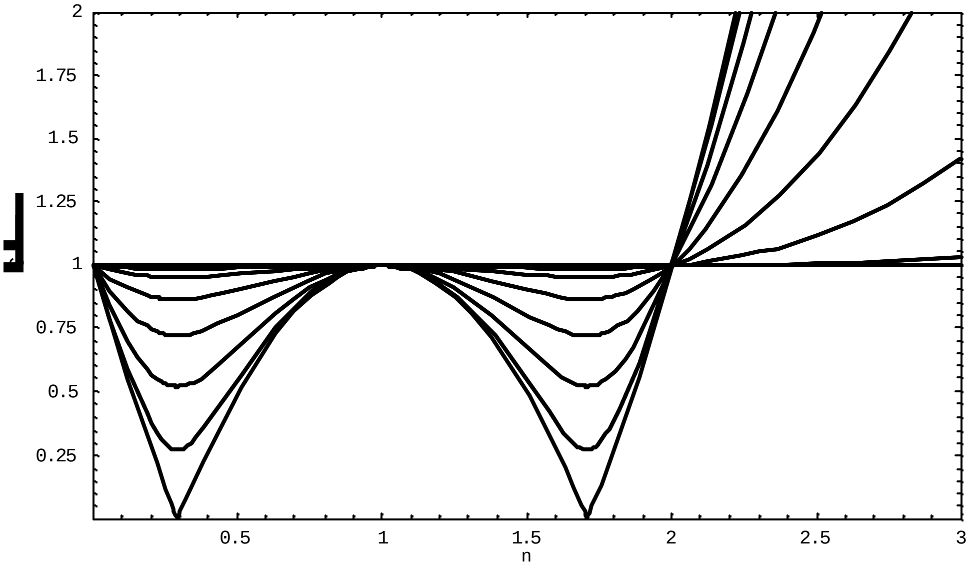

Figure 2. Amplification factor $|G(\nu,\theta)|$ for the Beam-Warming scheme evaluated at $\theta = m\pi/8$, $m = 0, 1, \ldots 16$.

and the step sizes satisfy the stability criterion.

1.1.5. Modified equations. We have established a number of methods for solving the advection equation (1.1). Up to now we have characterized the error of any one scheme by its truncation error. Though indicative of the overall quality of an approximation, the precise nature of the error in the scheme is not apparent. It has proved very fruitful in the development of better methods to more accurately describe how a numerical approximation differs from the exact solution. A question one can ask is whether a given numerical scheme is perhaps a more accurate discretization of another PDE than the one it was originally designed for. Let us exemplify using the upwind scheme for the advection equation with $u > 0$

$$Q_j^{n+1} = Q_j^n - \nu\left(Q_j^n - Q_{j-1}^n\right). \tag{1.74}$$

We know that this scheme is $O(k, h)$ accurate for the equation $q_t + uq_x = 0$. Suppose that the scheme is an exact discretization of some unknown PDE $Ls = 0$ with L an unknown differential operator and $s = s(x, t)$. Then we would have

$$s(x, t + k) = s(x, t) - \nu\left[s(x, t) - s(x - h, t)\right], \tag{1.75}$$

exactly. Let us carry out Taylor series expansion of s around (x, t)

$$s + ks_t + \frac{k^2}{2}s_{tt} + \frac{k^3}{6}s_{ttt} + \ldots = s - \frac{uk}{h}\left(hs_x - \frac{h^2}{2}s_{xx} + \frac{h^3}{6}s_{xxx} - \ldots\right). \tag{1.76}$$

To obtain a more concise notation the function arguments have been dropped. We obtain

$$s_t + us_x = -\frac{k}{2}s_{tt} + \frac{uh}{2}s_{xx} - \frac{k^2}{6}s_{ttt} - \frac{uh^2}{6}s_{xxx} + \ldots \tag{1.77}$$

This is of the form $As = E_{(h,k)}s$ with A the advection operator $A = \partial_t + u\partial_x$ and $E_{(h,k)}$ an operator giving the deviation of the modified equation from the advection equation. Note that if $k = h = 0$ we obtain the advection equation for which the scheme (1.74) is $O(h, k)$ accurate. We can interpret (1.77) as stating that the scheme (1.74) is:

