambda\to\infty} m(\lambda) = \int_{-\infty}^{\infty} \chi(\xi,u) - g(\xi)\,d\xi = u - u = 0.$$

Furthermore, by (3.82), we have that m is nondecreasing in the interval $(-\infty,u)$ and nonincreasing in the interval (u,∞). Hence $m(\lambda)$ is nonnegative. For every twice differentiable convex function $S(\lambda)$ we have

$$\int_{\mathbb{R}} S'(\lambda)\,(\chi(\lambda,u) - g(\lambda))\,d\lambda = -\int_{\mathbb{R}} S''(\lambda)m(\lambda)\,d\lambda \le 0.$$

Thus, for a strictly convex function S, the function $\chi(\cdot,u)$ is the unique minimizer of the problem: Find $g \in L^1(\mathbb{R})$ such that (3.82) holds and

$$\int_{\mathbb{R}} S'(\lambda)g(\lambda)\,d\lambda \qquad \text{is minimized.} \tag{3.83}$$

If $\{u_n\}_{n\in\mathbb{N}} \subset L^\infty(\Omega)$ is uniformly bounded, then $\{\chi(\cdot,u_n)\}_{n\in\mathbb{N}} \subset L^\infty(\mathbb{R}\times\Omega)$ is also uniformly bounded. Thus it has (modulo subsequences) a weak-star limit, which we call $f(\lambda,x,t)$. The next lemma gives some properties of this limit.

Lemma 3.23 *Let $f(\lambda,x,t)$ denote the weak-star limit of $\chi(\lambda,u_n)$. Then f is in $L^\infty(\mathbb{R}\times\Omega)$ and satisfies*

$$\int_{\mathbb{R}} f(\lambda,x,t)\,d\lambda = u(x,t) \tag{3.84}$$

for almost all (x,t). Furthermore,

$$|f(\lambda,x,t)| = \operatorname{sign}(\lambda)\,f(\lambda,x,t), \tag{3.85}$$

$$\frac{\partial}{\partial\lambda} f(\lambda,x,t) = \delta(\lambda) - \nu_{(x,t)}(\lambda), \tag{3.86}$$

where $\delta(\lambda)$ is the Dirac measure, and $\nu_{(x,t)}(\lambda)$ is a nonnegative measure in (λ,x,t) such that

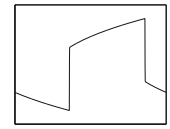
$$\int_{\mathbb{R}} \nu_{(x,t)}(\lambda)\,d\lambda = 1 \tag{3.87}$$

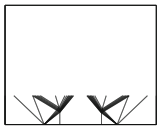
for almost all (x,t).

Remark 3.24 The derivative in (3.86) is to be interpreted in the distributional sense, i.e., (3.86) means that

$$\begin{aligned} -\int_{\mathbb{R}} f(\lambda,x,t)\varphi'(\lambda)\,d\lambda &= \int_{\mathbb{R}} \frac{\partial}{\partial\lambda} f(\lambda,x,t)\varphi(\lambda)\,d\lambda \\ &= \int_{\mathbb{R}} \big(\delta(\lambda) - \nu_{(x,t)}(\lambda)\big)\,\varphi(\lambda)\,d\lambda, \end{aligned}$$

for all $\varphi \in C_0^\infty(\mathbb{R})$.





Proof The first equality, (3.84) follows from the observation

$$u_n(x,t) = \int_{\mathbb{R}} \chi(\lambda, u_n(x,t))\, d\lambda.$$

To prove (3.85) we choose a test function of the form $\varphi(x,t)\psi(\lambda)$, where the ψ has support in $(0,\infty)$ and $\varphi \geq 0$. By definition of the weak-star limit,

$$\iint_{\Omega} \int_{\mathbb{R}} f(\lambda,x,t)\psi(\lambda)\varphi(x,t)\, d\lambda\, dx\, dt$$
$$= \lim_{n\to\infty} \iint_{\Omega} \int_{\mathbb{R}} \chi(\lambda, u_n(x,t))\psi(\lambda)\varphi(x,t)\, d\lambda\, dx\, dt \geq 0.$$

Thus $f \geq 0$ for $\lambda \geq 0$, and one similarly shows that $f \leq 0$ if $\lambda \leq 0$.

To prove (3.86), by Remark 3.24 we have that for all test functions $\varphi(\lambda,x,t)$,

$$\begin{aligned}
&\iint_{\Omega} \int_{\mathbb{R}} \frac{\partial}{\partial\lambda}\chi(\lambda,u_n)\varphi(\lambda,x,t)\, d\lambda\, dx\, dt \\
&\qquad = -\iint_{\Omega} \int_{\mathbb{R}} \chi(\lambda,u_n)\frac{\partial}{\partial\lambda}\varphi(\lambda,x,t)\, d\lambda\, dx\, dt \\
&\qquad = \iint_{\Omega} \big(\varphi(0,x,t) - \varphi(u_n,x,t)\big)\, dx\, dt \\
&\qquad = \iint_{\Omega} \int_{\mathbb{R}} \Big(\delta(\lambda)\varphi(\lambda,x,t) - \delta_{u_n}(\lambda)\varphi(\lambda,x,t)\Big)\, d\lambda\, dx\, dt,
\end{aligned}$$

where δ_{u_n} is the Dirac mass centered at u_n. Thus we define

$$\nu_{n,(x,t)}(\lambda) = \delta_{u_n}(\lambda),$$

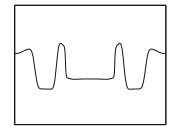
so that

$$\frac{\partial}{\partial\lambda}\chi(\lambda,u_n(x,t)) = \delta(\lambda) - \nu_{n,(x,t)}(\lambda).$$

The measure $\nu_{n,(x,t)}$ is a probability measure in the first variable, in the sense that it is nonnegative and has unit total mass. Thus we have that there exists a nonnegative measure $\nu_{(x,t)}$ such that

$$\int_{\mathbb{R}} \nu_{n,(x,t)}(\lambda)\psi(\lambda)\, d\lambda \to \int_{\mathbb{R}} \psi(\lambda)\, \nu_{(x,t)}(\lambda)\, d\lambda,$$

for all continuous functions ψ. In order to conclude, we must prove (3.87). Choose a test function of the form $\psi(\lambda)\varphi(x,t)$, where ψ has compact support and $\psi \equiv 1$



for $|\lambda| \le \|u_n\|_\infty$. Then

$$\begin{aligned} 0 &= -\iint_\Omega \int_{\mathbb{R}} \chi(\lambda, u_n)\psi'(\lambda)\varphi(x,t)\,d\lambda\,dx\,dt \\ &= \iint_\Omega \left(1 - \int_{\mathbb{R}} \nu_{n,(x,t)}(\lambda)\,d\lambda\right)\varphi(x,t)\,dx\,dt \\ &\to \iint_\Omega \left(1 - \int_{\mathbb{R}} \nu_{(x,t)}(\lambda)\,d\lambda\right)\varphi(x,t)\,dx\,dt \quad \text{as } n \to \infty. \end{aligned}$$

Thus (3.87) holds. □

If now $u_n \overset{*}{\rightharpoonup} u$ in L^∞, then we have

$$u_n(x,t) = \int_{\mathbb{R}} \chi(\lambda, u_n(x,t))\,d\lambda \to \int_{\mathbb{R}} f(\lambda,x,t)\,d\lambda = u(x,t).$$

Similarly, for every function $S(u)$ with S' bounded and $S(0) = 0$,

$$S(u_n) = \int_{\mathbb{R}} S'(\lambda)\chi(\lambda,u_n)\,d\lambda = \int_{\mathbb{R}} S(\lambda)\nu_{n,(x,t)}(\lambda)\,d\lambda.$$

Therefore, if $\bar{S}(x,t)$ denotes the weak-star limit of $S(u_n)$, then

$$\bar{S}(x,t) = \int_{\mathbb{R}} S'(\lambda) f(\lambda,x,t)\,d\lambda = \int_{\mathbb{R}} S(\lambda)\nu_{(x,t)}(\lambda)\,d\lambda. \tag{3.88}$$

The limit measure $\nu_{(x,t)}$ is called the Young measure associated with the sequence $\{u_n\}$. If S is strictly convex, then using (3.83), we obtain

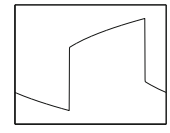
$$\bar{S}(x,t) = \int_{\mathbb{R}} S'(\lambda) f(\lambda,x,t)\,d\lambda \le \int_{\mathbb{R}} S'(\lambda)\chi(\lambda,u)\,d\lambda = S(u),$$

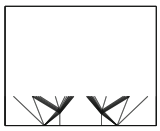
with equality if and only if $f(\lambda,x,t) = \chi(\lambda,u(x,t))$. Hence, $u_n \to u$ strongly, if and only if $\nu_{(x,t)}(\lambda) = \delta_u(\lambda)$.

