Asymptotic Behavior $f(t)$ is amyptotic to t means the following:

$$f(t) \sim t^2 \text{ as } t \to 0 \text{ means } t^{-2} f(t) \to 0 \text{ as } t \to 0$$

Equivalently, we write $f(t) = t^2 + o(t^2)$. Also, it could be the case:

$$f(t) = O(t^2) \text{ means } t^{-2} f(t) \text{ is bounded as } t \to 0$$

1 Floating Point Arithmetic

A **Floating Point Number System** is a finite subset of the reals defined by $\mathbb{F}(b, K, m, M)$ where b is the base of the system, K is the number of digits, m is the smallest exponent representable and M is the largest exponent representable.
If $y \in \mathbb{F}(b, K, m, M)$, then

$$y = \pm \left(0.d_1 d_2 d_3 .... d_K\right)_b \times b^E, m \leq M, d_1 \neq 0 \iff y = 0$$

1.1 Round-off Error

The error in representing $z \in \mathbb{R}$ by its nearest element in $\mathbb{F}(b, K, m, M)$. If $z \in \mathbb{R}$, then

$$fl(z) = \begin{cases} \pm(0.d_1 d_2 ... d_K)_b \times b^E & d_{k+1} < \frac{b}{2} \\ \pm\left[(0.d_1 d_2 ... d_K)_b + b^{-K}\right] \times b^E & d_{k+1} \geq \frac{b}{2} \end{cases}$$

IEEE Double precision (Used in MATLAB) is $b = 2$ and $K = 52$. In base 10, this is approximately $K \approx 16$, $m \approx -308$, and $M \approx 308$.

1.2 Relative error in rounding

Let $y \in \mathbb{R}, y \neq 0$. $fl(y) \in \mathbb{F}(b, K, m, M)$. Assume $d_{k+1} \leq \frac{b}{2}$

$$|fl(y)| = (0.d_1 d_2 ... d_K)_b \times b^E$$

The **Relative error** is

$$\text{Rel error: } \frac{|y - fl(y)|}{|y|}$$

Since $|y| = (0.d_1 d_2 ... d_k d_{k+1} ...)_b \times b^E \geq (0.1)_b \times b^E = b^{E-1}$
and $|y - fl(y)| = (0.d_{k+1 d_{k+2} .... b} \times b^{E-k} \leq \frac{1}{2} b^{E-k}$ thus

$$\text{Rel error: } \frac{|y - fl(y)|}{|y|} \leq \frac{1}{2} b^{1-K} =_{\text{machine}}$$

This **Machine Epsilon** is the smallest representable number In IEEE DP, $b = 2$, and $K = 52$, so $_{\text{machine}} = 2^{-52} \approx 2.2204 \times 10^{-16}$

1.3 Error in Computation

1.3.1 Finite Difference Operators

Recall
$$f''(x) - \frac{1}{h^2}\left(f(x-h) - 2f(x) + f(x+h)\right) = -\frac{1}{12}h^2 f^{(4)}(\xi)$$
Assume the round off error $e(x-h)$ is in the evaluation of function values $f(x-h) = \tilde{f}(x-h) + e(x-h)$.
$$f''(x) - \frac{1}{h^2}\left(\tilde{f}(x-h) - 2\tilde{f}(x) + \tilde{f}(x+h)\right) = E_h = -\frac{1}{12}h^2 f^{(4)}(\xi) + \frac{1}{h^2}\left(e(x-h) - 2e(x) + e(x+h)\right)$$
$$|E_h| \leq \frac{1}{12}h^2\left|f^{(4)}(\xi)\right| + \frac{1}{h^2}\left(|e(x-h)| + |2e(x)| + |e(x+h)|\right)$$
Assume $\left|f^{(4)}(\xi)\right| \leq M$ for $\xi \in [x-h, x+h]$ and assume $|e(x)| \leq$ for $x \in [x-h, x+h]$.
$$|E_h| \leq \frac{1}{12}h^2 M + \frac{1}{h^2}4$$
The first term shrinks but the second term blows up as $h \to 0$. One hopes to find the minimum at
$$h_{optimal} = \left(\frac{48}{M}\right)^{\frac{1}{4}}$$
We could take to be $\epsilon_{machine}$

2 Polynomial Approximations

2.1 Taylor Expansion Theorem

The Taylor series expansion for a function f centered at α evaluated at z is
$$f(z) = \sum_{k=0}^{\infty} a_k (z-\alpha)^k \qquad \text{where } a_k = \frac{f^{(k)}(\alpha)}{k!}$$
A Taylor polynomial is any finite truncation of this series:
$$f(z) = \sum_{k=0}^{N} a_k (z-\alpha)^k \qquad \text{where } a_k = \frac{f^{(k)}(\alpha)}{k!}$$
The Taylor series is the limit of the Taylor Polynomials, given that the limit exists.

