Math 228B, Lecture 5
January 20, 2009
Edward Tavernetti
Edited by Robert Hildebrand

Stability, Consistency, and Convergence
For a given a PDE: discretize temporal and spatial terms, come up with difference equations, solve numerically with a cpu, hope that the solution generated by the finite difference equations is a good approximate solution to actual the solution. We are going to need to consider what restrictions we should have on the time and space discretizations keeping in mind that convergence is the ultimate goal.

If we have a second order accurate LTE we would like to be able to say that we have a second order method, but consistency is not enough to ensure this. Consistency only implies that the LTE goes to zero as time and space discretization go to zero ($\Delta t, h \to 0$). Today we will show that stability is also required to ensure convergence

Lax Equivalence Theorem (fundamental theorem of FD-methods):
Stability + Consistency $\Rightarrow$ Convergence
A stronger statement (which we will not prove): A consistent finite difference scheme for a well-posed problem is convergent if and only if it is stable. But keep in mind that the Lax Equivalence Theorem only applies to

- Linear, well-posed PDEs
- Linear Schemes

Why is Lax Equivalence Theorem useful? Because, in general, showing convergence is hard, but it is easy to enforce consistency and then the work in proving convergence is in proving stability. Proving stability is more straightforward that proving convergence directly.

The natural question is then what does it mean to be stable?

Stability

Recall the example from a few weeks ago in which we saw growing oscillations in the solution to the diffusion equation when using forward Euler with a large time step. In some sense stability means that the solution should be bounded as $\Delta t, h \to 0$ (e.g. it does not blow-up), but what exactly do we mean by bounded? We give two equivalent definitions of stability, the first is a practical definition (useful for deriving stability constraints), the second gives more mathematical insight.

Practical definition of Stability:
Consider the linear update

$$u^{n+1} = B(\Delta t)u^n + b^n(\Delta t).$$

Which we may want to consider writing as

$$B_1(\Delta t)u^{n+1} = B_0(\Delta t) + \Delta(t)f.$$

Then it follows that

$$u^{n+1} = B_1^{-1}B_0u^n + \Delta t B_1^{-1}f.$$

The method is Lax-Richtmyer Stable if for each time T there is constant C_T independent of Δt such that

$$\|B^n(\Delta t)\| \leq C_T \text{ for all } n\Delta t \leq T.$$

Note that stability involves a norm. Most analysis is done using the 2-norm and max-norm, but for conservation laws, sometimes the 1-norm is more natural. Practically this definition means that the solutions may grow in time, but not in the number of time steps (e.g. the solutions are bounded as $n \to \infty$ with $n\Delta t = T$ fixed).

Intuitive definition of Stability (equivalent to the practical definition):
Let w^n and v^n be two different solutions to $u^{n+1} = B(\Delta t)u^n + b^n(\Delta t)$. That is w^n and v^n both solve the same equation (e.g. we are considering solutions with different initial conditions).

The method is Lax-Richtmyer Stable if for each time T there is constant K_T, independent of Δt and the initial conditions, w^0 and v^0, such that

$$\|w^n - v^n\| \leq K_T \left\|w^0 - v^0\right\| \text{ for all } n\Delta t \leq T.$$

Essentially the idea is that solutions starting *close* together stay *close* together as Δt is refined. That is, the distance between

w^n and v^n is bounded by a constant times the distance between w^0 and v^0 as $\Delta t \to 0$ with $n\Delta t$ fixed.

What does it mean for a solution to a differential equation to be well-posed? A Well-posed problem means that the solution exists, the solution is unique and the solution depends continuously on the initial data. Note that stability of a difference scheme looks just like continuous dependence on the initial data for differential equations.

Lax Equivalence Theorem: Consistency and stability imply convergence.

Proof: Let $u^{n+1} = B(\Delta t)u^n + b^n(\Delta t)$ be a consistent and stable discretization of a linear (well-posed) PDE.

Let u^n_{sol} be the actual solution to the PDE evaluated at time $t_n = n\Delta t$. Think of this as being the column vector:

$$u^n = (u_{sol}(x_1, t_n), u_{sol}(x_2, t_n), ..., u_{sol}(x_N, t_n))'$$

Applying the difference scheme to the solution gives

$$u^{n+1}_{sol} = Bu^{n+1}_{sol} + b^n + \Delta t \tau^n,$$

where τ^n is the LTE at step n. Why the Δt times the LTE? For example, consider Forward Euler for the heat equation:

$$\frac{u^{n+1}_{sol} + u^n_{sol}}{\Delta t} = Lu^n_{sol} + f^n + \tau^n.$$

Now multiplying by Δt and rearranging:

$$u^{n+1}_{sol} = (I + \Delta t L)u^n_{sol} + \Delta t f^n + \Delta t \tau^n$$

The error (e.g. the difference between the numerical solution and the actual solution) is

$$e^n = u^n - u^n_{sol}.$$

Then subtracting the difference scheme as applied to u^n from the difference scheme applied to u^n_{sol} it follows that (note: this is where linearity is important, and it is the only thing we are using to get the following result)

$$e^{n+1} = Be^n - \Delta t \tau^n.$$

The solution to this difference equation is

$$\begin{aligned}
e^0 &= 0 \text{ because the initial condition is exact,}\\
e^1 &= -\Delta t \tau^0,\\
e^2 &= -\Delta t B \tau^0 - \Delta t \tau^1,\\
e^3 &= -\Delta t B^2 \tau^0 - \Delta t B \tau^1 - \Delta t \tau^2,\\
&\vdots\\
e^n &= -\Delta t \sum_{k=1}^{n} B^{n-k} \tau^{k-1}
\end{aligned}$$