We have proved the following theorem:

Theorem 3.25 (Young's theorem) *Let $\{u_n\}$ be a sequence of functions from $\Omega = \mathbb{R}\times[0,\infty)$ with values in $[-K,K]$. Then there exists a family of probability measures $\{\nu_{(x,t)}(\lambda)\}_{(x,t)\in\Omega}$, depending weak-star measurably on (x,t), such that for every continuously differentiable function $S\colon[-K,K]\to\mathbb{R}$ with S' bounded and $S(0)=0$, we have*

$$S(u_n) \overset{*}{\rightharpoonup} \bar{S} \text{ in } L^\infty(\Omega) \text{ as } n \to \infty,$$





where

$$\bar{S}(x,t) = \int_{\mathbb{R}} S(\lambda)\, d\nu_{(x,t)}(\lambda) \text{ for a.e. } (x,t) \in \Omega,$$

and where the exceptional set possibly depends on S. Furthermore,

$$\text{supp } \nu_{(x,t)} \subset [-K, K] \text{ for a.e. } (x,t) \in \Omega.$$

We also have that $u_n \to u$ strongly in $L^1_{\text{loc}}(\Omega)$ if and only if $\nu_{(x,t)}(\lambda) = \delta_{u(x,t)}(\lambda)$.

◇ Example 3.26
Let us compute the Young measure associated with the sequence $\{\sin(nx)\}$. In this case the weak limit of $\chi(\lambda, \sin(nx))$ will be independent of x. If $\lambda > 0$, then

$$\int_a^b \chi(\lambda, \sin(nx))\, dx = \frac{\text{meas}\,\{x \in [a,b] \mid \sin(nx) > \lambda\}}{b-a},$$

and similarly, if $\lambda < 0$, then

$$\int_a^b \chi(\lambda, \sin(nx))\, dx = -\frac{\text{meas}\,\{x \in [a,b] \mid \sin(nx) < \lambda\}}{b-a}.$$

We have $\chi(\lambda, \sin(nx)) \overset{*}{\rightharpoonup} f(\lambda)$, where

$$f(\lambda) = \frac{1}{2\pi} \begin{cases} 2(\frac{\pi}{2} - \sin^{-1}(\lambda)) & 0 < \lambda \le 1, \\ -2(\frac{\pi}{2} + \sin^{-1}(\lambda)) & -1 \le \lambda \le 0, \\ 0 & \text{otherwise.} \end{cases}$$

This can be rewritten

$$f(\lambda) = \chi_{[-1,1]}(\lambda) \left(\frac{1}{2} \text{sign}\,(\lambda) - \frac{1}{\pi} \sin^{-1}(\lambda) \right).$$

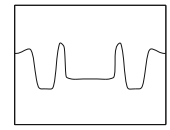
Thus from (3.86),

$$f'(\lambda) = \delta(\lambda) - \nu_x(\lambda) = \delta(\lambda) - \chi_{[-1,1]}(\lambda) \frac{1}{\pi\sqrt{1-\lambda^2}},$$

and we see that

$$\nu_x(\lambda) = \frac{\chi_{[-1,1]}(\lambda)}{\pi\sqrt{1-\lambda^2}}. \qquad \diamond$$

Theorem 3.25 is indeed the main reason why measure-valued solutions are easier to obtain than weak solutions, since for every bounded sequence of approximations to a solution of a conservation law we can associate (at least) one probability measure $\nu_{(x,t)}$ representing the weak-star limits of the sequence. Thus we avoid having to show that the method is TVD stable and use Helly's theorem to be able to work with the limit of the sequence. The measures associated with weakly convergent sequences are frequently called Young measures.



Intuitively, when we are in the situation that we have no knowledge of eventual oscillations in u_ε as $\varepsilon \to 0$, the Young measure $\nu_{(x,t)}(E)$ can be thought of as the probability that the "limit" at the point (x,t) takes a value in the set E. To be a bit more precise, define

$$\nu^{\varepsilon,r}_{(x,t)}(E) = \frac{1}{r^2}\,\mathrm{meas}\Big\{(y,s) \;\Big|\; |x-y|, |t-s| \le r \quad \text{and} \quad u_\varepsilon(y,s) \in E\Big\}.$$

Then for small r, $\nu^{\varepsilon,r}_{(x,t)}(E)$ is the probability that u^ε takes values in E near x. It can be shown that

$$\nu_{(x,t)} = \lim_{r\to 0}\lim_{\varepsilon\to 0} \nu^{\varepsilon,r}_{(x,t)};$$

see [10].

Measure-Valued Solutions

Now we can define measure-valued solutions. We use the notation

$$\langle \nu_{(x,t)}, g\rangle = \int_{\mathbb{R}} g(\lambda)\, d\nu_{(x,t)}(\lambda).$$

A probability measure $\nu_{(x,t)}$ is a measure-valued solution to (3.1) if

$$\big\langle \nu_{(x,t)}, \mathrm{Id}\big\rangle_t + \big\langle \nu_{(x,t)}, f\big\rangle_x = 0$$

in the distributional sense, where Id is the identity map, $\mathrm{Id}(\lambda) = \lambda$. As with weak solutions, we call a measure-valued solution compatible with the entropy pair (η, q) (recall that $q' = \eta' f'$) if

$$\big\langle \nu_{(x,t)}, \eta\big\rangle_t + \big\langle \nu_{(x,t)}, q\big\rangle_x \le 0 \tag{3.89}$$

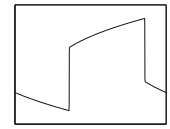
in the distributional sense. If (3.89) holds for *all* convex η, we call $\nu_{(x,t)}$ a measure-valued entropy solution. Clearly, weak entropy solutions are also measure-valued solutions, as we can see by setting

$$\nu_{(x,t)} = \delta_{u(x,t)}$$

for a weak entropy solution u. But measure-valued solutions are more general than weak solutions, since for every two measure-valued solutions $\nu_{(x,t)}$ and $\mu_{(x,t)}$ and $\theta \in [0,1]$, the convex combination

$$\theta\nu_{(x,t)} + (1-\theta)\mu_{(x,t)} \tag{3.90}$$

is also a measure-valued solution. It is not clear, however, what are the initial data satisfied by the measure-valued solution defined by (3.90). We would like our



measure-valued solutions initially to be Dirac masses, i.e., $\nu_{(x,0)} = \delta_{u_0(x)}$. Concretely, we shall assume the following:

$$\lim_{T\downarrow 0} \frac{1}{T} \int_0^T \int_{-A}^{A} \left\langle \nu_{(x,t)}, |\mathrm{Id} - u_0(x)| \right\rangle \, dx \, dt = 0 \tag{3.91}$$

for every A. For every Young measure $\nu_{(x,t)}$ we have the following lemma.