Analytic Functions A function that is equal to its Taylor Series in an open interval (or open disc in the complex plane), is known as an **Analytic Function**

Maclaurin Series If the Taylor series or Polynomial is centered at the origin ($\alpha = 0$), then it is also a MacLaurin series.

2.1.1 Important Taylor Series

The Maclaurin series for $(x-1)^{-1}$ is
$$(x-1)^{-1} = 1 + x + x^2 + x^3 + ... = \sum_{k=0}^{\infty} x^k$$

4 Solving $Ax = b$

4.1 Tridiagonal Solver

For a tridiagonal system of equations

$$A\vec{\mathbf{u}} = \vec{\mathbf{f}}, \qquad A \text{ tridiagonal}$$

take $b_1 = c_n = 0$ and

$$A = LU = \begin{pmatrix} a_1 & c_1 & & \\ b_2 & a_2 & c_2 & \\ & \ddots & \ddots & \ddots \\ & & b_n & a_n \end{pmatrix} = \begin{pmatrix} 1 & & & \\ \beta_2 & 1 & & \\ & \ddots & \ddots & \\ & & \beta_n & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 & c_1 & & \\ & \alpha_2 & c_2 & \\ & & \ddots & \ddots \\ & & & \alpha_n \end{pmatrix}$$

So to solve $LU\vec{\mathbf{u}} = \vec{\mathbf{f}}$

1. Solve $L\vec{\mathbf{v}} = \vec{\mathbf{f}}$ using Forward Substitution
2. Solve $U\vec{\mathbf{u}} = \vec{\mathbf{v}}$ using Backward Substitution

4.1.1 Pseudocode

INPUT: $\vec{\mathbf{a}}, \vec{\mathbf{b}}, \vec{\mathbf{c}}, \vec{\mathbf{f}}$, all length n LU decomposition:
$\alpha_1 = a_1$
for $k = 2$ to n
    $\beta_k = b_k \setminus \alpha_{k-1}$
    $\alpha_k = a_k - \beta_k c_{k-1}$
end
Forward Substitution:
$v_1 = f_1$
for $k = 2$ to n
    $v_k = f_k - \beta_k v_{k-1}$
end
Backward Substitution:
$u_n = v_n \setminus \alpha_n$
for $k = 2$ to n
    $j = (n+1) - k$
    $u_j = (v_j - c_j u_{j+1}) \setminus \alpha_j$
end
Operation count: $O(n)$

4.2 Spectral Decomposition Method

If $A \in \mathbb{C}^{m\times m}$ is Hermitian, we can do the following

1. Compute the spectral decomposition (not trivial when m is large)

$$A = UDU^*$$

$$A\vec{\mathbf{x}} = UDU^*\vec{\mathbf{x}} = \vec{\mathbf{b}}$$

So we see

$$\vec{\mathbf{x}} = \alpha_1\vec{\mathbf{u}}_1 + \alpha_2\vec{\mathbf{u}}_2 + ... + \alpha_m\vec{\mathbf{u}}_m \implies U^*\vec{\mathbf{x}} = \vec{\alpha}$$

and

$$\vec{\mathbf{b}} = \beta_1\vec{\mathbf{u}}_1 + \beta_2\vec{\mathbf{u}}_2 + ... + \beta_m\vec{\mathbf{u}}_m \implies \beta_k = \vec{\mathbf{u}}_k^*\vec{\mathbf{b}}$$

so

$$U^*\vec{\mathbf{b}} = \vec{\beta} \text{ and } D\vec{\alpha} = \vec{\beta} \implies \vec{\alpha} = D^{-1}\vec{\beta}$$

but since $D_{ii}^{-1} = \frac{1}{\lambda_i}$,

$$\alpha_k = \frac{\vec{\mathbf{u}}_k^*\vec{\mathbf{b}}}{\lambda_k} \implies \vec{\mathbf{x}} = \sum_{k=1}^{m}\left(\frac{\vec{\mathbf{u}}_k^*\vec{\mathbf{b}}}{\lambda_k}\right)\vec{\mathbf{u}}_k$$

5 Finite Differences

Finite Differences seeks to approximate an ODE or PDE over a **mesh** or **grid**. The steps involved are:

1. Discretize the PDE using a difference scheme.
2. Solve the discretized PDE by iterating and/or time stepping.

5.1 Meshes

5.1.1 Uniform Meshes

Given a closed domain $\Omega = \bar{R} \times [0, t_F]$, we divide it into a $(J+1) \times (N+1)$ grid of parallel lines. Assume $\bar{R} = [0, 1]$. Given the mesh sizes $\Delta x = \frac{1}{J}, \Delta t = \frac{1}{N}$, a **mesh point** is

$$(x_j, t_n) = (j\Delta x, n\Delta t) \qquad j = 0, ..., J \qquad n = 0, ..., N$$

and $x_0 = 0$, $x_n = 1$
An alternative convention uses a $(J+2) \times (N+2)$ grid with the mesh sizes $\Delta x = \frac{1}{J+1}, \Delta t = \frac{1}{N+1}$. $x_0 = 0$, $x_{n+1} = 1$ are the boundary points. and

$$(x_j, t_n) = (j\Delta x, n\Delta t) \qquad j = 0, ..., J+1 \qquad n = 0, ..., N+1$$

