AMATH 442 (Fall 2014 - 1149)

Numerical Solutions of Partial Differential Equations

Prof. L. Krivodonona
University of Waterloo

LaTeXer: W. KONG
http://wwkong.github.io
Last Revision: October 6, 2014

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

ACKNOWLEDGMENTS:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in AMATH 442. The formal prerequisite to this course is either AMATH 351 or AMATH 350.

Errata

6-7 Assignments Biweekly

25% Assignments, 25% Midterm, 50% Final Exam

Office hours: W, Th @ 2-3pm

Midterm: Oct. 21st @ 4-5:30pm

1 Introduction

We begin with a quick review of the theoretical bases of **partial differential equations**.

1.1 Classification of 2nd Order Linear PDEs

There are 3 types of (linear) PDEs:

1. Parabolic PDEs (e.g. heat equation, diffusion equation)
   (a) Has the form $u_t = \sigma u_{xx}$ or in the multivariate case, $u_t = \sigma(u_{xx} + u_{yy} + u_{zz})$
2. Elliptic PDEs (e.g. Laplace's equation, Poisson equation)
   (a) Has the form $u_{xx} + u_{yy} = f(x, y)$ or $\Delta u = f(x, y)$
3. Hyperbolic PDEs (e.g. 1st, 2nd order wave equations)
   (a) Has the form $u_{tt} - c^2 u_{xx} = 0$ or $u_t + a u_x = 0$

There are also **non-linear PDEs**:

- Burger's equation: $u_t + u u_x = 0$
- Non-linear heat equation: $u_t = (\sigma(u) u_x)_x$
- Higher-order PDEs: $u_t + u u_x = \sigma u_{xxx}$
- Mixed types

1.2 Examples of Linear PDEs

(1) Let's begin by looking at the classic **linear advection (a.k.a. wave) equation**. The basic form is

$$u_t + a u_x = 0, a \in \mathbb{R}, (x, t) \in \mathbb{R} \times \mathbb{R}$$

Claim 1.1. Any $\phi(x - at)$ is a solution.

Proof. Substitution and chain rule:

$$u = \phi(x - at) \implies u_t = \phi'(x - at)(-a), u_x = \phi'(x - at) \implies -a\phi' + a\phi' = 0$$

Therefore ϕ is a solution. □

With the initial condition $u(x,0) = u_0(x)$, the solution is $u = u_0(x - at)$. The PDE with the aforementioned initial condition is called the **Cauchy problem**. We can interpret the parameter a as a speed parameter.

Now suppose that we introduce a finite domain $\Omega = [\alpha, \beta]$ and **boundary conditions**. Saying $u(\beta, t) = b_{right}(t)$ might lead to contradiction since $u_0(x-at) \neq b_{right}(\beta, t)$ or $u_0 = b_{right}$ (no new information is given). Instead, we provide $u(\alpha, t) = b_{left}(t)$ if $a > 0$ and we provide $u(\beta, t) = b_{right}(t)$ if $a < 0$.[1]

Conclusion 1. Here are some conclusions regarding the above wave equation:

[1] The solution of (1) does not grow or decay over time.

[2] New extrema can be introduced only through boundary conditions.

(2) Moving on, we have the **diffusion (heat) equation**. The basic form is

$$u_t = \sigma u_{xx}, \sigma \in \mathbb{R}$$

Assume that the initial conditions (I.C.) and boundary conditions (B.C.) are such that

$$u(x,t) = \hat{u}(k,t) \sin kx$$

is a solution, with k fixed. By substitution,

$$\begin{aligned} u_t &= \hat{u}_t \sin kx \\ u_x &= k\hat{u} \cos kx \\ u_{xx} &= -k^2 \hat{u} \sin kx \end{aligned}$$

and so

$$\hat{u_t} \sin kx = -\sigma k^2 \hat{u} \sin kx \implies \hat{u}_t = -\sigma k^2 \hat{u} \implies \hat{u}(k,t) = ce^{-\sigma k^2 t} \implies u(x,t) = ce^{-\sigma k^2 t} \sin kx$$

If we set $c = 1$ then the I.C. should be

$$u(x,0) = e^{-\sigma k^2 \cdot 0} \sin kx = \sin kx$$

If the domain is $\Omega = [-1, 1]$ and $k = \pi$ then the B.C. is

$$\begin{cases} u(-1,t) &= 0 \\ u(1,t) &= 0 \end{cases}$$

Remark 1.1. Here are some remarks about the solution:

[1] If $\sigma > 0$ then $u(x,t)$ decays with time (proper heat equation) and if $\sigma < 0$ then $u(x,t)$ grows with time (inverse or backwards heat equation). For this course, we always assume that $\sigma > 0$.

[2] The larger the σ, the faster the decay with respect to time. We call σ the **diffusion coefficient**.

[3] The larger the k, the faster the decay $\implies$ high frequencies decay faster.

2 Finite Difference Methods

Recall that

$$\begin{aligned} u_x &:= \lim_{\Delta x \to 0} \frac{u(x + \Delta x, t) - u(x,t)}{\Delta x} \\ \Delta^+ u &:= \frac{u(x + \Delta x, t) - u(x,t)}{\Delta x} \\ \Delta^- u &:= \frac{u(x,t) - u(x - \Delta x, t)}{\Delta x} \end{aligned}$$

[1] $x = \alpha$ is called inflow while $x = \beta$ is called outflow.

where we call the last two the **1st forward difference** and **1st backward difference** respectively. By convention, $\Delta x > 0$ and $\Delta t > 0$. Note that Δx is finite which is where the name "finite difference" comes from. u_x will be approximated by $\Delta^+ u$ or $\Delta^- u$. Similarly,

$$u_t \approx \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t}$$

We then introduce a discretization of space where

$$\begin{aligned} \Delta x_j &= x_{j+1} - x_j \\ \Delta t_n &= t_{n+1} - t_n \end{aligned}$$