Then, by applying triangle inequality it follows that

$$\begin{aligned}
\|e^n\| &= \Delta t \left\| \sum_{k=1}^{n} B^{n-k} \tau^{k-1} \right\|\\
&\le \Delta t \sum_{k=1}^{n} \left\| B^{n-k} \tau^{k-1} \right\|\\
&\le \Delta t \sum_{k=1}^{n} \left\| B^{n-k} \right\| \left\| \tau^{k-1} \right\|
\end{aligned}$$

Let $T = n\Delta t$. By the definition of stability, $\left\|B^{n-k}\right\| \le C_T$ for all $(n-k)\Delta t \le n\Delta t = T$. Therefore,

$$\begin{aligned} \|e^n\| &\le \Delta t C_T \sum_{k=1}^{n} \left\|\tau^{k-1}\right\| \\ &\le (n\Delta t) C_T \max_{1\le k\le n} \left\|\tau^{k-1}\right\| \\ &= T C_T \max_{1\le k\le n} \left\|\tau^{k-1}\right\| \to 0 \text{ as } \Delta t, h \to 0 \text{ by consistency.} \end{aligned}$$

□

The main idea of the Lax equivalence theorem, which you can see in the proof: the numerical scheme is going to introduce small errors, but if the scheme is stable then these errors will not get amplified in time.

Example 1: Show stability of the Crank-Nicolson Method in 2-norm for the diffusion equation with Dirichlet BC's on [0,1] (note L is the discrete Laplacian):

$$\frac{u^{n+1} - u^n}{\Delta t} = \frac{1}{2}(Lu^n + Lu^{n+1}) + f^{n+\frac{1}{2}}$$

First we need to rearrange so that we can show that the 2-norm of the update matrix is bounded:

$$\left(I - \frac{\Delta t}{2} L\right) u^{n+1} = \left(I + \frac{\Delta t}{2} L\right) u^n + \Delta t f^{n+\frac{1}{2}}$$

To solve for u^{n+1} we have to invert the matrix $(I - \frac{\Delta t}{2} L)$ on the LHS. We know that $(I - \frac{\Delta t}{2} L)$ has an inverse because the eigenvalues of the discrete Laplacian are nonpositive. Therefore all the eigenvalues of this matrix are greater than or equal to one, and the matrix is invertible. Thus:

$$u^{n+1} = \left(I - \frac{\Delta t}{2} L\right)^{-1} \left(I + \frac{\Delta t}{2} L\right) u^n + \left(I - \frac{\Delta t}{2} L\right)^{-1} \Delta t f^{n+\frac{1}{2}} = Bu^n + b^{n+\frac{1}{2}}$$

Now we need to show that: $\|B^n\|_2 \le C_T$ independent of Δt.

Since L is symmetric, $\|L\|_2$ is the largest e-value in absolute value. B is symmetric and $\|B\|_2$ is the largest e-value in absolute value AND the e-vectors of B are the e-vectors of L. The eigenvalues of L are

$$\lambda_k = \frac{2}{h^2}\Big(\cos(k\pi h) - 1\Big) = \frac{-4}{h^2}\sin^2\left(\frac{k\pi h}{2}\right).$$

Let $\mu_k = k^{th}$ e-value of B. It follows that

$$|\mu_k| = \left|\frac{1 + \frac{\Delta t}{2}\lambda_k}{1 - \frac{\Delta t}{2}\lambda_k}\right| < 1.$$

This implies that

$$\|B^n\|_2 \le \|B\|_2^n < 1 \text{ that is: } C_T = 1.$$

So Crank-Nicolson is stable in the 2-norm. Since we also know that the method is second order accurate in space and time, we know we have a second order convergent FD-scheme from the Lax-equivalence theorem.

Example 2: Show stability of Forward Euler in max-norm for diffusion equation:

$$\frac{u^{n+1} - u^n}{\Delta t} = Lu^n + f^n \Rightarrow u^{n+1} = Bu^n + \Delta t f^n \text{ where } (I + \Delta t L) = B.$$

Writing out the form of the matrix B,

$$B = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} + \frac{\Delta t}{h^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & -2 & 1 \\ & & & 1 & -2 \end{bmatrix}.$$

$\|B\|_\infty$ is just the max row sum in absolute value:

$$\|B\|_\infty = \frac{\Delta t}{h^2} + \left|1 - \frac{2\Delta t}{h^2}\right| + \frac{\Delta t}{h^2}.$$

Thus, we require

$$1 - \frac{2\Delta t}{h^2} \geq 0,$$

which is equivalent to

$$\Delta t \leq \frac{h^2}{2}.$$

With this restriction, we have that $\|B\|_\infty \leq 1$. This implies stability because

$$\|B^n\|_\infty \leq \|B\|_\infty^n = 1.$$