Math 228B, Lecture 5
January 20, 2009
Edward Tavernetti

Talking about consistency and what means to be convergent. The game is as follows 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 goig 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 ($\Delta t, h \rightarrow 0$). Today we will show that stability is also required to ensure convergence

Lax Equivalence Theorem (fundamental theorem of FD-methods): Stability + Consistency $\Rightarrow$ convergence (or: A consistent finite difference scheme for a well-posed problem is convergent if and only if it is stable.) But keep in mind that Lax Equivalence Theorem only applies to:

- Linear PDE
- Linear Schemes
- Well-posed PDE

Why is Lax Equivalence Theorem useful? Because, in general, showing convergence is hard, but it is easy to enforce consistency and then to show stability is comparatively easy as compared with analyzing convergence directly.

The natural question is then what does it mean to show stability?

Stability

Recall oscillations growing example from few weeks ago... in some sense stability implies that the solution should be bounded as $\Delta t, h \rightarrow 0$ (e.g. it does not blow-up).

So we have deffered the question of stability to a question of boundedness, but what do we mean by Bounded? We give two notions, the first is a practical definition (more useful for proofs), the second is an intuitive definition (offering more mathematical insight).

Practical definition of Bounded:

Consider the linear update:

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

Which we may want to consider writing like:

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

Then it follows:

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

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

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

This is what it means to be stable relative to a given norm (most analysis is done in the 2-norm, for the conservation laws the 1-norm is the more natural choice). Practically this means that the solutions do not continue to grow as time is refined (e.g. the solutions are in some sense bounded as $n \rightarrow \infty$).

Intuitive definition of Bounded (equivelent 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 (L-R) stable if for each time T there is constant K_T, that is 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 n is refined. That is, if w^0 and v^0 close then as you take more steps in time the distance between w^n and v^n is bounded by a constant independent of Δt.

What does it mean for a solution to a differential equation to be well-posed? Well, a well-posed problem means: the solution exists, the solution is unique and the solution depends continuously on the initial data (in a reasonable topology). We will find that boundedness *looks like* continuous dependence on the initial data.

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 discretization of a linear (well-posed) PDE and suppose that it is stable.

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:

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

Consider now if we apply a difference scheme to the solution: $u^{n+1}_{sol} = Bu^{n+1}_{sol} + b^n + \Delta t\tau^n$.

For example, applying Forward Euler for the heat equation:

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

Now multiplying by Δt and rearranging:

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

(Aside: for another example, if we were to show this for Crank-Nicholson work from: $(I - \frac{\Delta t}{2}L)u^{n+1} = (I + \frac{\Delta t}{2}L)u^n + \Delta t f^n$

Now the error (e.g. the difference between the numerical solution and the actual solution) is given by :

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

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

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

Now let's assume that (note that we have exponenets on B and indices on τ):

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

Then, by applying triangle inequality it follows:

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

Let $T = n\Delta t$. Then since by definition of stability we know: $\|B^{n-k}\| \leq C_T$ for all $(n-k)\Delta t \leq n\Delta t = T$, we derive the result that:

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

□

Think: 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 stabilty of the Crank-Nicolson Method in 2-norm for the diffusion equation with Dirchlet 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 bounded:

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

To solve for u^{n+1} we have to invert differential operator $(I - \frac{\Delta t}{2}L)$ on the LHS. We know that $(I - \frac{\Delta t}{2}L)$ has an inverse because the discrete Laplacian is inveritble and all the eigenvalues are negative (negative semi-definite matrix). Since on the LHS we have: 1 - (negative numbers), $(I - \frac{\Delta t}{2}L)$ will be postive and the e-values are going to all be bigger than 1 $\Rightarrow$ invertibility. Thus:

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

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

Since L is symmetric and $\|L\|_2$ 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. It follows $\lambda_k = \frac{2}{h^2}(\cos(k\pi h) - 1) = \frac{-4}{h^2}\sin^2(\frac{k\pi h}{2})$ (note these are the e-values of the discrete laplacian). Now setting $\mu_k = k^{th}$ 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:

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

So Crank-Nicoleson 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-Richtmyer Theorem.

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

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

Rewriting this in the familiar form we have:

$$B = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} + \begin{bmatrix} -2 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -2 & 1 \\ 0 & 0 & 0 & \dots & 1 & -2 \end{bmatrix}$$

So, by noting that $\|B\|_\infty$ 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: $1 - \frac{2\Delta t}{h^2} \geq 0$ to ensure stability (Note by rearranging terms this is: $\Delta t \leq \frac{h^2}{2}$
With: $\|B\|_\infty = 1$, then to complete the proof we see that: $\|B^n\|_\infty \leq \|B\|_\infty^n = 1$.