Lecture 2

Convergence and Accuracy

This lecture is a bit on the technical side, but the concepts introduced are critical to the analysis of finite difference methods for ODE's.

2.1 Convergence and Global Accuracy

As the timestep is decreased, i.e. $\Delta t \to 0$, the approximation from a finite difference method should converge to the solution of the ODE. This concept is known as convergence and is stated mathematically as follows:

Definition 2.1 (Convergence) *A finite difference method for solving,*

$$\dot{u} = f(u,t) \qquad with \qquad u(0) = u_0$$

from $t = 0$ to T is convergent if

$$\max_{n=[0,T/\Delta t]} |v^n - u(n\Delta t)| \to 0 \qquad as \qquad \Delta t \to 0.$$

While convergence is a clear requirement for a good numerical method, the rate at which the method converges is also important. This rate is known as the global order of accuracy.

Definition 2.2 (Global Order of Accuracy) *A method has a global order of accuracy of p if,*

$$\max_{n=[0,T/\Delta t]} |v^n - u(n\Delta t)| \leq O(\Delta t^p) \qquad as \qquad \Delta t \to 0,$$

for any $f(u,t)$ that has p continuous derivatives (i.e. up to and including $\partial^p f/\partial t^p$ and $\partial^p f/\partial u^p$).

Thus, methods with higher p will converge to $u(t)$ more rapidly than those methods with lower p.