For simplicity, assume uniform discretization. That is, $\Delta x_j = \Delta x, \Delta t_n = \Delta t$ for all j and t. In general $\Delta x \neq \Delta t$. We will use the notation

$$\begin{aligned} (x_j, t_n) &\equiv (j, n) \\ u_j^n &\equiv u(x_j, t_n) \end{aligned}$$

Finally, we denote the numerical solution as $U_j^n \approx u_j^n$. Now recall the Taylor series expansion of u about (x_j, t_n) in x:

$$\begin{aligned} u(x_j + \Delta x, t_n) &= u(x_j, t_n) + \Delta x u_x(x_j, t_n) + \frac{\Delta x^2}{2} u_{xx}(x_j, t_n) + ... \\ u(x_j - \Delta x, t_n) &= u(x_j, t_n) - \Delta x u_x(x_j, t_n) + \frac{\Delta x^2}{2} u_{xx}(x_j, t_n) - ... \end{aligned}$$

or more compactly,

$$\begin{aligned} u_{j+1}^n &= u_j^n + \Delta x \, (u_x)_j^n + \frac{\Delta x^2}{2} (u_{xx})_j^n + ... \\ u_{j-1}^n &= u_j^n - \Delta x \, (u_x)_j^n + \frac{\Delta x^2}{2} (u_{xx})_j^n - ... \end{aligned}$$

If we solve for $(u_x)_j^n$ in the first equation, then we get

$$(u_x)_j^n = \frac{u_{j+1}^n - u_j^n}{\Delta x} - \frac{\Delta x}{2} (u_{xx})_{j+\xi}^n, 0 < \xi < 1$$

by the mean value theorem. We call $\tau_j^n = -\frac{\Delta x}{2} (u_{xx})_{j+\xi}^n$ the **discretization (truncation) error**. Similarly from the second equation,

$$(u_x)_j^n = \frac{u_{j+1}^n - u_j^n}{\Delta x} \underbrace{+ \frac{\Delta x}{2} (u_{xx})_{j-\xi}^n}_{\tau_j^n}, 0 < \xi < 1$$

If we subtract the two equations together, then

$$(u_x)_j^n = \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} \underbrace{- \frac{1}{6} (u_{xxx})_{j+\xi}^n \Delta x^2}_{\tau_j^n}, 0 < \xi < 1$$

We call the first term on the right side the **1st central difference**. Central difference is more accurate than forward and backward difference. More accuracy is achievable with more points x_{j+2}, x_{j+3}. Adding the two equations will give us

$$(u_{xx})_j^n = \frac{u_{j+1} - 2u_j^n + u_{j-1}^n}{\Delta x^2} - \frac{\Delta x^2}{12} (u_{xxxx})_{j+\eta}^n, 0 < \eta < 1$$

In general, higher derivatives and more accurate approximations require more points (i.e. larger **stencil**).

Using **big-O notation**, we can write:

$$\begin{aligned}(u_x)_j^n &= \frac{u_{j+1}^n - u_j^n}{\Delta x} + O(\Delta x)\\ (u_t)_j^n &= \frac{u_j^{n+1} - u_j^n}{\Delta t} + O(\Delta t)\\ (u_{xx})_j^n &= \frac{u_{j+1} - 2u_j^n + u_{j-1}^n}{\Delta x^2} + O(\Delta x^2)\end{aligned}$$

Example 2.1. Let's construct a finite difference (FD) scheme for the heat equation:

$$\begin{aligned}u_t &= \sigma u_{xx}, -\infty < x < \infty\\ u(x,0) &= \phi(x)\end{aligned}$$

We have

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} = \sigma \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{\Delta x^2} \implies u_j^{n+1} = ru_{j-1}^n + (1-2r)u_j^n + ru_{j+1}^n$$

where $r = \sigma \Delta t / \Delta x^2$. If we know u_j^n for all j then we can compute u_j^{n+1} for all j. We need u_j^0 so we set $u_j^0 = u(x_j, 0) = \phi(x_j)$ for all j.

Let's plug in some values. Suppose that $\sigma = 1$ and choose the I.C. such that $u_0^0 = 1, u_j^0 = 0, \forall j \neq 0$ and $\Delta x = 1, \Delta t = 1/4 \implies r = 4$. This gives us

$$u_j^{n+1} = 4u_{j-1}^n - 7u_j^n + 4u_{j+1}^n$$

From stability analysis (CS 476), you will see that:

1. u_j^n grows

2. u_j^n oscillates (+ve, -ve, +ve, -ve, ...)

Instead, let's try: $\Delta x = 1/4, \Delta t = 1/64 \implies r = 1/4$ with:

$$u_j^{n+1} = \frac{1}{4}u_{j-1}^n + \frac{1}{2}u_j^n + \frac{1}{4}u_{j+1}^n$$

This will provide reasonable results. In general, we want u_0^n to be a good approximation of u_j^n.

Definition 2.1. A scheme is **convergent** on $0 < t \leq T$ if

$$\|u^n - U^n\| \to 0$$

as $\Delta x \to 0, \Delta t \to 0, n \to \infty, n\Delta t \leq T$. Here, $\|\cdot\|$ is some norm with u^n as a vector of all the (u_j^n)'s. A scheme is **convergent of order** k if

$$\|u^n - U^n\| = O(\Delta x^k)$$

Fact 2.1. *Convergence is difficult to prove directly. Instead, we look at:*

- Stability
- Convergence

Going back to our last example, consider $\|u\|_\infty = \max_j |u_j|$. From the general equation

$$\begin{aligned}|u_j^{n+1}| &\leq |r||u_{j-1}^n| + |1-2r||u_j^n| + |r||u_{j+1}^n|\\ &\leq (|r| + |1-2r| + |r|)\|u^n\|_\infty\end{aligned}$$

If $0 < r < \frac{1}{2}$ then $|u_j^{n+1}| \leq \|u^n\|_\infty, \forall j \implies \|u^{n+1}\| \leq \|u^n\|_\infty$. If $r > \frac{1}{2}$, then

$$|r| + |1-2r| + |r| = 2r - 1 + 2r = 4r - 1 \geq 1$$

and hence

$$|u_j^{n+1}| \leq (4r-1)\|u^n\|_\infty$$

Definition 2.2. A scheme is **stable** if $\exists C > 0$ **independent** of $\Delta x, \Delta t, u^0$ such that

$$\|u^n\| \leq C\|u^0\|, \Delta x \to 0, \Delta t \to 0, n \to \infty, n\Delta t \leq T$$

Note 1. (1) We allow some growth in the solution. Don't confuse this definition of stability with stability in ODE theory.