We seek approximations to the solution at these mesh points, denoted by

$$U_j^n \approx u(x_j, t_n)$$

Where initial values are exact from the initial value function $u^0(x, t) = u(x, 0)$

$$U_j^0 = u^0(x_j) \qquad j = 1, ..., J-1$$

and boundary values are exact from the boundary value functions $f(t) = u(0, t)$ and $g(t) = u(1, t)$

$$U_0^n = f(t_n) \qquad U_J^n = g(t_n) \qquad n = 1, 2, ...,$$

5.2 Difference Coefficients

$$D_+, D_-$$

5.3 Explicit Scheme

A scheme is **explicit** if the solution at the next iteration (time level t_{n+1}) can be written as a single equation involving only previous time steps. This is, if it can be written in the form

$$U_j^{n+1} = \sum_i \sum_{k \le n} a_{i,k} U_i^k + b_{i,k} f_i^k$$

Example For the Heat Equation $u_t = u_{xx}$, using a forward difference in time and a centered difference in space, we get

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

Pseudocode :
At $n = 0$, $U_j^0 = u^0(x_j, 0)$
for $n = 1 : N$
  $U_0^n = 0, U_J^n = 0$
  for $j = 1 : (J-1)$
    $U_j^{n+1} = U_j^n + \mu(U_{j+1}^n - 2U_j^n + U_{j-1}^n)$
  end
end

The **stability** of the problem depends on μ.

5.4 Truncation Error

Example For our model problem (the Heat Equation), the trucation error is

$$T(x,t) \coloneqq \frac{D_{+t}u(x,t)}{\Delta t} - \frac{D_x^2 u(x,t)}{(\Delta x)^2}$$

So we see that since $u_t - u_{xx} = 0$,

$$T(x,t) = (u_t - u_{xx}) + \left(\frac{1}{2}u_{tt}\Delta t - \frac{1}{12}u_{xxxx}(\Delta x)^2\right) + ... = \left(\frac{1}{2}u_{tt}\Delta t - \frac{1}{12}u_{xxxx}(\Delta x)^2\right) + ...$$

If we truncate this Infinite Taylor series using $\eta \in (t, t + t\Delta t)$ and $\xi \in (x - \Delta x, x + \Delta x)$ and assume the boundry and initial data are consistent at the corners and are both sufficientl smooth, we can then estimate $|u_{tt}(x, \eta)| \le M_{tt}$ and $|u_{xxxx}(\xi, n)| \le M_{xxxx}$, so it follows that

$$|T(x,t)| \le \frac{1}{2}\Delta t \left(M_{tt} - \frac{1}{6\mu}M_{xxxx}\right)$$

We can assume these bounds will hold uniformly over the domain. We see that

$$|T(x,t)| \to 0 \text{ as } \Delta t, \Delta x \to 0 \; \forall \; (x,t) \in \Omega$$

and this result is independent of any relaton between the two mesh sizes. Thus this scheme is **unconditionally consistent** with the differential equation.
Since $|T(x,t)|$ will behave asymptotically like $O(\Delta t)$ as $\Delta t \to 0$, this scheme is said to have **first order accuracy**
Since $u_t = u_{xx}$, $u_{tt} = u_{xxxx}$ and so for $\mu = \frac{1}{6}$,

$$T(x,t) = \frac{1}{2}\Delta t \left(u_{tt} - \frac{1}{6\mu}u_{xxxx}\right) + O\left((\Delta t)^2\right) = O\left((\Delta t)^2\right)$$

and so the scheme is **second order accurate** for $\mu = \frac{1}{6}$
We can define notation: $T_j^n = T(x_j, t_n)$

5.5 Consistency

Does the difference scheme approximate the PDE as $\Delta x, \Delta t \to 0$?
A scheme is consistent if

$$T(x,t) \to 0 \text{ as } \Delta x, \Delta t \to 0$$

- A scheme is **unconditionally consistent** if the scheme is consistent for any relationship between Δx and Δt
- A scheme is **conditionally consistent** if the scheme is consistent only for certain relationships between Δx and Δt

5.6 Accuracy

Take μ finite, so $(\Delta t)^{\frac{a}{b}} = \mu(\Delta x)^{\frac{c}{d}}$, so $(\Delta t)^{\alpha} = (\Delta t)^{ad} = \mu^{bd}(\Delta x)^{cb}$ and we have

$$|T(x,t)| = O(\Delta t^{\alpha})$$

- If $\alpha = 1$, the scheme is **first order accurate**.
- If $\alpha = 2$, the scheme is **second order accurate**.
- etc...
- The scheme is α**-order accurate**.