Lemma 3.27 *Let $\nu_{(x,t)}$ be a Young measure with $\operatorname{supp} \nu_{(x,t)} \subset [-K, K]$, and let ω_ε be a standard mollifier in x and t. Then*:

(i) *there exists a Young measure $\nu^\varepsilon_{(x,t)}$ defined by*

$$\begin{aligned} \left\langle \nu^\varepsilon_{(x,t)}, g \right\rangle &= \left\langle \nu_{(x,t)}, g \right\rangle * \omega_\varepsilon \\ &= \iint \omega_\varepsilon(x-y)\omega_\varepsilon(t-s) \left\langle \nu_{(y,s)}, g \right\rangle \, dy \, ds. \end{aligned} \tag{3.92}$$

(ii) *For all $(x,t) \in \mathbb{R} \times [0,T]$ there exist bounded measures $\partial_x \nu^\varepsilon_{(x,t)}$ and $\partial_t \nu^\varepsilon_{(x,t)}$, defined by*

$$\begin{aligned} \left\langle \partial_t \nu^\varepsilon_{(x,t)}, g \right\rangle &= \partial_t \left\langle \nu^\varepsilon_{(x,t)}, g \right\rangle, \\ \left\langle \partial_x \nu^\varepsilon_{(x,t)}, g \right\rangle &= \partial_x \left\langle \nu^\varepsilon_{(x,t)}, g \right\rangle. \end{aligned} \tag{3.93}$$

Proof Clearly, the right-hand side of (3.92) is a bounded linear functional on $C_0(\mathbb{R})$, the set of compactly supported continuous functions, and hence the Riesz representation theorem guarantees the existence of $\nu^\varepsilon_{(x,t)}$. To show that $\|\nu^\varepsilon_{(x,t)}\|_{\mathcal{M}(\mathbb{R})} = 1$, where $\mathcal{M}(\mathbb{R})$ is the set of all Radon measures, we let $\{\psi_n\}$ be a sequence of test functions such that

$$\left\langle \nu_{(x,t)}, \psi_n \right\rangle \to 1, \quad \text{as } n \to \infty.$$

Then for all $1 > \kappa > 0$ we can find an N such that

$$\left\langle \nu_{(x,t)}, \psi_n \right\rangle > 1 - \kappa,$$

for $n \geq N$. Thus, for such n,

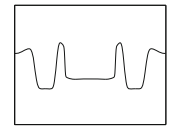
$$\left\langle \nu^\varepsilon_{(x,t)}, \psi_n \right\rangle \geq 1 - \kappa,$$

and therefore $\|\nu^\varepsilon_{(x,t)}\|_{\mathcal{M}(\mathbb{R})} \geq 1$. The opposite inequality is immediate, since

$$\left| \left\langle \nu^\varepsilon_{(x,t)}, \psi \right\rangle \right| \leq \left| \left\langle \nu_{(x,t)}, \psi \right\rangle \right|$$

for all test functions ψ. Therefore, $\nu^\varepsilon_{(x,t)}$ is a probability measure. Similarly, the existence of $\partial_x \nu^\varepsilon_{(x,t)}$ and $\partial_t \nu^\varepsilon_{(x,t)}$ follows by the Riesz representation theorem. Since $\nu_{(x,t)}$ is bounded, the boundedness of $\partial_x \nu^\varepsilon_{(x,t)}$ and $\partial_t \nu^\varepsilon_{(x,t)}$ follows for each fixed $\varepsilon > 0$. □

Now that we have established the existence of the "smooth approximation" to a Young measure, we can use this to prove the following lemma.



Lemma 3.28 *Assume that f is a Lipschitz continuous function and that $\nu_{(x,t)}(\lambda)$ and $\sigma_{(x,t)}(\mu)$ are measure-valued solutions with support in $[-K, K]$. Then*

$$\partial_t \left\langle \nu_{(x,t)} \otimes \sigma_{(x,t)}, |\lambda - \mu| \right\rangle + \partial_x \left\langle \nu_{(x,t)} \otimes \sigma_{(x,t)}, q(\lambda, \mu) \right\rangle \leq 0, \tag{3.94}$$

in the distributional sense, where

$$q(\lambda, \mu) = \operatorname{sign}(\lambda - \mu)\,(f(\lambda) - f(\mu)),$$

and $\nu_{(x,t)} \otimes \sigma_{(x,t)}$ denotes the product measure $d\nu_{(x,t)} d\sigma_{(x,t)}$ on $\mathbb{R} \times \mathbb{R}$.

Proof If $\nu^\varepsilon_{(x,t)}$ and $\sigma^\varepsilon_{(x,t)}$ are defined by (3.92), and $\varphi \in C_0^\infty(\mathbb{R} \times [0, T])$, then we have that

$$\begin{aligned} \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu_{(x,t)}, g \right\rangle \partial_t \left(\varphi * \omega_\varepsilon\right) dx\, dt &= \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu^\varepsilon_{(x,t)}, g \right\rangle \partial_t \varphi \, dx\, dt \\ &= - \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \partial_t \nu^\varepsilon_{(x,t)}, g \right\rangle \varphi \, dx\, dt, \end{aligned}$$

and similarly,

$$\iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu_{(x,t)}, g \right\rangle \partial_x \left(\varphi * \omega_\varepsilon\right) dx\, dt = - \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \partial_x \nu^\varepsilon_{(x,t)}, g \right\rangle \varphi \, dx\, dt,$$

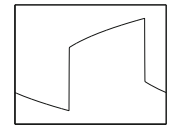
and analogous identities also hold for $\sigma_{(x,t)}$. Therefore,

$$\left\langle \partial_t \nu^\varepsilon_{(x,t)}, |\lambda - \mu| \right\rangle + \left\langle \partial_x \nu^\varepsilon_{(x,t)}, q(\lambda, \mu) \right\rangle \leq 0, \tag{3.95}$$

$$\left\langle \partial_t \sigma^\varepsilon_{(x,t)}, |\lambda - \mu| \right\rangle + \left\langle \partial_x \sigma^\varepsilon_{(x,t)}, q(\lambda, \mu) \right\rangle \leq 0. \tag{3.96}$$

Next, we observe that for every continuous function g,

$$\begin{aligned} \partial_t \left\langle \nu^\varepsilon_{(x,t)} \otimes \sigma^\varepsilon_{(x,t)}, g(\lambda, \mu) \right\rangle &= \int\limits_{\mathbb{R}} \partial_t \left(\int\limits_{\mathbb{R}} g(\lambda, \mu)\, d\nu^\varepsilon_{(x,t)}(\lambda) \right) d\sigma^\varepsilon_{(x,t)}(\mu) \\ &\quad + \int\limits_{\mathbb{R}} \partial_t \left(\int\limits_{\mathbb{R}} g(\lambda, \mu)\, d\sigma^\varepsilon_{(x,t)}(\mu) \right) d\nu^\varepsilon_{(x,t)}(\lambda) \\ &= \int\limits_{\mathbb{R}} \left\langle \partial_t \nu^\varepsilon_{(x,t)}, g(\lambda, \mu) \right\rangle d\sigma^\varepsilon_{(x,t)}(\mu) \\ &\quad + \int\limits_{\mathbb{R}} \left\langle \partial_t \sigma^\varepsilon_{(x,t)}, g(\lambda, \mu) \right\rangle d\nu^\varepsilon_{(x,t)}(\lambda), \end{aligned}$$



and an analogous equality holds for

$$\partial_x \left\langle \nu^{\varepsilon}_{(x,t)} \otimes \sigma^{\varepsilon}_{(x,t)}, g(\lambda,\mu) \right\rangle.$$

Therefore, we find that

$$\begin{aligned}
&\iint\limits_{\mathbb{R}\times[0,T]} \left[\left\langle \nu^{\varepsilon_1}_{(x,t)} \otimes \sigma^{\varepsilon_2}_{(x,t)}, |\lambda-\mu| \right\rangle \varphi_t + \left\langle \nu^{\varepsilon_1}_{(x,t)} \otimes \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle \varphi_x(x,t) \right] dx\,dt \\
&\quad = -\iint\limits_{\mathbb{R}\times[0,T]} \Big(\int\limits_{\mathbb{R}} \left\langle \partial_t \nu^{\varepsilon_1}_{(x,t)}, |\lambda-\mu| \right\rangle + \left\langle \partial_x \nu^{\varepsilon_1}_{(x,t)}, q(\lambda,\mu) \right\rangle d\sigma^{\varepsilon_2}_{(x,t)}(\mu) \Big) \varphi\, dx\,dt \\
&\qquad - \iint\limits_{\mathbb{R}\times[0,T]} \Big(\int\limits_{\mathbb{R}} \left\langle \partial_t \sigma^{\varepsilon_2}_{(x,t)}, |\lambda-\mu| \right\rangle + \left\langle \partial_x \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle d\nu^{\varepsilon_1}_{(x,t)}(\lambda) \Big) \varphi\, dx\,dt \\
&\quad \geq 0,
\end{aligned}$$

for every nonnegative test function φ. Now we would like to conclude the proof by sending ε_1 and ε_2 to zero. Consider the second term:

$$\begin{aligned}
I^{\varepsilon_1,\varepsilon_2} &= \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu^{\varepsilon_1}_{(x,t)} \otimes \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle \varphi_x(x,t)\, dx\,dt \\
&= \iint\limits_{\mathbb{R}\times[0,T]} \iiint \left\langle \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle d\nu_{(y,s)} \\
&\qquad \times \omega_{\varepsilon_1}(x-y)\omega_{\varepsilon_1}(t-s)\varphi_x(x,t)\, dy\,ds\,dx\,dt.
\end{aligned}$$