(2) Scheme is usually stable only for fixed values of some parameters. For example, Δt as a function of Δx or r.

In our example above, we showed that it was a stable scheme for the heat equation when $r < \frac{1}{2}$.

Definition 2.3. Alternatively, if u^n, v^n are solutions with $u^0 = \phi, v^0 = \psi$ (same problem, different I.C.), then a scheme is stable if $\exists C > 0$ independent of $\Delta x, \Delta t, u^0$ such that

$$\|u^n - v^n\| \leq C\|u^0 - v^0\|, \Delta x \to 0, \Delta t \to 0, n \to \infty, n\Delta t \leq T$$

Example 2.2. Going back to heat equation, suppose we choose I.C.

$$u^0 = (..., -1, 1, -1, 1, ...) \implies u_j^0 = (-1)^\gamma$$

and hence

$$\begin{aligned} u_j^1 &= 2r(-1)^{j+1} + (1-2r)(-1)^\gamma \\ &= (-1)^\gamma(-2r+1-2r) \\ &= -(4r-1)(-1)^\gamma \\ u_j^n &= (-1)^{\gamma+1}(4r-1)^n \end{aligned}$$

Taking norms, we have

$$\|u^n\|_\infty = (4r-1)^n\|u^0\|_\infty = (4r-1)^n$$

We call this **exponential growth** in the case of $r > \frac{1}{2}$. As $\Delta x, \Delta t \to 0$ with fixed T and $n \to \infty$, we have

$$\|u^n\|_\infty \to \infty$$

So with $r > \frac{1}{2}$, the results are **unstable**.

Remark 2.1. Stability for numerical methods is equivalent to **well-posedness** for PDEs:

- Solution exists given suitable I.C. and B.C.
- Solution is unique
- Solution is continuously independent on initial data

2.1 Consistency

We now change our notation so that U^n is the finite difference estimate and u^n is the exact solution. We want to know how much $u(x,t)$ satisfies the below equation

$$(2)\ \frac{U_j^{n+1} - U_n^n}{\Delta t} = \sigma\frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2}$$

which is a discretization of the heat equation. Note that $u(x,t)$ only exactly solves

$$(1)\ u_t = \sigma u_{xx}$$

Define

$$P(v) = \frac{v_j^{n+1} - v_n^n}{\Delta t} - \sigma\frac{v_{j+1}^n - 2v_j^n + v_{j-1}^n}{\Delta x^2}$$

We have $P(U^n) = 0$. Define $\tau_j^n \equiv P(u)$ where $u = u(x,t)$. We call τ_j^n the **truncation or discretization error**. Plug $u(x,t)$ into (2) to get

$$\begin{aligned}\tau_j^n &= \frac{u_j^n + \Delta t(u_t)_j^n + \frac{\Delta t^2}{2}(u_{tt})_j^n + O(\Delta t^3) - u_j^n}{\Delta t} - \\ &\sigma\frac{u_j^n + \Delta x(u_x)_j + \frac{\Delta x^2}{2}(u_{xx})_j^n + \frac{\Delta x^3}{6}(u_{xxx})_j^n + \frac{\Delta x^4}{24}(u_{xxxx})_j^n + O(\Delta x^5)}{\Delta x^2} - \\ &\sigma\frac{-2u_j^n + \left(u_j^n - \Delta x(u_x)_j + \frac{\Delta x^2}{2}(u_{xx})_j^n - \frac{\Delta x^3}{6}(u_{xxx})_j^n + \frac{\Delta x^4}{24}(u_{xxxx})_j^n + O(\Delta x^5)\right)}{\Delta x^2}\end{aligned}$$

This can be reduced to

$$\begin{aligned}\tau_j^n &= \underbrace{(u_t)_j^n - \sigma(u_{xx})_j^n}_{=0} + \frac{\Delta t}{2}(u_{tt})_j^n - \sigma\frac{\Delta x^2}{12}(u_{xxxx})_j^n + O(\Delta t^2) \\ &= \frac{\Delta t}{2}(u_{tt})_{j+\xi}^n - \sigma\frac{\Delta x^2}{12}(u_{xxxx})_{j+\eta}^n\end{aligned}$$

Suppose that the function $u(x,t)$ is smooth enough such that there exists M with the property $|u_{tt}|, |u_{xxxx}| \leq M$. We then get

$$|\tau_j^n| \leq M\left(\frac{\Delta t}{2} + \frac{\Delta x^2}{12}\right)$$

and hence $(\tau_j^n) = O(\Delta t, \Delta x^2)$. Since we need $r < \frac{1}{2}$ for stability, we have

$$\sigma\frac{\Delta t}{\Delta x^2} < \frac{1}{2} \implies \Delta t < \frac{\Delta x^2}{2\sigma} \implies (\tau_j^n) = O(\Delta x^2) \text{ if } r < \frac{1}{2}$$

Definition 2.4. A scheme is called **consistent** if $\tau_j^n \to 0$ as $\Delta x \to 0, \Delta t \to 0$. A scheme is called **consistent of order** k in Δx and m in Δt if

$$\tau_j^n = O(\Delta x^k, \Delta t^m)$$

Remark 2.2. Regarding the truncation error:

1. τ_j^n measures how far (2) is from (1)
2. τ_j^n is a purely analytical tool. Don't try to find it in your code!
3. τ_j^n is easy to compute $\implies$ the reason it is used
4. For many schemes (all ours) if $\tau_j^n = O(\Delta x^k, \Delta t^m)$ then

$$\|e^n\| = \|u^n - U^n\| = O(\Delta x^k, \Delta t^m)$$

Note 2. We note that

$$\begin{cases} U_j^{n+1} &= rU_{j-1}^n + (1-2r)U_j^n + rU_{j+1}^n \\ u_j^{n+1} &= ru_{j-1}^n + (1-2r)u_j^n + ru_{j+1}^n + \Delta\tau_j^n\Delta t \end{cases}$$

and hence

$$\begin{aligned} e_j^{n+1} &= re_{j-1}^n + (1-2r)e_j^n + re_{j+1}^n + \tau_j^n\Delta t \\ |e_j^{n+1}| &\leq (|r| + |1-2r| + |r|)\|e^n\| + \Delta t\|\tau^n\| \end{aligned}$$

and $r < \frac{1}{2}$ gives us

$$\begin{aligned}\|e^{n+1}\| &\leq \|e^n\| + \Delta t\|\tau^n\| \\ &\leq \|e^{n-1}\| + \Delta t(\|\tau^n\| + \|\tau^{n-1}\|) \\ &\leq \|e^0\| + \Delta t\sum_{k=1}^{n}\|\tau^k\|\end{aligned}$$

Since $\|e^0\| = 0$ because $U_j^0 = u_j^0$ if we let $\tau = \max \|\tau^k\|$, then

$$\|e^{n+1}\| \leq \Delta t \sum_{k=1}^{n} \tau = \Delta t \cdot n \cdot \tau = t_n \cdot \tau$$

and hence

$$\|e^{n+1}\| = \|u^n - U^n\| \leq \underbrace{t_n}_{\text{finite}} \cdot C(\Delta x^2 + \Delta t) \leq \bar{C}\Delta x^2$$

for some constants $C, \bar{C}$. We should expect quadratic convergence on smooth solutions of (1) using (2).

Remark 2.3. If $\tau_j^n = 0$ then $e_j^n = 0$ and hence if $u(x,t)$ is linear in time and cubic in space, then (2) solves (1) exactly.

Recall

1. Stability doesn't grow with time uncontrollably
2. Consistency gives the convergence (and its rate)
3. Convergence is *good*

Theorem 2.1. *(Lax Equivalence Theorem) We have*

$$\textit{Stability+Consistency} \iff \textit{Convergence}$$

The forward direction is easy to prove, while the reverse direction is difficult to prove. This is true for most (and of all of our) methods.

2.2 Convergence in Practice and Error Estimation

From the previous section, we saw that

$$\|e_j^n\| \sim C\Delta x^2 \implies \|e_j^n\| = O(\Delta x^2)$$

Suppose we have two meshes with Δx and $\frac{\Delta x}{2}$ and we know that in general $e_j^n = O(\Delta x^k)$. We then have

$$\begin{cases} \|e_{\Delta x}^n\| & \sim C_1 \Delta x^k \\ \|e_{\frac{\Delta x}{2}}^n\| & \sim C_2 \left(\frac{\Delta x}{2}\right)^k \end{cases} \implies \frac{\|e_{\Delta x}^n\|}{\|e_{\frac{\Delta x}{2}}^n\|} \sim \frac{C_1 \Delta x^k}{C_2 \left(\frac{\Delta x}{2}\right)^k} \implies \log_2 \frac{\|e_{\Delta x}^n\|}{\|e_{\frac{\Delta x}{2}}^n\|} \sim k, \ (C_1 \approx C_2)$$

Convergence Tests when $u(x,t)$ is not known

Suppose that we have three solutions $\{U_{\Delta x}, U_{\Delta x/2}, U_{\Delta x/4}\}$ which are methods of order k.

1. Pick a very fine mesh, say $\Delta x/64$ (arbitrary) and view if as an exact solution.
2. Consider

$$\log_2 \frac{\|U_{\Delta x} - U_{\Delta x/2}\|}{\|U_{\Delta x/2} - U_{\Delta x/4}\|} \leq \log_2 \frac{\|U_{\Delta x} - u\| + \|u - U_{\Delta x/2}\|}{\|U_{\Delta x/2} - u\| + \|u - U_{\Delta x/4}\|} \sim \log_2 \frac{C_1 \Delta x^k + C_2 \left(\frac{\Delta x}{2}\right)^k}{C_2 \left(\frac{\Delta x}{2}\right)^k + C_3 \left(\frac{\Delta x}{4}\right)^k}$$

   and simplifying with $C = C_1 \sim C_2 \sim C_3$, we we get

$$\log_2 \frac{C_1 \Delta x^k + C_2 \left(\frac{\Delta x}{2}\right)^k}{C_2 \left(\frac{\Delta x}{2}\right)^k + C_3 \left(\frac{\Delta x}{4}\right)^k} \sim \log_2 2^k = k$$

Richardson Extrapolation (Error Estimation)

Suppose we have two solutions $U_{\Delta x}^n, U_{\Delta x/2}^n$. Then,

$$U_{\Delta x}^n - U_{\Delta x/2}^n = (U_{\Delta x}^n - u^n) + (u_n - U_{\Delta x/2}^n) \approx c_1 \Delta x^k - c_2 \left(\frac{\Delta x}{2}\right)^k \approx C\left(1 - \frac{1}{2^k}\right)\Delta x^k$$

The error of the Δx grid is $e^n_{\Delta x} \approx C\Delta x^k$ and hence

$$e^n_{\Delta x} \sim C\Delta x^k \approx \frac{U^n_{\Delta x} - U^n_{\Delta x/2}}{1 - \frac{1}{2^k}}$$

Similarly for the $\Delta x/2$ grid, we have $e^n_{\Delta x/2} \approx C\left(\frac{\Delta x}{2}\right)^k$ and hence

$$\frac{U^n_{\Delta x/2} - U^n_{\Delta x/4}}{2^k\left(1 - \frac{1}{2^k}\right)} = \frac{U^n_{\Delta x/2} - U^n_{\Delta x/4}}{2^k - 1} \sim e^n_{\Delta x/2}$$

So $e_{\Delta x}$ is more reliable than $e_{\Delta x/2}$ but $e_{\Delta x/2}$ is an estimate for a better solution.