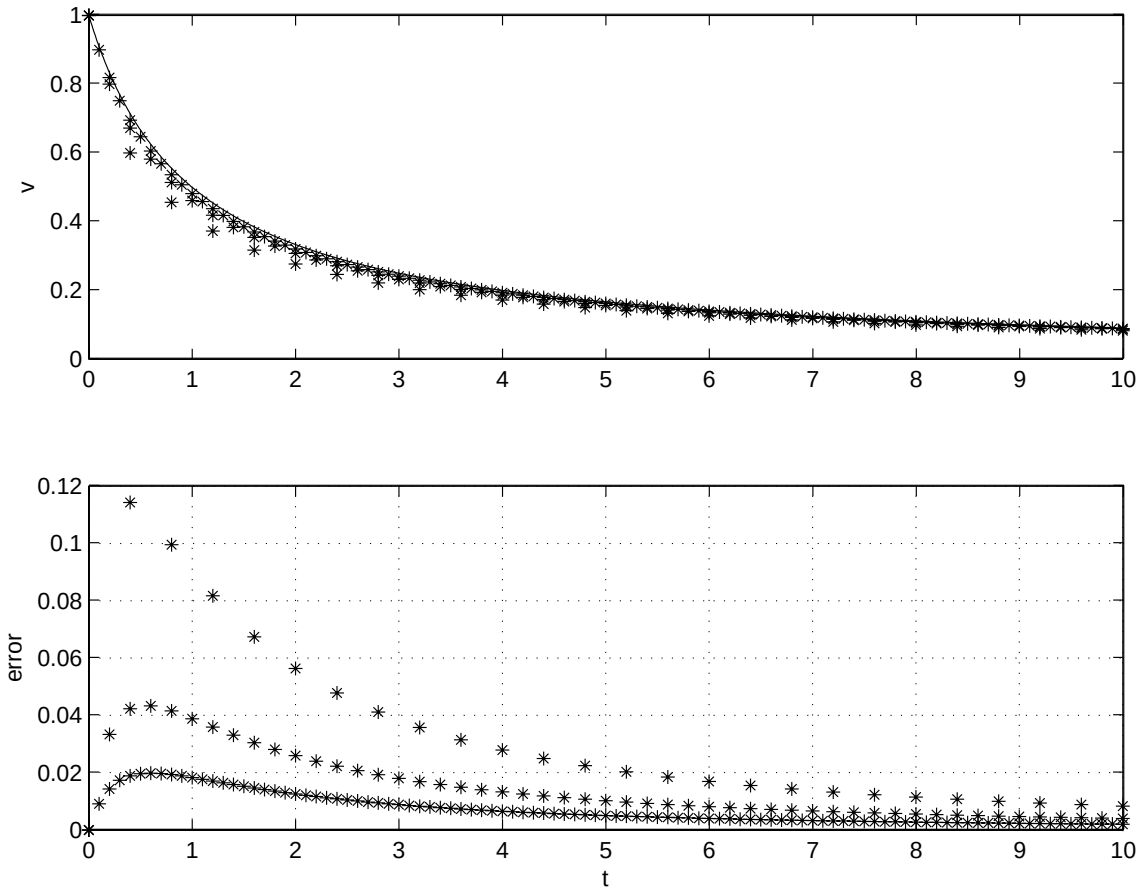

Figure 2.1: Forward Euler solution for $\dot{u} = -u^2$ with $u(0) = 1$ with $\Delta t = 0.1$, 0.2, and 0.4. Forward Euler (symbols) and exact solution (line) are shown in first plot. Error is shown in second plot.

Example 2.1 *To demonstrate the ideas of global accuracy, we will consider an ODE with* $f = -u^2$ *and an initial condition of* $u(0) = 1$. *The solution to this ODE is* $u = (1+t)^{-1}$. *Now, let us apply the forward Euler method to solving this problem for* $t = 0$ *to* 10. *The approximate solutions for a range of* Δt *are shown Figure 2.1 along with the exact solution. The forward Euler solutions are clearly approaching the exact solution as* Δt *decreases. Furthermore, the error appears to be decreasing by approximately a factor of 2 for every factor of 2 decrease in* Δt. *For example, if we look at* $t = 4$, *the error is seen to be* 0.028, 0.013, *and* 0.0065 *for* $\Delta t = 0.4$, 0.2, *and* 0.1, *respectively. Thus, from these results, we would conclude that the global accuracy of the forward Euler method is* $p = 1$ *since the error is proportional to* Δt.

Example 2.2 *Now, let's apply the midpoint method on the problem from Example 2.1. Similar to the results observed in Example 1.5, the midpoint method shows an oscillatory behavior (this may be a little hard to see because of scale of the figure, but the midpoint results are basically oscillated about the exact solution, with the oscillations reducing for the smaller timesteps). Note that the timesteps used in these results are a factor of 10 smaller than those used with the forward Euler method in Example 2.1. Since the midpoint and the for-*

ward Euler method require essentially the same work per timestep, the midpoint results took about a factor of 10 more work than the forward Euler method for this problem. Another interesting aspect of these results is that the error is actually increasing as t increases (in the forward Euler results in Figure 2.1, the error decreased as t increased). Regardless, the method does appear convergent since as the timestep decreases, so are the errors. In fact, it appears that the errors are decreasing by a factor of 4 for a factor of 2 decrease in Δt*. For example, if we look at* $t = 4$*, the error (averaged to remove the oscillations) is seen to be approximately* 0.02, 0.005, *and* 0.00125 *for* $\Delta t = 0.04$, 0.02, *and* 0.01, *respectively. Thus, from these results, we would conclude that the global accuracy of the midpoint method is* $p = 2$ *since the error is proportional to* Δt^2*.*