Since

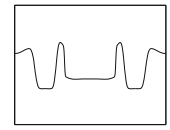
$$\begin{aligned}
&\iiint \left\langle \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle d\nu_{(y,s)} \omega_{\varepsilon_1}(x-y)\omega_{\varepsilon_1}(t-s)\varphi_x(x,t)\, dy\,ds \\
&\qquad \to \int \left\langle \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle d\nu_{(x,t)} \varphi_x(x,t) < \infty
\end{aligned}$$

for almost all (x,t) as $\varepsilon_1 \to 0$, we can use the Lebesgue dominated convergence theorem to conclude that

$$\lim_{\varepsilon_1\to 0} I^{\varepsilon_1,\varepsilon_2} = \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu_{(x,t)} \otimes \sigma^{\varepsilon_2}_{(x,t)}, q(\lambda,\mu) \right\rangle \varphi_x(x,t)\, dx\,dt.$$

We can apply this argument once more for ε_2, obtaining

$$\lim_{\varepsilon_2\to 0} \lim_{\varepsilon_1\to 0} I^{\varepsilon_1,\varepsilon_2} = \iint\limits_{\mathbb{R}\times[0,T]} \left\langle \nu_{(x,t)} \otimes \sigma_{(x,t)}, q(\lambda,\mu) \right\rangle \varphi_x(x,t)\, dx\,dt. \tag{3.97}$$



Similarly, we obtain

$$\lim_{\varepsilon_2\to 0}\lim_{\varepsilon_1\to 0}\iint\limits_{\mathbb{R}\times[0,T]}\left\langle \nu^{\varepsilon_1}_{(x,t)}\otimes\sigma^{\varepsilon_2}_{(x,t)},|\lambda-\mu|\right\rangle\varphi_t(x,t)\,dx\,dt = \iint\limits_{\mathbb{R}\times[0,T]}\left\langle \nu_{(x,t)}\otimes\sigma_{(x,t)},|\lambda-\mu|\right\rangle\varphi_t(x,t)\,dx\,dt. \tag{3.98}$$

This concludes the proof of the lemma. □

Let $\{u_\varepsilon\}$ and $\{v_\varepsilon\}$ be the sequences associated with $\nu_{(x,t)}$ and $\sigma_{(x,t)}$, respectively, and assume that for $t \le T$, the support of $u_\varepsilon(\cdot,t)$ and $v_\varepsilon(\cdot,t)$ is contained in a finite interval I. Then both $u_\varepsilon(\cdot,t)$ and $v_\varepsilon(\cdot,t)$ are in $L^1(\mathbb{R})$ uniformly in ε. This means that both

$$\left\langle \nu_{(x,t)},|\lambda|\right\rangle \quad\text{and}\quad \left\langle \sigma_{(x,t)},|\lambda|\right\rangle$$

are in $L^1(\mathbb{R})$ for almost all t. Using this observation and the preceding lemma, Lemma 3.28, we can continue. Define a smooth approximation to the characteristic function of $[t_1,t_2]$ by

$$\phi_\varepsilon(t)=\int_0^t\big(\omega_\varepsilon(s-t_1)-\omega_\varepsilon(s-t_2)\big)\,ds,$$

where $t_2>t_1>0$ and ω_ε is the usual mollifier. Also define

$$\psi_n(x)=\begin{cases}1 & \text{for } |x|\le n,\\ 2(1-x/(2n)) & \text{for } n<|x|\le 2n,\\ 0 & \text{otherwise,}\end{cases}$$

and set $\psi_{\varepsilon,n}=\psi_n*\omega_\varepsilon(x)$. Hence

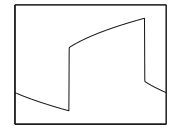
$$\varphi(x,t)=\phi_\varepsilon(t)\psi_{\varepsilon,n}(x)$$

is an admissible test function. Furthermore, $\left|\psi'_{\varepsilon,n}\right|\le 1/n$, and $\phi_\varepsilon(t)$ tends to the characteristic function of the interval $[t_1,t_2]$ as $\varepsilon\to 0$. Therefore,

$$-\lim_{\varepsilon\to 0}\iint\limits_{\mathbb{R}\times[0,T]}\Big[\left\langle \nu_{(x,t)}\otimes\sigma_{(x,t)},|\lambda-\mu|\right\rangle\varphi_t + \left\langle \nu_{(x,t)}\otimes\sigma_{(x,t)},q(\lambda,\mu)\right\rangle\varphi_x\Big]\,dx\,dt\le 0.$$

Set

$$A_n(t)=\int_{\mathbb{R}}\left\langle \nu_{(x,t)}\otimes\sigma_{(x,t)},|\lambda-\mu|\right\rangle\psi_n(x)\,dx.$$



Using this definition, we find that

$$A_n(t_2) - A_n(t_1) \le \int_{t_1}^{t_2} \int_{\mathbb{R}} \left\langle \nu_{(x,t)} \otimes \sigma_{(x,t)}, |\lambda - \mu| \right\rangle |\psi_n'(x)| \, dx \, dt. \tag{3.99}$$

The right-hand side of this is bounded by

$$\|f\|_{\mathrm{Lip}} \frac{1}{n} \left(\left\| \left\langle \nu_{(x,t)}, |\lambda| \right\rangle \right\|_{L^1(\mathbb{R})} + \left\| \left\langle \sigma_{(x,t)}, |\mu| \right\rangle \right\|_{L^1(\mathbb{R})} \right) \to 0$$

as $n \to \infty$. Since $\nu_{(x,t)}$ and $\sigma_{(x,t)}$ are probability measures, for almost all t, the set

$$\left\{ x \mid \langle \nu_{(x,t)}, 1 \rangle \ne 1 \text{ and } \langle \sigma_{(x,t)}, 1 \rangle \ne 1 \right\}$$

has zero Lebesgue measure. Therefore, for almost all t,

$$\begin{aligned} A_n(t) &\le \int_{\mathbb{R}} \left\langle \nu_{(x,t)} \otimes \sigma_{(x,t)}, |\lambda - u_0(x)| + |\mu - u_0(x)| \right\rangle dx \\ &= \int_{\mathbb{R}} \left\langle \nu_{(x,t)}, |\lambda - u_0(x)| \right\rangle dx + \int_{\mathbb{R}} \left\langle \sigma_{(x,t)}, |\mu - u_0(x)| \right\rangle dx. \end{aligned}$$

Integrating (3.99) with respect to t_1 from 0 to T, then dividing by T and sending T to 0, using (3.91), and finally sending $n \to \infty$, we find that

$$\iint_{\mathbb{R}\times\mathbb{R}} |\lambda - \mu| \, d\nu_{(x,t)} \, d\sigma_{(x,t)} = 0, \quad \text{for } (x,t) \notin E, \tag{3.100}$$

where the Lebesgue measure of the (exceptional) set E is zero. Suppose now that for $(x,t) \notin E$ there is a $\bar{\lambda}$ in the support of $\nu_{(x,t)}$ and a $\bar{\mu}$ in the support of $\sigma_{(x,t)}$ and $\bar{\lambda} \ne \bar{\mu}$. Then we can find positive functions g and h such that

$$0 \le g \le 1, \quad 0 \le h \le 1,$$

and

$$\bar{\lambda} \in \operatorname{supp}(g), \quad \bar{\mu} \in \operatorname{supp}(h), \quad \operatorname{supp}(g) \cap \operatorname{supp}(h) = \emptyset.$$