3 Von-Neumann Stability Analysis

This is a general tool applicable to schemes other than finite difference methods.

Review. Recall **Euler's formula**

$$e^{\beta i} = \cos\beta + i\sin\beta \implies \cos\beta = \frac{1}{2}(e^{\beta i} - e^{-\beta i}), \sin\beta = \frac{i}{2}(e^{-\beta i} - e^{\beta i})$$

Given

$$g(t) = e^{(\alpha + i\beta)} = e^{\alpha t}(\cos\beta t + i\sin\beta t)$$

we have that α is responsible for the growth in $g(t)$ w.r.t. (with respect to) time and β is the frequency.

Review. Suppose that $f(x)$ is on $[-\pi, \pi]$. Then the **Fourier series** (F.S.) of $f(x)$ is

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx)$$

Theorem 3.1. *If $f(x)$ is periodic and C^1 then the Fourier series of $f(x)$ converges to $f(x)$ in the infinity and L_2 norms.*

Consider the exponential for the F.S. using Euler's formula as a substitution:

$$\begin{aligned} f(x) &\sim \frac{a_0}{2} + \sum_{k=1}^{\infty}\frac{a_k}{2}(e^{kxi} + e^{-kxi}) + \sum_{k=1}^{\infty}\frac{ib_k}{2}(e^{-kxi} - e^{kxi}) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty}\underbrace{\frac{1}{2}(a_k - ib_k)}_{c_k}e^{kxi} + \sum_{k=1}^{\infty}\underbrace{\frac{1}{2}(a_k + ib_k)}_{c_{-k}}e^{-kxi} \\ &= \sum_{k=-\infty}^{\infty} c_k e^{kxi} \end{aligned}$$

It is easy to show that $c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-kxi}$. In the discrete version, we first choose a function $[-\pi, \pi] \mapsto [0, J]$ using

$$x(\xi) = \frac{2\pi}{J}\xi - \pi$$

Then

$$e^{kxi} = e^{\frac{2\pi\xi k}{J}i}e^{-\pi ki} = (-1)^k e^{\frac{2\pi k}{J}\xi i}$$

If we substitute this into the exponential form, then we get the F.S. on $[0, J]$:

$$f(\xi) \sim \sum_{k=-\infty}^{\infty} c_k(-1)^k e^{\frac{2\pi k}{J}\xi i} = \sum_{k=-\infty}^{\infty}\hat{c}_k e^{\frac{2\pi k}{J}\xi i}, \hat{c}_k = c_k(-1)^k$$

Now U_j^n is a discrete function defined at $x = x_j, j \in [0, J], \xi = \delta$. We claim that

$$(*)\ U_j = \sum_{k=0}^{J-1} A_k e^{\frac{2\pi k}{J} ji}$$

where we will call A_k the **discrete Fourier coefficients**. For the justification of (*), remark that:

1. The summation should be finite (stops at $k = J - 1$) because

   $$e^{\frac{2\pi J}{J} ji} = e^{2\pi ji} = e^{0ji} = 1$$

   Similar reasoning can be applied for any $k = J + s, 0 < s < J$.

2. If we rewrite (*) for U_j^n, then

   $$(**)\ U_j^n = \sum_{k=0}^{J-1} A_k^n w_j^k, w_j^k = e^{\frac{2\pi k}{J} ji}$$

   where A_k^n is time-indexed with a superscript and w_j^k is of degree k (power k).

3. (Orthogonality relation) Note that

   $$\sum_{j=0}^{J-1} w_j^k \bar{w}_j^m = \begin{cases} J & k \equiv m \mod J \\ 0 & \text{otherwise} \end{cases}$$

   Multiply $(**)$ by $\bar{w}_j^m$ and sum over j (m is fixed) to get

   $$\sum_{j=0}^{J-1} U_j^n \bar{w}_j^m = \sum_{j=0}^{J-1} \sum_{k=0}^{J-1} A_k^n w_j^k \bar{w}_j^m = \sum_{j=0}^{J-1} A_k^n \sum_{k=0}^{J-1} w_j^k \bar{w}_j^m = J A_m^n$$

   and hence

   $$A_m^n = \frac{1}{J} \sum_{j=0}^{J-1} U_j^n \bar{w}_j^m$$

4. (Discrete Parseval's Relation) It follows from above that

   $$\|U^n\|_2^2 = J \|A^n\|_2^2$$

   which follows from orthogonality. Compare this with the continuous case (very similar).

Remark 3.1. For the general heat equation $u_t = \sigma u_{xx} + f(x)$, if $u(x, t)$ tends to the $\bar{u}(x)$, called the **steady state**, then

$$\frac{\partial}{\partial t} u(x, t) = \frac{\partial}{\partial t} \bar{u}(x) = 0$$

and $\sigma u_{xx} = -f(x)$ which is an elliptic equation. Elliptic equations can be viewed as a steady state of parabolic equations.