5.7 Convergence

A scheme is **convergent** if as $\Delta t, \Delta x \to 0$ for any fixed point (x^*, t^*) in the domain,

$$x_j \to x^*, t_n \to t^* \implies U_j^n \to u(x^*, t^*)$$

It suffices to show this for mesh points for sufficiently refined meshes, as convergence at all other points will follow from continuity of $u(x,t)$. We suppose that we can find a bound for the error $\bar{T}$:

$$\left|T_j^n\right| \leq \bar{T} < \infty$$

We denote the **error**

$$e_j^n := U_j^n - u(x_j, t_n)$$

Taking the difference between the scheme and $u(x_j, t^{n+1})$ in terms the truncation error and the exact solution at previous time steps yields the error at e_j^{n+1}. If the RHS of our difference scheme is represented by D, then

$$e_j^{n+1} = DU_j^n - (Du(x_j, t_n) + T(x_j, t_n)\Delta t) = De_j^n - T_j^n \Delta t$$

Choose μ such that the coefficients of the RHS are positive so that you may estimate $E^n := \max\left\{\left|e_j^n\right|, j = 0, ..., J\right\}$ and so

$$E^{n+1} \leq E^n + \bar{T}\Delta t \text{ s.t. } E^0 = 0$$

and thus

$$E^n \leq n\bar{T}\Delta t \qquad n = 0, 1, 2, ...$$

and considering the domain

$$E^n \leq \bar{T} t_F \qquad n = 0, 1, 2, ..., N$$

and since $\bar{T} \to 0$ as $\Delta t, dx \to 0$, $E^n \to 0$

Example If we replace U_j^n with $u(x_j, t_n)$ in the definition of T_j^n we obtain

$$e_j^{n+1} = e_j^n + \mu D_x^2 e_j^n - T_j^n \Delta t$$

which is

$$e_j^{n+1} = (1-2\mu)e_j^n + \mu e_{j+1}^n + \mu e_{j-1}^n - T_j^n \Delta t$$

For $\mu \leq \frac{1}{2}$, define $E^n \coloneqq \max\left\{\left|e_j^n\right|, j = 0, ..., J\right\}$ and so

$$\left|e_j^{n+1}\right| \leq E^n + \bar{T}\Delta t \implies E^{n+1} \leq E^n + \bar{T}\Delta t$$

Since $E^0 = 0$ (the initial values are exact), we have

$$E^n \leq n\bar{T}\Delta t = \frac{1}{2}\Delta t\left(M_{tt} - \frac{1}{6\mu}M_{xxxx}\right)t_F \to 0 \text{ as } t \to 0$$

5.7.1 Refinement Path

A **refinement path** is a sequence of pairs of mesh sizes each which tends to zero

$$\text{refinement path} \coloneqq \{((\Delta x)_i, (\Delta t)_i), i = 0, 1, 2, ...; (\Delta x)_i, (\Delta t)_i \to 0\}$$

We can specify particular paths by requiring certain relationships between the mesh sizes.

Examples $(\Delta t)_i \sim (\Delta x)_i$ or $(\Delta t)_i \sim (\Delta x)_i^2$

Theorem For the heat equation, $\mu_i = \frac{(\Delta t)_i}{(\Delta x)_i^2}$ and if $\mu_i \leq \frac{1}{2}\ \forall\ i$ and if for all sufficiently large values of i and the positive numbers n_i, j_i are such that

$$n_i(\Delta t)_i \to t > 0, j_i(\Delta x)_i \to x \in [0,1]$$

and if $|u_{xxxx}| \leq M_{xxxx}$ uniformly on Ω, then the approximations $U_{j_i}^{n_i}$ generated by the explicit scheme for $i = 0, 1, ...$ converge to the solution $u(x,t)$ of the differential equation uniformly in the region.
This means that arbitrarily good accuracy can be attained by use of a sufficiently fine mesh.

5.8 Error: Fourier Analysis

Let

$$U_j^n = (\lambda)^n e^{ik(j\Delta x)}$$

where $\lambda(k)$ is known as the **amplification factor** of the **Fourier Node** U_j^n. Place this into the difference equation of your scheme and solve for λ. We then have another numerical approximation

$$U_j^n = \sum_{-\infty}^{\infty} A_m e^{-im\pi(j\Delta x)}\left(\lambda(k)\right)^n$$

which can be compared to the Fourier expansion approximating the exact solution.

Example For $U_j^n = U_j^n + \mu(U_{j+1}^n - 2U_j^n + U_{j-1}^n)$, we see

$$\lambda(k) = 1 + \mu(e^{ik\Delta x} - 2 + e^{-ik\Delta x}) = 1 - 4\mu\sin^2\left(\frac{1}{2}k\Delta x\right)$$

So now

$$e^{-k^2\Delta t} - \lambda(k) = \left(1 - k^2\Delta t + \frac{1}{2}k^4\Delta t(\Delta t)^2 - ...\right) - \left(1 - k^2\Delta t + \frac{1}{12}k^4\Delta t(\Delta x)^2 - ...\right) = \left(\frac{(\Delta t)^2}{2} - \frac{\Delta t(\Delta x)^2}{12}\right)k^4 - ...$$

Thus we have first order accuracy in general but second order accuracy if $(\Delta x)^2 = 6\Delta t$.