(1) first order accurate for

$$s_t + us_x = 0 \tag{1.78}$$

(2) second order accurate for

$$s_t + us_x = -\frac{k}{2}s_{tt} + \frac{uh}{2}s_{xx} \tag{1.79}$$

(3) third order accurate for

$$s_t + us_x = -\frac{k}{2}s_{tt} + \frac{uh}{2}s_{xx} - \frac{k^2}{6}s_{ttt} - \frac{uh^2}{6}s_{xxx} \tag{1.80}$$

The equations obtained above are called modified equations. These statements can be verified by explicit computation of the truncation error. For example let us compute the truncation error in applying (1.74) to (??)

$$\tau_j^n = \left(\bar{D} - D\right) s(x_j, t^n) \tag{1.81}$$

The finite difference approximation operator is

$$\bar{D}s(x_j, t^n) = \frac{s_j^{n+1} - s_j^n}{k} + \frac{u}{h}\left[s_j^n - s_{j-1}^n\right] \tag{1.82}$$

The exact operator for the modified equation (??) is

$$D = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + \frac{k}{2}\frac{\partial^2}{\partial t^2} - \frac{uh}{2}\frac{\partial^2}{\partial x^2} \tag{1.83}$$

and we have $Ds = 0$. We now expand s_j^{n+1}, s_{j-1}^n, s_{j+1}^n from (1.82) around (x_j, t^n) and obtain

$$\tau_j^n = \frac{1}{k}\left(s + ks_t + \frac{k^2}{2}s_{tt} + \dots\right) - \frac{1}{k}s + \frac{u}{h}\left(s - s + hs_x - \frac{h^2}{2}s_{xx} + \frac{h^3}{6}s_{xxx} + \dots\right) \tag{1.84}$$

$$\tau_j^n = s_t + \frac{k}{2}s_{tt} + us_x + \frac{uh}{2}s_{xx} + O(k^2, h^2) = Ds + O(k^2, h^2) = O(k^2, h^2) \tag{1.85}$$

so the truncation error is indeed of second order.

Now let us show the benefits of looking at the modified equation by using (??) for which the upwind scheme (1.77) is second order accurate. First we recast (??) so as to eliminate higher order derivatives in time. We can rewrite (??) as

$$s_t = -us_x - \frac{k}{2}s_{tt} + \frac{uh}{2}s_{xx} \tag{1.86}$$

and differentiate with respect to t to obtain

$$s_{tt} = -us_{xt} - \frac{k}{2}s_{ttt} + \frac{uh}{2}s_{xxt} \tag{1.87}$$

Replacing (1.87) in (??) gives

$$s_t + us_x = -\frac{k}{2}\left(-us_{xt} - \frac{k}{2}s_{ttt} + \frac{uh}{2}s_{xxt}\right) + \frac{uh}{2}s_{xx} \tag{1.88}$$

$$= \frac{uk}{2}s_{xt} + \frac{uh}{2}s_{xx} + O(k^2, h^2, kh) \tag{1.89}$$

We can neglect the higher order terms since this is consistent with the order of accuracy used in obtaining (??). Differentiating (??) with respect to x yields

$$s_{tx} = -us_{xx} - \frac{k}{2}s_{ttx} + \frac{uh}{2}s_{xxx} \tag{1.90}$$

and replacing this in (??) gives

$$s_t + us_x = -\frac{u^2k}{2}s_{xx} + \frac{uh}{2}s_{xx} = \frac{uh}{2}(1-\nu)s_{xx} \tag{1.91}$$

Equation (1.91) is the usual way to express the modified equation for the upwind scheme applied to the advection equation to second order. It shows that the upwind scheme does indeed model the advection equation in the limit $h \to 0$. For finite step sizes however the upwind scheme more accurately models the equation (1.91). The difference between (1.91) and the advection equation is the term

$$\frac{uh}{2}(1-\nu)s_{xx} \tag{1.92}$$

Note that this is a diffusive term whose effect is to smooth out any variations in $s(x,t)$ as long as $|1-\nu| \geq 0$ as has been seen in the study of the heat equation. The condition $|1-\nu| \geq 0$ is exactly the stability criterion for the upwind scheme. Indeed if $\nu > 1$ then we would obtain a negative diffusion coefficient for which the initial value problem is ill-defined. We can see that at exactly $\nu = 1$ there is no diffusion indicating that for $\nu = 1$ the upwind scheme achieves higher order accuracy for the advection equation. When $\nu < 1$ the error in the upwind scheme with the respect to the true solution $q(x,t)$ of the advection equation will be diffusive: gradients will be smoothed out instead of being simply advected.