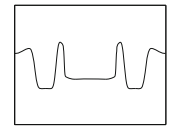
Furthermore,

$$\left\langle \nu_{(x,t)}, g \right\rangle > 0 \quad \text{and} \quad \left\langle \sigma_{(x,t)}, h \right\rangle > 0.$$

Thus

$$\begin{aligned} 0 &< \iint_{\mathbb{R}\times\mathbb{R}} g(\lambda) h(\mu) \, d\nu_{(x,t)} \, d\sigma_{(x,t)} \\ &\le \sup_{\lambda,\mu} \left| \frac{g(\lambda)h(\mu)}{\lambda - \mu} \right| \iint_{\mathbb{R}\times\mathbb{R}} |\lambda - \mu| \, d\nu_{(x,t)} \, d\sigma_{(x,t)} = 0. \end{aligned}$$

This contradiction shows that both $\nu_{(x,t)}$ and $\sigma_{(x,t)}$ are unit point measures with support at a common point. Precisely, we have proved the following theorem:



Theorem 3.29 *Let $u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$.*

(i) *Suppose that $\nu_{(x,t)}$ is a measure-valued entropy solution to the conservation law*

$$u_t + f(u)_x = 0$$

such that $\nu_{(x,t)}$ satisfies the initial condition (3.91), and that $\langle \nu_{(x,t)}, |\lambda| \rangle$ is in $L^\infty([0,T]; L^1(\mathbb{R}))$. Then there exists a function $u \in L^\infty([0,T]; L^1(\mathbb{R})) \cap L^\infty(\mathbb{R} \times [0,T])$ such that

$$\nu_{(x,t)} = \delta_{u(x,t)}, \quad \text{for almost all } (x,t).$$

(ii) *Assume that $\sigma_{(x,t)}$ is (another) measure-valued entropy solution satisfying the same regularity assumptions as $\nu_{(x,t)}$. Then*

$$\nu_{(x,t)} = \sigma_{(x,t)} = \delta_{u(x,t)}, \quad \text{for almost all } (x,t).$$

In order to avoid checking (3.91) directly, we can use the following lemma.

Lemma 3.30 *Let $\nu_{(x,t)}$ be a probability measure, and assume that for all test functions $\varphi(x)$ we have*

$$\lim_{\tau \to 0^+} \frac{1}{\tau} \int_0^\tau \int \langle \nu_{(x,t)}, \mathrm{Id} \rangle \varphi(x)\, dx\, dt = \int u_0(x) \varphi(x)\, dx, \tag{3.101}$$

and that for all nonnegative $\varphi(x)$ and for at least one strictly convex continuous function η,

$$\limsup_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int \langle \nu_{(x,t)}, \eta \rangle \varphi(x)\, dx\, dt \le \int \eta\,(u_0(x))\, \varphi(x)\, dx. \tag{3.102}$$

Then (3.91) holds.

Proof We shall prove

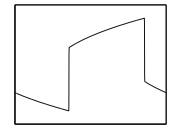
$$\lim_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{-A}^{A} \langle \nu_{(x,t)}, \big(\mathrm{Id} - u_0(x)\big)^+ \rangle\, dx\, dt = 0, \tag{3.103}$$

from which the desired result will follow from (3.101) and the identity

$$|\lambda - u_0(x)| = 2\big(\lambda - u_0(x)\big)^+ - \big(\lambda - u_0(x)\big),$$

where $a^+ = \max\{a, 0\}$ denotes the positive part of a. To get started, we write η'_+ for the right-hand derivative of η. It exists by virtue of the convexity of η; moreover,

$$\eta(\lambda) \ge \eta(y) + \eta'_+(y)(\lambda - y)$$



for all λ. Whenever $\varepsilon > 0$, write

$$\zeta(y,\varepsilon) = \frac{\eta(y+\varepsilon) - \eta(y)}{\varepsilon} - \eta'_+(y).$$

Since η is *strictly* convex, $\zeta(y,\varepsilon) > 0$, and this quantity is an increasing function of ε. In particular, if $\lambda > y + \varepsilon$, then $\zeta(y, \lambda - y) > \zeta(y,\varepsilon)$, or

$$\eta(\lambda) > \eta(y) + \eta'_+(y)(\lambda - y) + \zeta(y,\varepsilon)(\lambda - y).$$

In *every* case, then,

$$\eta(\lambda) > \eta(y) + \eta'_+(y)(\lambda - y) + \zeta(y,\varepsilon)\big((\lambda - y)^+ - \varepsilon\big). \tag{3.104}$$

On the other hand, whenever $y < \lambda < y + \varepsilon$, then $\zeta(y, \lambda - y) > \zeta(y,\varepsilon)$, so

$$\eta(\lambda) < \eta(y) + \eta'_+(y)(\lambda - y) + \varepsilon\zeta(y,\varepsilon) \qquad (y \le \lambda < y + \varepsilon). \tag{3.105}$$

Let us now assume that $\varphi \ge 0$ is such that

$$\varphi(x) \ne 0 \Rightarrow y \le u_0(x) < y + \varepsilon. \tag{3.106}$$

We use (3.104) on the left-hand side and (3.105) on the right-hand side of (3.102), and get

$$\begin{aligned}\limsup_{\tau\to 0+} \frac{1}{\tau}\int_0^\tau \int_{\mathbb{R}} \langle v_{(x,t)}, \big[\eta(y) + \eta'_+(y)(\mathrm{Id} - y) \\ + \zeta(y,\varepsilon)\big((\mathrm{Id} - y)^+ - \varepsilon\big)\big]\rangle \varphi(x)\,dx\,dt \\ \le \int_{\mathbb{R}} \big[\eta(y) + \eta'_+(y)\big(u(x_0) - y\big) + \varepsilon\zeta(y,\varepsilon)\big]\varphi(x)\,dx.\end{aligned}$$

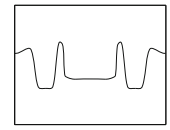
Here, thanks to (3.101) and the fact that $v_{(x,t)}$ is a probability measure, all the terms not involving $\zeta(y,\varepsilon)$ cancel, and then we can divide by $\zeta(y,\varepsilon) \ne 0$ to arrive at

$$\limsup_{\tau\to 0+} \frac{1}{\tau}\int_0^\tau \int_{\mathbb{R}} \langle v_{(x,t)}, (\mathrm{Id} - y)^+\rangle \varphi(x)\,dx\,dt \le 2\varepsilon \int_{\mathbb{R}} \varphi(x)\,dx.$$

Now, remembering (3.106), we see that whenever $\varphi(x) \ne 0$ we have $(\lambda - y)^+ \le (\lambda - u_0(x))^+ + \varepsilon$, so the above implies

$$\limsup_{\tau\to 0+} \frac{1}{\tau}\int_0^\tau \int_{\mathbb{R}} \langle v_{(x,t)}, \big(\mathrm{Id} - u_0(x)\big)^+\rangle \varphi(x)\,dx\,dt \le 3\varepsilon \int_{\mathbb{R}} \varphi(x)\,dx$$

whenever (3.106) holds.



It remains only to divide up the common support $[-M, M]$ of all the measures $\nu_{(x,t)}$, writing $y_i = -M + i\varepsilon$ for $i = 0, 1, \ldots, N-1$, where $\varepsilon = 2M/N$. Let φ_i be the characteristic function of $[-A, A] \cap u_0^{-1}([y_i, y_i + \varepsilon))$, and add together the above inequalities, one for each i, to arrive at

$$\limsup_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{-A}^{A} \langle \nu_{(x,t)}, \big(\mathrm{Id} - u_0(x)\big)^+ \rangle \varphi(x)\, dx\, dt \le 3\varepsilon \int_{-A}^{A} \varphi(x)\, dx.$$

Since ε can be made arbitrarily small, (3.103) follows, and the proof is complete. □

Remark 3.31 We *cannot* conclude that[3]

$$\lim_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{\mathbb{R}} \langle \nu_{(x,t)}, |\mathrm{Id} - u_0(x)| \rangle\, dx\, dt = 0 \tag{3.107}$$

from the present assumptions. Here is an example to show this.