A scheme is **stable** if there exists a constant K such that

$$|(\lambda)|^n \leq K, \qquad n\Delta t \leq t_F, \ \forall \ k$$

That is, if the difference in the solutions of the DE and the numerical DE is bounded uniformly in the domain for any amount of time less than t_F. Thus

$$|\lambda(k)| \leq 1 + K'\Delta t$$

This is necessary and sufficient.

5.10 Implicit Scheme

If the scheme cannot be written in a form that has U_j^{n+1} explicitly computed given values U_j^n, $j = 0, 1, ..., J$, it is implicit. Implicit schemes involve more work but often have higher accuracy and/or stability, and thus much larger time steps allow us to reach the solution much more quickly.

Example

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

which can be written as

$$\Delta_{-t} U_j^{n+1} = \mu \delta_x^2 U_j^{n+1} \qquad \mu = \frac{\Delta t}{(\Delta x)^2}$$

This involves solving a system of linear equations. However, Fourier analysis for the stability shows

$$\lambda = \frac{1}{1 + 4\sin^2\left(\frac{1}{2}k\Delta t\right)}$$

Since $\lambda < 1$ for any positive μ, this scheme is **unconditionally stable**

5.11 Other Conditions

If an equation obeys extra conditions such as a Maximum Principle, uniqueness condition, or a physical constraint, the numerical scheme must also obey such conditions else it may not converge.

6 Methods

6.1 Weighted Average θ method

Given two schemes, you can weight one θ and the other with $(1-\theta)$ and add them together. Then stability, covergence, and accuracy may depend on θ, and it can be chosen to

Example For the explicit and implicit first order accurate schemes for the heat equation are averaged, we have

$$U_j^{n+1} - U^n = \mu\left(\theta\delta_x^2 U_j^{n+1} + (1-\theta)\delta_x^2 U_j^n\right)$$

$\theta = 0$ yields the explicit scheme and $\theta = 1$ yields the implicit scheme.

7 General Boundary Conditions

Boundary conditions like

$$u_x = \alpha(t)u + g(t) \qquad x = 0$$

Can be handled like

$$\frac{U_1^n - U^n)0}{\Delta x} = \alpha^n U_0^n + g^n \implies U_0^n = \beta^n U_1^n - \beta^n g^n \Delta t \qquad \beta^n = \frac{1}{1 + \alpha^n \Delta x}$$

Dirichlet conditions are trivial

$$u(0,t) = 0 \implies U_0^n = 0$$

$$D_x^2 y_i = D_+ D_- y_i = \frac{1}{h^2}(y_{i+1} - 2y_i + y_{i-1})$$

$$y_{i\pm1} = y_i \pm h y_i' + \frac{1}{2}h^2 y_i'' \pm \frac{1}{6}h^3 y_i''' + \frac{1}{24}h^4 y_i''''...$$

$$D_x^2 y_i = \frac{1}{h^2}\left(y_i + hy_i' + h^2 y'' + \frac{1}{6}h^3 y_i''' + \frac{1}{24}h^4 y_i'''' - 2y_i + y_i - hy_i' + h^2 y'' - \frac{1}{6}h^3 y_i''' + \frac{1}{24}h^4 y_i'''' + ...\right)$$

which simplifies to

$$D_x^2 y_i = y_i'' + O(h^2)$$

Let $u_i \approx y_i = y(x_i)$

$$u_{xx} + q(x)u = f(x)$$

with $u(0) = \alpha$ and $u(1) = \beta$ becomes

$$\frac{-1}{h^2}\left(u_{i+1} - 2u_i + u_{i-1}\right) + q_i u_i = f_i$$

or

$$\begin{cases} -u_2 + (2 + h^2 q_1)u_1 = h^2 f_1 + \alpha & i = 1 \\ -u_{i+1} + (2 + h^2 q_i)u_i - u_{i-1} = h^2 f_i & 2 \leq i \leq n \\ (2 + h^2 q_n)u_n - u_{n-1} = h^2 f_n + \beta & i = n \end{cases}$$

where $u_0 = \alpha$ and $u_{n+1} = \beta$ we must solve

$$A_n \vec{\mathbf{u}}_n = \begin{pmatrix} 2 + h^2 q_1 & -1 & & \\ -1 & 2 + h^2 q_2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 + h^2 q_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} h^2 f_1 + \alpha \\ h^2 f_2 \\ \vdots \\ h^2 f_n + \beta \end{pmatrix} = \vec{\mathbf{f}}_n$$

Note: A_n is tridiagonal and symmetric. We can solve this by using $A_n = LU$

$$L\vec{\mathbf{v}}_n = \vec{\mathbf{f}}_n \qquad U\vec{\mathbf{u}}_n = \vec{\mathbf{v}}_n$$