Example. Here is the Von Neumann analysis applied to

$$(1)\ u_t + a u_x = 0$$

with periodic boundary conditions. In this problem, we want to find the stability condition for (1) (if any). Recall that

$$U_j^n = \sum_{k=0}^{J-1} A_k^n w_j^k, 0 \le j \le J$$

Note that we require periodic boundary conditions to allow the Fourier series to converge. One of the many FDMs for (1) is

$$(2)\ \frac{U_j^{n+1} - U_j^n}{\Delta t} + a \frac{U_j^n - U_{j-1}^n}{\Delta x} = 0, U_0^n = U_j^n$$

We call this scheme is FTBS. Rewriting, we have

$$U_j^{n+1} = (1-\alpha)U_j^n + \alpha U_{j-1}^n, \alpha = \frac{a\Delta t}{\Delta x}$$

Plug in the F.S. expansion to get

$$\sum_{k=0}^{J-1} A_k^{n+1} w_j^k = \sum_{k=0}^{J-1} \left((1-\alpha)A_k^n w_j^k + \alpha A_k^n w_{j-1}^k\right)$$

Collect terms with w_j^k with the fact that

$$w_{j-1}^k = e^{\frac{2\pi i}{J}kj} e^{-\frac{2\pi i}{J}k} = w_j^k e^{-\frac{2\pi i}{J}k}$$

to get

$$\sum_{k=0}^{J-1} \left(A_k^{n+1} - \left[(1-\alpha)A_k^n + \alpha A_k^n e^{\frac{2\pi i}{J}k}\right]\right) w_j^k = 0 \implies A_k^{n+1} = \underbrace{\left[(1-\alpha) + \alpha e^{-\frac{2\pi i}{J}k}\right]}_{M_k} A_k^n$$

by linear independence.[2] By recurrence,

$$A_k^{n+1} = (M_k)^{n+1} A_k^0 \implies U_j^n = \sum_{k=0}^{J-1} (M_k)^n A_k^0 w_j^k$$

and hence by Parseval's identity

$$\|U^n\|_2^2 = J \sum_{k=0}^{J-1} |M_k|^{2n} |A_k^0|^2$$

If $|M_k| \leq 1$ for all k then

$$\|U^n\|_2^2 \leq J \sum_{k=0}^{J-1} |A_k^0| = \|U^0\|_2^2$$

So U^n is stable in 2-norm with $C = 1$. Since the exact solution of (1) doesn't grow in time, it is reasonable to require the same from U_j^n, i.e. $C = 1$.

$$\|U^n\| \leq C\|U^0\|$$

Note 3. Say we have $k = \hat{k}$ such that $M_{\hat{k}} > 1$ and $M_k \leq 1, \forall k \neq \hat{k}$. Then the corresponding wave will grow in amplitude and dominate over the other smaller waves.

So now we want to find α such that $|M_k| \leq 1$. Instead, look for $|M_k|^2 \leq 1$ with

$$\begin{aligned}
|M_k|^2 = M_k \bar{M}_k &= (1 - \alpha + \alpha\cos\theta)^2 + (\alpha \sin\theta)^2, \theta = \frac{2\pi k}{J} \\
&= 1 - 2\alpha + \alpha^2 + 2\alpha(1-\alpha)\cos\theta + \alpha^2\cos^2\theta + \alpha^2 \sin\theta \\
&= 1 - 2\alpha + 2\alpha^2 + 2\alpha(1-\alpha)\cos\theta \\
&= 1 - 2\alpha(1-\alpha) + 2\alpha(1-\alpha)\cos\theta \\
&= 1 - 2\alpha(1-\alpha)(1-\cos\theta)
\end{aligned}$$

So

$$0 \leq |M_k|^2 = 1 - 4\alpha(1-\alpha)\sin^2\frac{\theta}{2} \leq 1$$

Since $\alpha = \frac{a\Delta t}{\Delta x}$ then our stability condition is

$$(*)\ \Delta t \leq \frac{\alpha \Delta x}{a}, 0 < \alpha \leq 1$$

for (1)-(2) if $a > 0$. If $a < 0$, then (2) is unstable for all $\Delta x, \Delta t$. We call (*) the **CFL** (Courant-Friedrichs-Lewy) condition which is a stability restriction on time step size Δt for hyperbolic problems.

Remark 3.2. Consider $\alpha = 1$ where $|M_k|^2 = 1 \implies$ no amplitude loss and and exact propagation of the initial profile. That

[2] This is due to the fact that the $w_j's$ form a linear independent basis.

is, $U_j^{n+1} = U_{j-1}^n$. If $a < 0$, we can show that

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + a\frac{U_{j+1}^n - U_j^n}{\Delta x} = 0$$

is stable with (*) under $\Delta t \leq \frac{\alpha \Delta x}{|a|}$ with $0 \leq \alpha \leq 1$.

Definition 3.1. A FDM (finite difference method) satisfies the **Von Neumann condition** if $\exists C > 0$ independent of $\Delta x, \Delta t, k$ such that

$$|M_k| \leq 1 + C\Delta t, \forall \Delta t \leq \overline{\Delta t}, \Delta x \leq \overline{\Delta x}$$

Theorem 3.2. *A constant coefficient scalar one-level FDM is stable in the 2-norm iff it satisfies the Von Neumann conditions.*

Proof. ($\Longleftarrow$) Suppose U^n satisfies the Von Neumann conditions. Then,

$$\begin{aligned} \|U^n\|_2^2 &= J\sum_{k=0}^{J-1} |M_k|^{2n} |A_k^0|^2 \\ &\leq (1 + c\Delta t)^{2n} \|U^0\|_2^2 \end{aligned}$$

Now recall that

$$e^x = 1 + x + \frac{x^2}{2} + ... \implies 1 + x \leq e^x$$

and hence

$$(1 + c\Delta t)^{2n} \|U^0\|_2^2 \leq e^{2c\underbrace{\Delta t n}_{\Delta t_n}} \|U^0\|_2^2 \leq e^{2cT} \|U^0\|_2^2 \leq \bar{C}\|U^0\|_2^2$$

where $0 \leq t_n \leq T$ where T is the final time. □

($\Longrightarrow$) Suppose the scheme is stable and $\exists k = k^*$ such that $|M_{k^*}| > (1 + c\Delta t), \forall c$. Choose I.C. such that $A_{k^*}^0 \neq 0, A_k^0 = 0, k \neq k^*$. Then $U_j^0 = A_{k^*}^0 w_k^{k^*}$ and

$$U_j^n = (M_{k^*})^n A_{k^*}^0 w_j^{k^*} = (M_{k^*})^n U_j^0$$

Hence,

$$\|U^n\|_2^2 = |M_{k^*}|^{2n} \|U^0\|_2^2 > (1 + c\Delta t)^{2n} \|U^0\|_2^2$$

which implies that $\|U^n\|_2^2$ cannot be bounded by $\bar{C}\|U^0\|_2^2$ and hence is unstable. This is impossible and thus the scheme must satisfy the Von Neumann condition.