Let $\nu_{(x,t)} = \mu_{\gamma(x,t)}$, where $\mu_\beta = \frac{1}{2}(\delta_{-\beta} + \delta_\beta)$ and γ is a continuous, nonnegative function with $\gamma(x, 0) = 0$. Let $u_0(x) = 0$ and $\eta(y) = y^2$.

Then (3.101) holds trivially, and (3.102) becomes

$$\limsup_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{\mathbb{R}} \gamma(x,t)^2 \varphi(x)\, dx\, dt = 0,$$

which is also true due to the stated assumptions on γ.

The desired conclusion (3.107), however, is now

$$\limsup_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{\mathbb{R}} \gamma(x,t)\, dx\, dt = 0.$$

But the simple choice

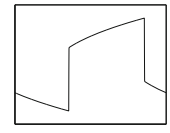
$$\gamma(x,t) = te^{-(xt)^2}$$

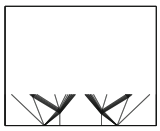
yields

$$\limsup_{\tau \to 0+} \frac{1}{\tau} \int_0^\tau \int_{\mathbb{R}} \gamma(x,t)\, dx\, dt = \sqrt{\pi}.$$

We shall now describe a framework that allows one to prove convergence of a sequence of approximations without proving that the method is TV stable. Unfortunately, the application of this method to concrete examples, while not very

[3] Where the integral over the compact interval $[-A, A]$ in (3.91) has been replaced by an integral over the entire real line.





difficult, involves quite large calculations, and will be omitted here. Readers are encouraged to try their hands at it themselves.

We give one application of these concepts. The setting is as follows. Let u^n be computed from a conservative and consistent scheme, and assume uniform boundedness of u^n. Young's theorem states that there exists a family of probability measures $\nu_{(x,t)}$ such that $g(u^n) \overset{*}{\rightharpoonup} \langle \nu_{(x,t)}, g \rangle$ for Lipschitz continuous functions g. We assume that the CFL condition, $\lambda \sup_u |f'(u)| \leq 1$, is satisfied. The next theorem states conditions, strictly weaker than TVD, for which we prove that the limit measure $\nu_{(x,t)}$ is a measure-valued solution of the scalar conservation law.

Theorem 3.32 *Let $u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$. Assume that the sequence $\{u^n\}$ is the result of a conservative, consistent method, and define $u_{\Delta t}$ as in (3.27). Assume that $u_{\Delta t}$ is uniformly bounded in $L^\infty(\mathbb{R} \times [0,T])$, $T = n\Delta t$. Let $\Delta t_n \to 0$ be a sequence such that $u_{\Delta t_n} \overset{*}{\rightharpoonup} u$, and let $\nu_{(x,t)}$ be the Young measure associated with $u_{\Delta t_n}$, and assume that u_j^n satisfies the estimate*

$$(\Delta x)^\beta \sum_{n=0}^{N} \sum_j \left| u_{j+1}^n - u_j^n \right| \Delta t \leq C(T), \tag{3.108}$$

for some $\beta \in [0,1)$ and some constant $C(T)$. Then $\nu_{(x,t)}$ is a measure-valued solution to (3.1).

Furthermore, let (η, q) be a strictly convex entropy pair, and let Q be a numerical entropy flux consistent with q. Write $\eta_j^n = \eta(u_j^n)$ and $Q_{j+1/2}^n = Q(u^n)_{j+1/2}$. Assume that

$$\frac{1}{\Delta t}\left(\eta_j^{n+1} - \eta_j^n\right) + \frac{1}{\Delta x}\left(Q_{j+1/2}^n - Q_{j-1/2}^n\right) \leq R_j^n \tag{3.109}$$

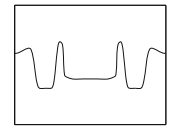
for all n and j, where R_j^n satisfies,

$$\lim_{\Delta t \to 0} \sum_{n=0}^{N} \sum_j \varphi_j^n R_j^n \Delta x \Delta t = 0 \tag{3.110}$$

for all nonnegative $\varphi \in C_0^1$ where $\varphi_j^n = \varphi(j\Delta x, n\Delta t)$. Then $\nu_{(x,t)}$ is a measure-valued solution compatible with (η, q), and the initial data is assumed in the sense of (3.101), (3.102). If (3.109) and (3.110) hold for all entropy pairs (η, q), then $\nu_{(x,t)}$ is a measure-valued entropy solution to (3.1).

Remark 3.33 For $\beta = 0$, (3.108) is the standard TV estimate, while for $\beta > 0$, (3.108) is genuinely weaker than a TV estimate.

Proof We start by proving the first statement in the theorem, assuming (3.108). As before, we obtain (3.28) by rearranging. For simplicity, we now write $F_{j+1/2}^n =$



$F(u^n)_{j+1/2}$, $f_j^n = f(u_j^n)$, and observe that $F_{j+1/2}^n = f_j^n + \left(F_{j+1/2}^n - f_j^n\right)$, getting

$$\begin{aligned} \iint \left(u_{\Delta t} D_+^t \varphi_j^n + f(u_{\Delta t}) D_+ \varphi_j^n\right) dx\, dt \\ = \sum_{j,n} D_+ \varphi_j^n \left(F_{j+1/2}^n - f_j^n\right) \Delta t\, \Delta x. \end{aligned} \tag{3.111}$$

Here we use the notation

$$u_{\Delta t} = u_j^n \quad \text{for} \quad (x,t) \in [j\,\Delta x, (j+1)\Delta x) \times [n\Delta t, (n+1)\Delta t)\,,$$

and

$$\begin{aligned} D_+^t \varphi_j^n &= \frac{1}{\Delta t}\left(\varphi_j^{n+1} - \varphi_j^n\right), \\ D_+ \varphi_j^n &= \frac{1}{\Delta x}\left(\varphi_{j+1}^n - \varphi_j^n\right). \end{aligned}$$

The first term on the left-hand side in (3.111) reads

$$\begin{aligned} \iint u_{\Delta t} D_+^t \varphi_j^n\, dx\, dt &= \iint \langle \nu_{(x,t)}, \mathrm{Id}\rangle \varphi_t\, dx\, dt + \iint \left(u_{\Delta t} - \langle \nu_{(x,t)}, \mathrm{Id}\rangle\right) \varphi_t\, dx\, dt \\ &\quad + \iint u_{\Delta t} \left(D_+^t \varphi_j^n - \varphi_t\right) dx\, dt. \end{aligned} \tag{3.112}$$

The third term on the right-hand side of (3.112) clearly tends to zero as Δt goes to zero. Furthermore, by definition of the Young measure $\nu_{(x,t)}$, the second term tends to zero as well. Thus the left-hand side of (3.112) approaches $\iint \langle \nu_{(x,t)}, \mathrm{Id}\rangle \varphi_t\, dx\, dt$.

One can use a similar argument for the second term on the left-hand side of (3.111) to show that the (whole) left-hand side of (3.111) tends to

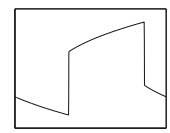
$$\iint \left(\langle \nu_{(x,t)}, \mathrm{Id}\rangle \varphi_t + \langle \nu_{(x,t)}, f\rangle \varphi_x\right) dx\, dt \tag{3.113}$$

as $\Delta t \to 0$. We now study the right-hand side of (3.111). Mimicking the proof of the Lax–Wendroff theorem, we have

$$\left|F_{j+1/2}^n - f_j^n\right| \le C \sum_{k=-p}^{q} \left|u_{j+k}^n - u_j^n\right|.$$

Therefore,

$$\begin{aligned} &\left|\sum_{j,n} D_+ \varphi_j^n \left(F_{j+1/2}^n - f_j^n\right) \Delta t \Delta x\right| \\ &\qquad \le C\,\|\varphi\|_{\mathrm{Lip}}(p+q+1) \sum_{n=0}^{N} \sum_{j} \left|u_{j+1}^n - u_j^n\right| \Delta t \Delta x \\ &\qquad \le C\,\|\varphi\|_{\mathrm{Lip}}(p+q+1)(\Delta x)^{1-\beta}, \end{aligned} \tag{3.114}$$



using the assumption (3.108). Thus the right-hand side of (3.114), and hence also of (3.111), tends to zero. Since the left-hand side of (3.111) tends to (3.113), we conclude that $\nu_{(x,t)}$ is a measure-valued solution. Using similar calculations, and (3.110), one shows that $\nu_{(x,t)}$ is also an entropy measure-valued solution.