8 New Notes

9 Finite Difference Coefficients

Pierson Guthrey

9.1 Centered Differences

For the difference scheme for the α derivative of f,

$$f^{(\alpha)}(x_i) \approx D_h^\alpha f = \frac{1}{d}\frac{a_{i-4}f(x_{i-4}) + ... + a_i f(x_i) + ... + a_{i+4}f(x_{i+4})}{h^\alpha}$$

where d is the denominator to make the coefficients a_i integers. If we want the scheme to have accuracy β,

$$f^{(\alpha)}(x_i) = D_h^\alpha f + Oh^\beta$$

Then the difference coefficients are given by

α	β	d	a_{i-4}	a_{i-3}	a_{i-2}	a_{i-1}	a_i	a_{i+1}	a_{i+2}	a_{i+3}	a_{i+4}
1	2	2				-1	0	1			
	4	12			1	-8	0	8	-1		
	6	60		-1	9	-45	0	45	-9	1	
	8	840	3	-32	168	-672	0	672	-168	32	-3
2	2	1				1	-2	1			
	4	12			-1	16	-30	16	-1		
	6	180		2	-27	270	-490	270	-27	2	
	8	5040	-9	128	-1008	8064	-14350	8064	-1008	128	-9
3	2	2			-1	2	0	-2	1		
	4	8		1	-8	13	0	-13	8	-1	
	6	240	-7	72	-338	488	0	-488	338	-72	7
4	2	1			1	-4	6	-4	1		
	4	6		-1	12	-39	56	-39	12	-1	
	6	240	7	-96	676	-1952	2730	-1952	676	-96	7

9.2 Forward/Backwards Differences

For the difference scheme for the α derivative of f,

$$f^{(\alpha)}(x_i) \approx D_\pm^\alpha f = \frac{1}{d}\frac{a_i f(x_i) + ... + a_{i\pm 4}f(x_{i\pm 4}) + ... + a_{i\pm 8}f(x_{i\pm 8})}{h^\alpha}$$

where d is the denominator to make the coefficients a_i integers. If we want the scheme to have accuracy β,

$$f^{(\alpha)}(x_i) = D_\pm^\alpha f + Oh^\beta$$

Then the difference coefficients are given by

α	β	d	a_i	a_{i+1}	a_{i+2}	a_{i+3}	a_{i+4}	a_{i+5}	a_{i+6}	a_{i+7}	a_{i+8}
1	1	1	∓ 1	± 1							
	2	2	∓ 3	± 4	∓ 1						
	3	6	∓ 11	± 18	∓ 9	± 2					
	4	12	∓ 25	± 48	∓ 36	± 16	∓ 3				
	5	60	∓ 137	± 300	∓ 300	± 200	∓ 75	± 12			
	6	60	∓ 147	± 360	∓ 450	± 400	∓ 225	± 72	∓ 10		
2	1	1	1	-2	1						
	2	1	2	-5	4	-1					
	3	12	35	-104	114	-56	11				
	4	12	45	-154	214	-156	61	-10			
	5	180	812	-3132	5265	-5080	2970	-972	137		
	6	180	938	-4014	7911	-9490	7389	-3616	1019	-126	
3	1	1	∓ 1	± 3	∓ 3	1					
	2	2	∓ 5	± 18	∓ 24	± 14	∓ 3				
	3	4	∓ 17	± 71	∓ 118	± 98	∓ 41	± 7			
	4	8	∓ 49	± 232	∓ 461	± 496	∓ 307	± 104	∓ 15		
	5	120	∓ 967	± 5104	∓ 11787	± 15560	∓ 12725	± 6432	∓ 1849	± 232	
	6	240	∓ 2403	± 13960	∓ 36706	± 57384	∓ 58280	± 39128	∓ 16830	± 4216	∓ 469
4	1	1	1	-4	6	-4	1				
	2	1	3	-14	26	-24	11	-2			
	3	6	35	-186	411	-484	321	-114	17		
	4	6	56	-333	852	-1219	1056	-555	164	-21	
	5	240	3207	-21056	61156	-102912	109930	-76352	33636	-8576	967

New Notes

10 Two Dimensional Problems

Given a problem

$$u_t = b(u_{xx} + u_{yy} \qquad \Omega = [0, X] \times [0, Y]$$

With initial conditions on Ω for $t = 0$ and boundary conditions on $\partial\Omega$

So for a uniform grid mesh

$$U_{i,j}^N \approx u(x_i, y_j, t_n) = u(i\Delta x, j\Delta y, n\Delta t), i = 0, 1, ...I, j = 0, 1, ...J, n = 0, 1, ...N$$

Explicit Scheme

$$\frac{U_{i,j}^{n+1} - U_{i,j}^n}{\Delta t} = b\left(\frac{\delta_x^2 U_{i,j}^n}{\Delta x^2} + \frac{\delta_y^2 U_{i,j}^n}{\Delta y^2}\right)$$

Consistency

$$T_{ij}^n = \left(\frac{1}{2}\Delta t u_{tt} - \frac{1}{12} b\left(\Delta x^2 u_{xxxx} + \Delta y^2 u_{yyyy}\right)\right)_{ij}^n + ...$$

$$T_{ij}^n \approx O(\Delta t + \Delta x^2 + \Delta y^2)$$

Stability

$$U_{ij}^n \approx (\lambda)^n e^{i(k_x i\Delta x + k_y i\Delta y)}$$

????