Now that we have seen the nature of the error introduced by the upwind scheme applied to the advection equation, we can also use this information to derive better schemes. Since the error is known to be given by (1.92) we can change the upwind scheme

$$Q_j^{n+1} = Q_j^n - \nu\left(Q_j^n - Q_{j-1}^n\right) \tag{1.93}$$

to counteract the known error by including a discretization of (1.92)

$$Q_j^{n+1} = Q_j^n - \nu\left(Q_j^n - Q_{j-1}^n\right) - k\frac{uh}{2}(1-\nu)\frac{Q_{j+1}^n - 2Q_j^n + Q_{j-1}^n}{h^2} \tag{1.94}$$

Working this through leads to the scheme

$$Q_j^{n+1} = Q_j^n - \frac{\nu}{2}\left(Q_{j+1}^n - Q_{j-1}^n\right) + \frac{\nu^2}{2}\left(Q_{j+1}^n - 2Q_j^n + Q_{j-1}^n\right). \tag{1.95}$$

Thus we have obtained the Lax-Wendroff scheme (1.30) via another route.

The procedure can be continued to higher orders. We can now ask what is the modified equation more accurately described by the Lax-Wendroff scheme. Repeating the same procedures as above we first write

(1.96)
$$s(x,t+k) = s(x,t) - \frac{\nu}{2}\left[s(x+h,t) - s(x-h,t)\right] + \frac{\nu^2}{2}\left[s(x+h,t) - 2s(x,t) + s(x-h,t)\right]$$

and then carry out Taylor series expansions around (x,t) to obtain

$$s + ks_t + \frac{k^2}{2}s_{tt} + \frac{k^3}{6}s_{ttt} + \ldots = s - \frac{uk}{2h}\left[2hs_x + \frac{h^3}{3}s_{xxx} + \ldots\right] + \tag{1.97}$$

$$\frac{u^2k^2}{2h^2}\left[h^2 s_{xx} + \frac{h^4}{12}s_{xxx}\right] \tag{1.98}$$

from where

$$s_t + us_x = -\frac{k}{2}\left(s_{tt} - u^2 s_{xx}\right) - \frac{k^2}{6}s_{ttt} - \frac{uh^2}{6}s_{xxx} + O(k^3,h^3) \tag{1.99}$$

Note the appearance of the $O(k)$ term. Had we carried out the Taylor expansion for the advection equation this term would have been proportional to $q_{tt} - u^2 q_{xx}$ which is zero according to (1.26). Here we cannot assume that $s_{tt} - u^2 s_{xx}$ is zero a priori. We must carry the term in the ensuing computations, expecting that it will give a higher order correction. Let us neglect the $O(k^3,h^3)$ contributions and proceed with our technique of replacing higher order time derivatives with spatial derivatives using

$$s_{tt} = -us_{xt} - \frac{k}{2}\left(s_{ttt} - u^2 s_{xxt}\right) \tag{1.100}$$

$$s_{ttt} = -us_{xtt}. \tag{1.101}$$

Higher order terms have been dropped since they would lead to $O(k^3,h^3)$ contributions in (1.99). Our intermediate result is

$$s_t + us_x = -\frac{k}{2}\left[-us_{xt} - \frac{k}{2}\left(s_{ttt} - u^2 s_{xxt}\right) - u^2 s_{xx}\right] + \frac{k^2}{6}us_{xtt} - \frac{uh^2}{6}s_{xxx} \tag{1.102}$$

and we continue by eliminating mixed derivatives. In the above formula we wish to express s_{xtt} in terms of x derivatives to $O(1)$

$$s_{xtt} = s_{ttx} = (s_t)_{tx} = (-us_x)_{tx} = -u\,(s_t)_{xx} = u^2 s_{xxx}\ . \tag{1.103}$$