It remains to show consistency with the initial condition, i.e., (3.101) and (3.102). Let $\varphi(x)$ be a test function, and we use the notation $\varphi(j\,\Delta x) = \varphi_j$. From the definition of u_j^{n+1}, after a summation by parts, we have that

$$\sum_j \varphi_j \left(u_j^{n+1} - u_j^n\right) \Delta x = \Delta t \sum_j F_{j+1/2}^n D_+ \varphi_j \, \Delta x \leq \mathcal{O}(1)\, \Delta t,$$

since u_j^n is bounded. Recall that $\varphi = \varphi(x)$, we get

$$\left| \sum_j \varphi_j \left(u_j^n - u_j^0\right) \Delta x \right| \leq \mathcal{O}(1)\, n \Delta t. \tag{3.115}$$

Let $t_1 = n_1 \Delta t$ and $t_2 = n_2 \Delta t$. Then (3.115) yields

$$\left| \frac{1}{(n_2 + 1 - n_1)\, \Delta t} \sum_{n=n_1}^{n_2} \sum_j \varphi_j \left(u_j^n - u_j^0\right) \Delta x \Delta t \right| \leq \mathcal{O}(1)\, t_2,$$

which implies that the Young measure $\nu_{(x,t)}$ satisfies

$$\left| \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \varphi(x) \left\langle \nu_{(x,t)}, \mathrm{Id} \right\rangle dx\, dt - \int \varphi(x) u_0(x)\, dx \right| \leq \mathcal{O}(1)\, t_2. \tag{3.116}$$

We let $t_1 \to 0$ and set $t_2 = \tau$ in (3.116), obtaining

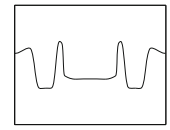
$$\left| \frac{1}{\tau} \int_0^{\tau} \int \varphi(x) \left\langle \nu_{(x,t)}, \mathrm{Id} \right\rangle dx\, dt - \int \varphi(x) u_0(x)\, dx \right| \leq \mathcal{O}(1)\, \tau, \tag{3.117}$$

which proves (3.101). Now for (3.102). We have that there exists a strictly convex entropy η for which (3.109) holds. Now let $\varphi(x)$ be a nonnegative test function. Using (3.109), and proceeding as before, we obtain

$$\left| \sum_j \left(\eta_j^n - \eta_j^0\right) \varphi_j \, \Delta x \right| \leq \mathcal{O}(1)\, n \Delta t + \sum_{l=0}^{n} \sum_j R_j^l \varphi_j \, \Delta t \, \Delta x.$$

Using this estimate and the assumption on R_j^l, (3.110), we can use the same arguments as in proving (3.117) to prove (3.102). The proof of the theorem is complete. □

A trivial application of this approach is found by considering monotone schemes. Here we have seen that (3.108) holds for $\beta = 0$, and (3.109) for $R_j^n = 0$. The theorem then gives the convergence of these schemes without using Helly's theorem. However, in this case the application does not give the existence of a solution, since we must have this in order to use DiPerna's theorem. The main usefulness of the method is for schemes in several space dimensions, where TV bounds are more difficult to obtain.



3.6 Notes

The Lax–Friedrichs scheme was introduced by Lax in 1954; see [124]. Godunov discussed what has later become the Godunov scheme in 1959 as a method to study gas dynamics; see [80]. The CFL condition was introduced in the seminal paper [50]; see also [57].

The Lax–Wendroff theorem, Theorem 3.4, was first proved in [128]. Theorem 3.8 was proved by Oleĭnik in her fundamental paper [145]; see also [169]. Several of the key results concerning monotone schemes are due to Crandall and Majda [53], [52]. Theorem 3.10 is due to Harten, Hyman, and Lax; see [84]. Harten's lemma, Lemma 3.12, can be found in [83]. See also [148].

The error analysis is based on the fundamental analysis by Kuznetsov, [119], where one also can find a short discussion of the examples we have analyzed, namely the smoothing method, the method of vanishing viscosity, as well as monotone schemes. Our presentation of the a priori estimates follows the approach due to Cockburn and Gremaud; see [44] and [45], where also applications to numerical methods are given.

The concept of measure-valued solutions is due to DiPerna, and the key results can be found in [62], while Lemma 3.30 is to be found in [61]. Our presentation of the Young measure follows the exposition of Perthame, [150]. For further information regarding the functional-analytic framework, see, e.g., [34] and references therein. The proof of Lemma 3.30 and Remark 3.31 are due to H. Hanche-Olsen. Our presentation of the uniqueness of measure-valued solutions, Theorem 3.29, is taken mainly from Szepessy, [173]. Theorem 3.32 is due to Coquel and LeFloch, [48]; see also [49], where several extensions are discussed. For numerical schemes that satisfy the criteria in Theorem 3.32, see [49] and [65].

3.7 Exercises

3.1 Consider the difference scheme (3.4). Show that if u^0 is given by

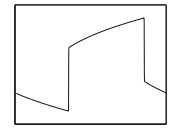
$$u_j^0 = \begin{cases} 0 & \text{for } j < 0, \\ 1 & \text{for } j \geq 0, \end{cases}$$

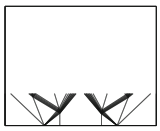
then $u^n = u^0$ for all n, thus indicating the solution $u(x,t) = \chi_{[0,\infty)}$. Determine the weak entropy solution.

3.2 Show that the Lax–Wendroff and the MacCormack methods are of second order.

3.3 The Engquist–Osher (or generalized upwind) method, see [63], is a conservative difference scheme with a numerical flux defined as follows:

$$F_{j+1/2}(u) = f^{\mathrm{EO}}\left(u_j, u_{j+1}\right), \quad \text{where}$$
$$f^{\mathrm{EO}}(u,v) = \int_0^u \max\{f'(s),0\}\,ds + \int_0^v \min\{f'(s),0\}\,ds + f(0).$$





(a) Show that this method is consistent and monotone.
(b) Find the order of the scheme.
(c) Show that the Engquist–Osher flux f^{EO} can be written

$$f^{\mathrm{EO}}(u,v) = \frac{1}{2}\left(f(u) + f(v) - \int_u^v |f'(s)|\, ds \right).$$

(d) If $f(u) = u^2/2$, show that the numerical flux can be written

$$f^{\mathrm{EO}}(u,v) = \frac{1}{2}\left(\max\{u,0\}^2 + \min\{v,0\}^2\right).$$

Generalize this simple expression to the case that $f''(u) \neq 0$ and $\lim_{|u|\to\infty} |f(u)| = \infty$.

3.4 Why does the method

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{2\Delta x}\left(f\left(u_{j+1}^n\right) - f\left(u_{j-1}^n\right)\right)$$

not give a viable difference scheme?

3.5 In the derivation of the Godunov scheme it is assumed that $\Delta t \max_u |f'(u)| \leq \frac{1}{2}\Delta x$, yet it is stated that the method is well defined if the CFL condition $\Delta t \max_u |f'(u)| \leq \Delta x$ is satisfied; see (3.9). Please explain.

3.6 Show that (3.24) is the model equation for the Lax–Friedrichs scheme.

3.7 Show that the Lax–Friedrichs scheme is monotone also in the case that the flux function is assumed only to be Lipschitz continuous.