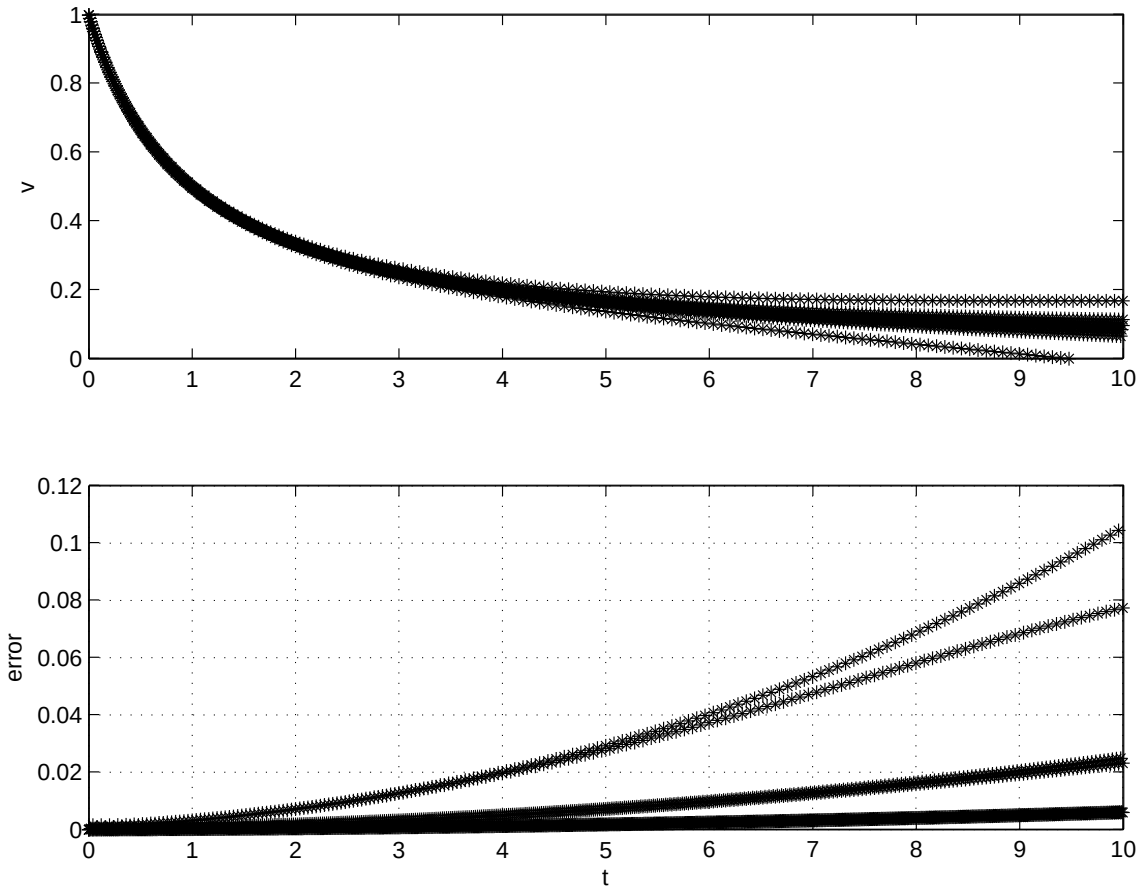

Figure 2.2: Forward Euler solution for $\dot{u} = -u^2$ with $u(0) = 1$ with $\Delta t = 0.01$, 0.02, and 0.04. Midpoint method (symbols) and exact solution (line) are shown in first plot. Error is shown in second plot.

2.2 Local Accuracy

The analysis of convergence and global accuracy usually relies on the analsysi of consistency and local accuracy. Both convergence/global accuracy and consistency/local accuracy are

related to the behavior of the error as $\Delta t \to 0$. However, while convergence/global accuracy is associated with the behavior of the error over a finite time (i.e. from $t = 0$ to T), consistency/local accuracy is associated with the behavior of the error for a single timestep. If we can quantify how much the error changes in a single timestep, then we will have an indication of how much the error could change over a series of timesteps. Specifically, let's write the solution error at $t = T$ as a sum of the change in error at each timestep,

$$e(T) = u(T) - v^{T/\Delta t} = \sum_{n=1}^{T/\Delta t} \Delta e^n,$$

where Δe^n is the change in the error from iteration $n-1$ to n (i.e. the local error). Suppose the local error is $O(\Delta t^{p+1})$, then the global error might be expected to behave as,

$$\begin{aligned} e(T) &= \sum_{n=1}^{T/\Delta t} \Delta e^n, \\ &= \sum_{n=1}^{T/\Delta t} O(\Delta t^{p+1}), \\ &= \frac{T}{\Delta t} O(\Delta t^{p+1}) \\ &= O(\Delta t^p). \end{aligned}$$