4 Implicit Methods

Recall the stability condition for the discretized heat equation

$$(1)\ u_t = \sigma u_{xx}$$

which was

$$(2)\ \frac{U_j^{n+1} - U_j^n}{\Delta t} = \sigma \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2}, r = \frac{\sigma \Delta t}{\Delta x^2} \leq \frac{1}{2} \implies \Delta t \leq \frac{\Delta x^2}{2\sigma}$$

This is very restrictive and seldom used in practice. For example, if $\Delta x = 10^{-3}$ then $\Delta t \approx 10^{-6}$ and if $T = 1$ then $N = 10^6$. Consider

$$(3)\ \frac{U_j^{n+1} - U_j^n}{\Delta t} = \sigma \frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{\Delta x^2}$$

where

$$(4)\ U_j^n = -rU_{j+1}^{n+1} + (1 + 2r)U_j^{n+1} - rU_{j-1}^{n+1}, \tau_j^{n+1} = O(\Delta x^2, \Delta t)$$

Substitute $U_j^n = \sum_{k=0}^{J-1} A_k^n w_j^k$ into (4) to get

$$\sum_{k=0}^{J-1} A_k^n w_j^k = \sum_{k=0}^{J-1} \left(-rA_k^{n+1} w_{j+1}^k + (1+2r)A_k^{n+1} w_j^k - rA_k^{n+1} w_{j-1}^k\right)$$

and factoring out w_j^k (and matching coefficients) we get

$$A_k^n = \underbrace{\left(-re^{\frac{2\pi k}{J}i} + (1+2r) - re^{-\frac{2\pi k}{J}i}\right)}_{\equiv M_k^{-1}} A_k^{n+1} \implies A_k^{n+1} = M_k A_k^n$$

where

$$\begin{aligned} M_k^{-1} &= (1+2r) - 2r\cos\frac{2\pi k}{J} \\ &= 1 + 2r\left(1 - \cos\frac{2\pi k}{J}\right) = 1 + 4r\left(\sin\frac{\pi}{J}\right)^2 \end{aligned}$$

So $M_k^{-1} \geq 1, \forall r > 0, r = \frac{\sigma\Delta t}{\Delta x^2} \implies M_k \leq 1$ and (3) will be unconditionally stable. In practice, Δt is taken to be $O(\Delta x)$ for unconditionally stable schemes. However, (2) is $O(\Delta x^2)$ accurate ($r < 1/2$) and (3) is only $O(\Delta x)$ with $\Delta t \approx \Delta x$.

4.1 Crank-Nicholson Method

The **Crank-Nicholson** (CN) method is defined as

$$(5)\ \frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{\sigma}{2}\left[\frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{\Delta x^2} + \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2}\right]$$

CN is unconditionally stable (see notes). It can be shown that $\tau_j^n = O(\Delta x^2, \Delta t^2)$:

$$\begin{aligned} \tau_j^n &= (u_t)_j^n + \frac{\Delta t}{2}(u_{tt})_j^n + O(\Delta t^2) - \frac{\sigma}{2}\left[(u_{xx})_j^{n+1} + O(\Delta x^2) + (u_{xx})_j^n + O(\Delta x^2)\right] \\ &= (u_t)_j^n + \frac{\Delta t}{2}(u_{tt})_j^n - \frac{\sigma}{2}\left[(u_{xx})_j^n + \Delta t(u_{xxt})_j^n + (u_{xx})_j^n + O(\Delta x^2) + O(\Delta t^2)\right] + O(\Delta t^2) \end{aligned}$$

Now $u_t = \sigma u_{xx} \implies u_{tt} = \sigma u_{xxt}$ and the result follows.

Solution Algorithm for Crank-Nicholson (CN) Method

Consider the heat equation with the following conditions:

$$\begin{aligned} (1)\ u_t = \sigma u_{xx} &\quad \text{on } x \in (\alpha, \beta) \\ u(x,0) = u_0(x) &\quad \text{I.C} \\ \begin{cases} u(\alpha,t) = f_l(t) \\ u(\beta,t) = f_r(t) \end{cases} &\quad \text{Diriclet B.C} \end{aligned}$$

and the scheme

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{\sigma}{2}\left[\frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{\Delta x^2} + \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2}\right]$$

Rewrite CN as

$$(2)\ U_j^{n+1} - \frac{r}{2}\left(U_{j-1}^{n+1} - 2U_j^{n+1} + U_{j+1}^{n+1}\right) = U_j^n + \frac{r}{2}\left(U_{j-1}^n - U_j^n + U_{j+1}^n\right), 1 \leq j \leq J-1$$

$$U_0^n = f_l(t_n) = f_l^n; U_J^n = f_r(t_n) = f_r^n$$

Rewrite (2) as a matrix where (2) is a system of $J-1$ linear equations. Let

$$C_{ij} = M_{ij} = \begin{cases} 1 & |i-j| = 1 \\ -2 & i = j \\ 0 & \text{otherwise} \end{cases}, f^n = \begin{cases} f_l^n & i = 1 \\ f_r^n & i = J \\ 0 & \text{otherwise} \end{cases}$$

Then the system (2) can be rewritten as

$$\begin{aligned} U^{n+1} - \frac{r}{2}CU^{n+1} - \frac{r}{2}f^{n+1} &= U^n + \frac{r}{2}CU^n + \frac{r}{2}f^n \\ \implies \underbrace{\left(I - \frac{r}{2}C\right)}_{A} U^{n+1} &= \underbrace{\left(I + \frac{r}{2}C\right)U^n + \frac{r}{2}\left(f^n + f^{n+1}\right)}_{F} \\ \implies AU^{n+1} &= F \end{aligned}$$

and the last equation is solvable using linear algebra methods. Remark that A is a sparse tridiagonal matrix with $\sim 3(J-1)$ non-zero elements. This will make A^{-1} dense with $(J-1)^2$ non-zero elements.