We also need to express s_{xt} in terms of x derivatives to $O(k,h)$

$$s_{xt} = s_{tx} = -us_{xx} - \frac{k}{2}\left(s_{ttx} - u^2 s_{xxx}\right) \tag{1.104}$$

and s_{ttt}, s_{xxt} to $O(1)$

$$s_{ttt} = -u^3 s_{xxx},\quad s_{xxt} = -us_{xxx}\ . \tag{1.105}$$

Replacing in (1.102) leads to

(1.106)
$$s_t + us_x = -\frac{k}{2}\left\{-u\left[-us_{xx} - \frac{k}{2}\left(s_{ttx} - u^2 s_{xxx}\right)\right] - u^2 s_{xx} - \frac{k}{2}\left(-u^3 s_{xxx} + u^3 s_{xxx}\right)\right\} + \frac{u}{6}\left(k^2u^2 - h^2\right)s_{xxx}$$

which simplifies to

$$s_t + us_x = \frac{k^2}{4}\left(s_{ttx} - u^2 s_{xxx}\right) + \frac{u}{6}\left(k^2u^2 - h^2\right)s_{xxx} \tag{1.107}$$

Since $s_{ttx} = u^2 s_{xxx}$ to $O(1)$ we obtain in final

$$s_t + u s_x = -\frac{uh^2}{6}\left(1 - \nu^2\right) s_{xxx}\ . \tag{1.108}$$

The third order derivative now obtained shows that the Lax-Wendroff scheme introduces a dispersive error with different wave numbers traveling at different speeds. As expected, the dispersive error is proportional to h^2 since the Lax-Wendroff scheme is second order. A scheme more accurate than Lax-Wendroff could be obtained by adding a correction term modeling the dispersive error. Since this involves a third-order derivative the stencil of the scheme would become wider by at least one unit thereby entailing more computational work.

1.2. Non-linear scalar equations. We have introduced a number of finite difference methods for the simple constant-velocity advection equation (1.1). Of course, there is hardly much need for a numerical method in order to solve (1.1). Rather we have used (1.1) as a model problem to study the properties of numerical schemes on a simple case. We now proceed to consider more complicated problems and investigate how the methods already derived apply to these problems.

A general first order, hyperbolic scalar equation is given by

$$q_t + u(x,t,q) q_x = \sigma(x,t,q) \tag{1.109}$$

where u may be interpretred as local advection velocity that depends in general upon x, t and q. In a wide range of problems equations of the form

$$q_t + f(q)_x = \sigma(x,t,q) \tag{1.110}$$

arise where f is known as the flux function. If f is differentiable we can write

$$q_t + f_q q_x = \sigma(x,t,q) \tag{1.111}$$

so f_q plays the role of the local advection velocity. Generally u, f depend on q so that the equations become non-linear in q. Equation (1.110) is said to be in conservative form as opposed to (1.109) which is said to be in non-conservative form. Generally we say that a PDE is in conservative form when it can be expressed as the space-time divergence of a vector field. For equation (1.110) the vector field would be $(q, f(q))$ and the space-time divergence is $\nabla_{(t,x)} \cdot = (\partial_t, \partial_x) \cdot$ so another way of writing (1.110) is

$$\nabla_{(t,x)} \cdot (q, f(q)) = 0\ . \tag{1.112}$$

An initial value problem is defined by specifying a solution domain along x and an initial condition $q_0(x)$.

1.2.1. *Solution by characteristics.* We can solve (1.109) using the moethod of characteristics. We again ask whether there are any special curves Γ within the (x,t) plane on which (1.109) reduces to a simpler form. Along the curves specified by the differential system

$$\frac{dt}{ds} = 1,\ \frac{dx}{ds} = u(x,t,q) \tag{1.113}$$

we do indeed obtain the simpler form

$$\frac{dq}{ds} = \sigma. \tag{1.114}$$