Convergence The error estimate

$$e_{ij}^n = U_{ij}^n - u_{ij}^n$$

$$U_{ij}^n = U_{ij}^n + b\left(\mu_x \delta_x^2 U_{ij}^n + \mu_y \delta_y^2 U_{ij}^n\right)$$

$$e_y^{n+1} = e_y^n + b\left(\mu_x \delta_x^2 e_{ij}^n + \mu_y \delta_y^2 e_{ij}^n\right) - \Delta t T_{ij}^n$$

$$\mu_x = b\frac{\Delta t}{\Delta x^2}, \mu_y = b\frac{\Delta t}{\Delta y^2}$$

10.0.1 θ Scheme

Accuracy is $O(\Delta t + \Delta x^2 + \Delta y^2)$, but if $\theta = \frac{1}{2}$ we have the Crank-Nicholson Scheme with accuracy $O(\Delta t^2 + \Delta x^2 + \Delta y^2)$.

This requires to solve a system

$$A\vec{\mathbf{U}}^{n+1} = \vec{\mathbf{b}}^n$$

Where $U^n ij$ is reshaped into a vector.

10.1 Alternating Direction Implicit (ADI) Methods

We see to modify the 2D problem so that we solve several 1D problems. We approximate the 2D Crank-Nicholson scheme

$$\left(1-\frac{1}{2}\mu_x\delta_x^2\right)\left(1-\frac{1}{2}\mu_x\delta_y^2\right)U_{ij}^{n+1}=\left(1-\frac{1}{2}\mu_x\delta_x^2\right)\left(1-\frac{1}{2}\mu_x\delta_y^2\right)U_{ij}^{n} \tag{1}$$

Note that

$$\left(1-\frac{1}{2}\mu_x\delta_x^2\right)\left(1-\frac{1}{2}\mu_x\delta_y^2\right)=1-\frac{1}{2}\mu_x\delta_x^2-\frac{1}{2}\mu_y\delta_y^2+\frac{1}{4}\mu_x\mu_y\delta_x^2\delta_y^2\approx O(\Delta t T_{ij}^{n+\frac{1}{2}}$$

We solve for the intermediate solution

$\left(1-\frac{1}{2}\mu_x\delta_x^2\right)U_{ij}^{n+\frac{1}{2}}=\left(1+\frac{1}{2}\mu_y\delta_y^2\right)U_{ij}^{n} using a system of equations. Then for the solution at the next step we must solve another s$

Stability is based on Fourier Analysis on equation (1) and shows that this scheme is unconditionally stable.

Maximum principle on (**??**) yields that we require $\mu_x \leq 1$. The same analysis on (**??**) yields that we require $\mu_y \leq 1$. Thus we require $\max\{\mu_x,\mu_y\}\leq 1$.

Consistency

$$T_{ij}^{n+\frac{1}{2}}=\left(\frac{1}{24}\Delta t^2 u_{ttt}-\frac{1}{12}u_{xxxx}-\frac{1}{12}\Delta y^2 u_{yyyy}-\frac{1}{8}\Delta t^2 u_{xxtt}-\frac{1}{8}\Delta t^2 u_{yytt}+\frac{1}{4}\Delta t^2 u_{xxyyt}\right)_{ij}^{n+\frac{1}{2}}+...\approx O(\Delta t^2+\Delta x^2+\Delta y^2)$$

10.2 Locally One Dimensional (LOD) Scheme

We can expand this to 3D

$$u_t=b(u_{xx}+u_{yy}+u_{zz})$$

$$\begin{cases}\left(1-\frac{1}{2}\mu_x\delta_x^2\right)U_{ij}^{n+*}=\left(1+\frac{1}{2}\mu_x\delta_x^2\right)U_{ij}^{n}\\ \left(1-\frac{1}{2}\mu_y\delta_y^2\right)U_{ij}^{n+**}=\left(1+\frac{1}{2}\mu_y\delta_y^2\right)U_{ij}^{n+*}\\ \left(1-\frac{1}{2}\mu_z\delta_z^2\right)U_{ij}^{n+1}=\left(1+\frac{1}{2}\mu_z\delta_z^2\right)U_{ij}^{n+**}\end{cases}$$

11 First Order Problems

$$F(Du,u,x)=0$$

In general there is no classical solution globally. Weak solutions may exist.