Thus, the global error would be one order less than the local error because the local errors sum for $T/\Delta t$ timesteps. However, the local errors do not have to sum this way if the numerical method is not stable. But, if a numerical method is both consistent and stable, this will be enough to guarantee convergence. For now, we concentrate on quantifying the local accuracy and leave the discussion of consistency and stability for another lecture.

The local error (usually called the local truncation error) is the difference between the approximate solution and the exact solution when using the exact solution for all of the required data. Let's consider the forward Euler method as an example. Recall, the forward Euler method is,

$$v^{n+1} = v^n + \Delta t f(v^n, t^n).$$

Thus, for the forward Euler method, $v^{n+1} = v^{n+1}(v^n, \Delta t, t^n)$. Then, if we substitute the exact solution into the right-hand side, we find,

$$v^{n+1}(u^n, \Delta t, t^n) = u^n + \Delta t f(u^n, t^n).$$

Recall our notation that u is the exact solution; in this discussion we use the superscript notation $u^n = u(n\Delta t)$ realizing that $u = u(t)$. The local truncation error for the forward Euler method is then,

$$\text{Local truncation error} \equiv v^{n+1}(u^n, \Delta t, t^n) - u^{n+1}. \tag{2.1}$$

Substitution gives,

$$\text{Local truncation error} = u^n + \Delta t f(u^n, t^n) - u^{n+1}.$$

The local order of accuracy is then found using a Taylor series expansion about $t = t^n$. Recall that $f(u^n, t^n) = \dot{u}(t^n)$ and

$$u(t^{n+1}) = u(t^n) + \Delta t \dot{u}(t^n) + \frac{1}{2}\Delta t^2 \ddot{u}(t^n) + O(\Delta t^3).$$

Substitution gives the local truncation error as,

$$\begin{aligned} \text{Local truncation error} &= u^n + \Delta t f(u^n, t^n) - u^{n+1}, \\ &= u(t^n) + \Delta t \dot{u}(t^n) - \left[u(t^n) + \Delta t \dot{u}(t^n) + \frac{1}{2}\Delta t^2 \ddot{u}(t^n) + O(\Delta t^3)\right] \\ &= -\frac{1}{2}\Delta t^2 \ddot{u}(t^n) + O(\Delta t^3). \end{aligned}$$

Thus, the leading term of the local trunction error for the forward Euler method is $-\frac{1}{2}\Delta t^2 \ddot{u}(t^n) = O(\Delta t^2)$. Based on our previous argument, we expect that the global accuracy of the forward Euler method should be $O(\Delta t)$ (i.e. first order accuracy). This was in fact observed in Example 2.1.

In-class Discussion 2.1 (Local accuracy of the midpoint method)

Definition 2.3 (Local Order of Accuracy) *Suppose we are given a numerical method for solving $\dot{u} = f(u, t)$ which we write in the following form,*

$$v^{n+1} = N(v^{n+1}, v^n, v^{n-1}, \ldots, \Delta t)$$

For simplicity, the possible dependence on t at various n has been omitted in the definition of N (though it should be there). The local truncation error, τ, is defined as,

$$\tau \equiv N(u^{n+1}, u^n, u^{n-1}, \ldots, \Delta t) - u^{n+1},$$

and the local order of accuracy p is,

$$|\tau| = O(\Delta t^{p+1}) \qquad as \qquad \Delta t \to 0.$$

Note: the local order of accuracy is defined to be one less than the order of the leading term of the local truncation error so that the local and global accuracy will be the same.

By the definition of the local order of accuracy, we see that the forward Euler method is first order ($p = 1$) and the midpoint method is second order ($p = 2$).