Tridiagonal Algorithm

Suppose we have $AX = F$ with $A \in \mathbb{R}^{N\times N}$ being tridiagonal. Consider the LU decomposition

$$(1)\ L\underbrace{UX}_{y} = F, A = LU \implies (2)\ Ly = F,\ (3)\ UX = y$$

Suppose that

$$A_{ij} = \begin{cases} b_{ij} & j - i = 1 \\ a_{ij} & i = j \\ c_{ij} & i - j = 1 \\ 0 & \text{otherwise} \end{cases} \implies L_{ij} = \begin{cases} l_{ij} & j - i = 1 \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases}, U_{ij} = \begin{cases} u_{ij} & i = j \\ v_{ij} & i - j = 1 \\ 0 & \text{otherwise} \end{cases}$$

We need to find u_j, v_j, l_j. The first row and first column, which we denote by $R_1 \cdot C_1$, of the LU decomposition implies

$$u_1 \cdot 1 = a_1$$

Similarly,

$$R_2 \cdot C_1 \implies l_2 u_1 = b_2 \implies l_2 = b_2/u_1$$

and in general,

$$R_j \cdot C_1 \implies l_j = b_j/u_{j-1}$$

With the first row and second column, using the same logic,

$$R_1 \cdot C_2 \implies v_1 \cdot 1 = c_1$$

$$R_2 \cdot C_j \implies l_2 v_1 + u_2 = a_2 \implies u_2 = a_2 - l_2 v_1$$

$$R_j \cdot C_2 \implies u_j = a_j - l_j v_{j-1}$$

This is LU factorization that takes into account the sparsity of A.

Note 4. Pivoting is usually not necessary for matrices arising in FD & FE (**finite element**) discretizations. For the number of operations, we have

$$\begin{aligned} 3 \text{ arithmetic} \times N &\sim 3N \\ (3 \text{ assignments of values} &\sim 3N) \end{aligned}$$

So it is $O(N)$ in the number of operations. Forward substitution for the heat equation gives us

$$y_j = f_j - l_j y_{j-1}$$

and we can use backward substitution to solve $UX = y$.

Summary 1. We have the following description of our discretizations:

	Explicit	Implicit	CN
Accuracy	$O(\Delta x^2, \Delta t)$	$O(\Delta x^2, \Delta t)$	$O(\Delta x^2, \Delta t^2)$
Stability	$\frac{\sigma \Delta t}{\Delta x^2} < \frac{1}{2}$	Unconditionally Stable	Unconditionally Stable
Work Per Time Step	$O(J)$	$O(J)$	$O(J)$
Total Work	$O(J \times N) = O(J^3)$	$O(J \times N) = O(J^2)$	$O(J \times N) = O(J^2)$
Reason (above)	$r < \frac{1}{2} \implies \Delta t \sim \Delta x^2$	since$\Delta t \sim \Delta x$	since$\Delta t \sim \Delta x$

Note 5. Even though CN is unconditionally stable, Δt should be about Δx for:

1) Accuracy

2) Convergence of iterative solvers for $AX = F$

Conclusion 2. Implicit schemes are more expensive per time step[3], but they might be more efficient (more than explicit schemes) overall if the number of time steps is smaller.

Boundary Conditions

1. Dirichlet B.C. are B.C. on $u(x,t)$
2. Neumann B.C. are B.C. on u_x
   (a) $u_x(-1,t) = u(1,t) = 0$ imply that the ends are insulated and no heat enters or leaves
   (b) Consider $u_x(\alpha, t) = g_e(t)$ for $u_t = \sigma u_{xx}$
   (c) Method 1:
      i. $(U_x)_0^n = g_e^n$, $\frac{U_1^n - U_0^n}{\Delta x} = g_e^n \implies U_0^n = U_1^n - \Delta x g_e^n$
      ii. At $j = 1$,

$$\begin{aligned} U_0^n - 2U_1^n + U_2^n &= U_1^n - \Delta x g_e^n - 2U_1^n + U_2^n \\ &= -U_1^n + U_2^n - \Delta g_e^n \end{aligned}$$

      iii. The first line in C is $AU + f$ where

$$A_{ij} = \begin{cases} -1 & (i,j) = (1,1) \\ 1 & |i-j| = 1 \\ -2 & i = j \end{cases}, f_i = \Delta x g_e^n$$

   (d) Method 2:
      i. Use higher order approximation for u_x where

$$(U_x)_0^n = \frac{-3U_0^n + 4U_1^n - U_2^n}{2\Delta x} = g_e^n \implies U_0^n = \frac{-(2\Delta x g_e^n - 4U_1^n + U_2^n)}{3}$$

      Eliminating U_0^n from the approximation of $(*)$ u_{xx} as in $(*)$.
3. Robin B.C. is in the form $\alpha u(1,t) + \beta u_x(1,t) = f(t)$
4. Mixed B.C. is when you have one half Dirichlet and one half Neumann

[3]This is due to the fact that in implicit schemes, you need to create a system of equations and solve it per time step. You will need to code this and it WILL take quite a bit of time. So efficiency here refers to amount of time invested.

Appendix

How to Check your Code

1. Manufacture a problem for which you know the exact solution and which you should be able to solve exactly.
   (a) In the heat equation, we could try $u(x,t) = 1$ with I.C. $u(x,0) = 1$ and B.C. $u(1,t) = 1, u(0,t) = 1$.