3.8 Show that Heun's method is unstable.

3.9 We study a nonconservative method for Burgers's equation. Assume that $u_j^0 \in [0,1]$ for all j. Then the characteristic speed is nonnegative, and we define

$$u_j^{n+1} = u_j^n - \lambda u_j^{n+1}\left(u_j^n - u_{j-1}^n\right), \quad n \geq 0, \tag{3.118}$$

where $\lambda = \Delta t/\Delta x$.

(a) Show that this yields a monotone method, provided that a CFL condition holds.
(b) Show that this method is consistent and determine the truncation error.

3.10 Assume that $f'(u) > 0$ and that $f''(u) \geq 2c > 0$ for all u in the range of u_0. We use the upwind method to generate approximate solutions to

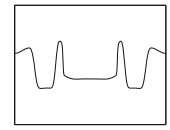
$$u_t + f(u)_x = 0, \quad u(x,0) = u_0(x); \tag{3.119}$$

i.e., we set

$$u_j^{n+1} = u_j^n - \lambda\left(f(u_j^n) - f(u_{j-1}^n)\right).$$

Set

$$v_j^n = \frac{u_j^n - u_{j-1}^n}{\Delta x}.$$



(a) Show that

$$v_j^{n+1} = \left(1 - \lambda f'(u_{j-1}^n)\right) v_j^n + \lambda f'(u_{j-1}^n) v_{j-1}^n - \frac{\Delta t}{2}\left(f''(\eta_{j-1/2})\left(v_j^n\right)^2 + f''(\eta_{j-3/2})\left(v_{j-1}^n\right)^2\right),$$

where $\eta_{j-1/2}$ is between u_j^n and u_{j-1}^n.

(b) Next, assume inductively that

$$v_j^n \le \frac{1}{(n+2)c\Delta t}, \quad \text{for all } j,$$

and set $\hat{v}^n = \max\{\max_j v_j^n, 0\}$. Then show that

$$\hat{v}^{n+1} \le \hat{v}^n - c\Delta t\,(\hat{v}^n)^2.$$

(c) Use this to show that

$$\hat{v}^n \le \frac{\hat{v}^0}{1 + \hat{v}^0 cn\Delta t}.$$

(d) Show that this implies that

$$u_i^n - u_j^n \le \Delta x(i-j)\frac{\hat{v}^0}{1 + \hat{v}^0 cn\Delta t},$$

for $i \ge j$.

(e) Let u be the entropy solution of (3.119), and assume that

$$0 \le \max_x u_0'(x) = M < \infty.$$

Show that for almost every x, y, and t we have that

$$\frac{u(x,t) - u(y,t)}{x - y} \le \frac{M}{1 + cMt}. \tag{3.120}$$

This is the Oleĭnik entropy condition for convex scalar conservation laws.

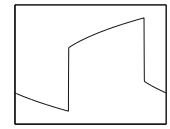
3.11 Assume that f is as in the previous exercise, and that u_0 is periodic with period p.

(a) Use uniqueness of the entropy solution to (3.119) to show that the entropy solution $u(x,t)$ is also periodic in x with period p.

(b) Then use the Oleĭnik entropy condition (3.120) to deduce that

$$\sup_x u(x,t) - \inf_x u(x,t) \le \frac{Mp}{1 + cMt}.$$

Thus $\lim_{t\to\infty} u(x,t) = \bar{u}$ for some constant $\bar{u}$.



(c) Use conservation to show that

$$\bar{u} = \frac{1}{p}\int_0^p u_0(x)\,dx.$$

3.12 Let $u_n\colon [0,1) \to [-1,1]$ be defined as

$$u_n(x) = \begin{cases} 1 & x \in [2k/2n, (2k+1)/2n), \\ -1 & x \in [(2k+1)/2n, (2k+2)/2n), \end{cases} \qquad \text{for } k = 0, \ldots, n-1,$$

for $n \in \mathbb{N}$. Find the weak limit of u_n as $n \to \infty$, and the associated Young measure.

3.13 We shall consider a scalar conservation law with a "fractal" function as the initial data. Define the set of piecewise linear functions

$$\mathcal{D} = \{\phi(x) = Ax + B \mid x \in [a,b], A, B \in \mathbb{R}\},$$

and the map

$$F(\phi) = \begin{cases} 2D(x-a) + \phi(a) & \text{for } x \in [a, a + L/3], \\ -D(x-a) + \phi(a) & \text{for } x \in [a + L/3, a + 2L/3], \\ 2D(x-b) + \phi(b) & \text{for } x \in [a + 2L/3, b], \end{cases}$$

for $\phi \in \mathcal{D}$, where $L = b - a$ and $D = (\phi(b) - \phi(a))/L$. For a nonnegative integer k introduce $\chi_{j,k}$ as the characteristic function of the interval $I_{j,k} = [j/3^k, (j+1)/3^k]$, $j = 0, \ldots, 3^{k+1} - 1$. We define functions $\{v_k\}$ recursively as follows. Let

$$v_0(x) = \begin{cases} 0 & \text{for } x \le 0, \\ x & \text{for } 0 \le x \le 1, \\ 1 & \text{for } 1 \le x \le 2, \\ 3 - x & \text{for } 2 \le x \le 3, \\ 0 & \text{for } 3 \le x. \end{cases}$$

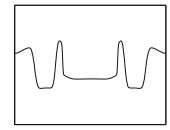
Assume that $v_{j,k}$ is linear on $I_{j,k}$ and let

$$v_k = \sum_{j=-3^k}^{3^k-1} v_{j,k}\chi_{j,k}, \tag{3.121}$$

and define the next function v_{k+1} by

$$v_{k+1} = \sum_{j=0}^{3^{k+1}-1} F(v_{j,k})\chi_{j,k} = \sum_{j=0}^{3^{k+2}-1} v_{j,k+1}\chi_{j,k+1}. \tag{3.122}$$

In the left part of Fig. 3.9 we show the effect of the map F, and on the right we show $v_5(x)$ (which is piecewise linear on $3^6 = 729$ segments).



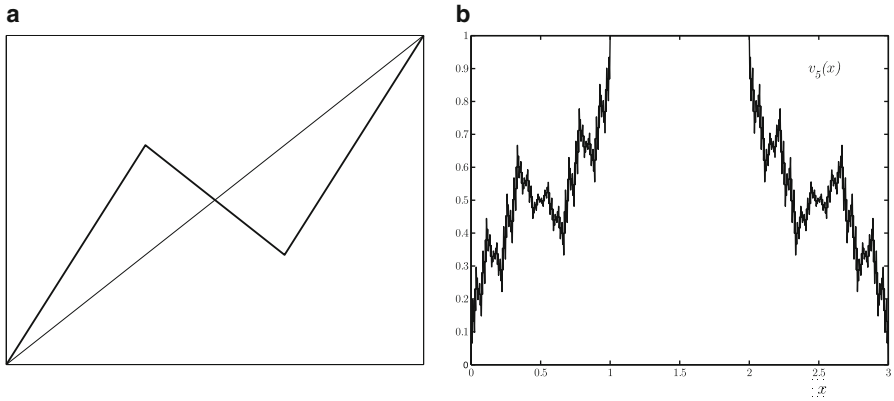


Fig. 3.9 **a** The construction of $F(\phi)$ from ϕ. **b** $v_5(x)$

(a) Show that the sequence $\{v_k\}_{k>1}$ is a Cauchy sequence in the supremum norm, and hence we can define a continuous function v by setting

$$v(x) = \lim_{k\to\infty} v_k(x).$$

(b) Show that v is not of bounded variation, and determine the total variation of v_k.

(c) Show that

$$v(j/3^k) = v_k(j/3^k),$$

for all integers $j = 0, \ldots, 3^{k+1}$, $k \in \mathbb{N}$.

(d) Assume that f is a C^1 function on $[0, 1]$ with $0 \leq f'(u) \leq 1$. We are interested in solving the conservation law

$$u_t + f(u)_x = 0, \quad u_0(x) = v(x).$$

To this end we shall use the upwind scheme defined by (3.10), with $\Delta t = \Delta x = 1/3^k$, and

$$u_j^0 = v(j\,\Delta x).$$

Show that $u_{\Delta t}(x, t)$ converges to an entropy solution of the conservation law above.