11.1 Method of Characteristics

$$z(s)u(x(s)),\vec{\mathbf{p}}(s)=Du(x(s))$$

$$\begin{cases}\dot{x}(s)=D_{\vec{\mathbf{p}}}F(\vec{\mathbf{p}}(s),z(s),x(s))\\ \dot{z}(s)=D_{\vec{\mathbf{p}}}F(\vec{\mathbf{p}}(s),z(s),x(s))\cdot\vec{\mathbf{p}}(s)\\ \dot{\vec{\mathbf{p}}}(s)=-D_xF-D_zF\cdot\vec{\mathbf{p}}(s)\end{cases}$$

Model problem: Transport equation/advection equaton

$$u_t+a(x,t)u_x \qquad u(x,t=0)=u^0(x)$$

If a constant:

Upwind Scheme

$$\text{If } (a)=\pm 1,\ \frac{U_j^{n+1}-U_j^n}{\Delta t}+a\frac{\pm U_j^n\mp U_{j\mp 1}^n}{\Delta x}=0$$

11.2 Schemes for Parabolic Equations

Courant Friedrich Lowry (CFL) Condition for convergence

$$|v| \leq 1 \qquad v = \frac{a\Delta t}{\Delta x} \text{ (CFL number)}$$

- The CFL is necessary but not sufficient for convergence.
- This ensures the domain of dependence of the scheme is a subset of the domain of dependence of the equation.

Characteristic Ray Tracing Method (Semi-Lagrangian Method)

11.2.1 Euler Schemes

Upwind Differencing Interpolate $U(x^*, t_n)$ with $\left\{U_j^n\right\}_{j=0}^J$

Forward Time - Backward Difference Scheme

....?

Higher Dimensions

$$u_t + au_x + bu_y = 0, \qquad a, b > 0$$

Let $v_x = a\frac{\Delta t}{\Delta x}$, $v_y = b\frac{\Delta t}{\Delta y}$

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

- CFL conditions: $|\nu_x| \leq 1$, $|\nu_y| \leq 1$
- We find that we require $\nu_x + \nu_y \leq 1$

Backward Time - Forward Difference Scheme

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

Since the computational domain of dependence is a rectangle, CFL will be satisfied.

- Unconditionally stable

Forward Time - Central Difference Scheme

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

- Unconditionally unstable

Backward Time - Central Difference Scheme

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

- Unconditionally stable
- Accuracy is $O((\Delta x)^2)$

$$\left(-\frac{\nu^2}{2}-\frac{\nu}{2}\right)U_{j-1}^{n+1}+\left(1+\nu^2\right)U_j^{n+1}+\left(-\frac{\nu^2}{2}-\frac{\nu}{2}\right)U_{j+1}^{n+1}=U_j^n$$

- Unconditionally Stable
- Accuracy $O((\Delta x)^2+(\Delta t)^2)$

Crank-Nicolson

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

- Unconditionally Stable with $|\lambda|=1$, so may become unstable due to roundoff error
- Accuracy $O((\Delta x)^2+(\Delta t)^2)$

Lax Friedrichs Higher Dimensions

$$U_{i,j}^{n+1}=\frac{1}{4}\left(U_{i+1,j}^n+U_{i-1,j}^n+U_{i,j-1}^n+U_{i,j+1}^n\right)-\frac{1}{2}\nu_x\left(U_{i+1,j}^n-U_{i-1,j}^n\right)-\frac{1}{2}\nu_y\left(U_{i,j+1}^n-U_{i,j-1}^n\right)$$

- $\lambda=\frac{1}{2}(\cos(\xi)+\cos(\eta))-i(\nu_x\sin(\xi)+\nu_y\sin(\eta))$
- $\nu_x^2+\nu_y^2\leq 1$

Euler Scheme

$$U_{ij}^{n+\frac{1}{2}}=U^{ij-\nu_x\Delta x_0 U_{ij}^n+\frac{1}{2}\nu_x^2\delta_x^2U_{ij}^n}$$

$$U_{ij}^{n+1}=U_{ij}^{n+\frac{1}{2}}-\nu_y\Delta y_0U_{ij}^{n+\frac{1}{2}}+\frac{1}{2}\nu_y^2\delta_y^2U_{ij}^{n+1}$$

- Accuracy $O((\Delta t)^2+(\Delta x)^2+(\Delta y)^2)$
- $\max\{|\nu_x|,|\nu_y|\}\leq 1$

Leap Frog

Beam Wamming?

12 2.20.2014

$$u_t+au_x+bu_y=0$$

Method of Characteristics tells us

$$u(x,y,t)=u^0(x-at,y-bt),$$

Forward Time Upwind Scheme

$$U_{i,j}^{n+1}=U^{i,j-\frac{1}{2}\nu_x\left(U_{i,j}^n-U_{i-1,j}^n\right)-\frac{1}{2}\nu_y\left(U_{i,j}^n-U_{i,j-1}^n\right)}$$

- CFL condition: $\max\{|\nu_X|,|\nu_y|\}\leq 1$
- Stability (Fourier analysis) $|\nu_x|+|\nu_y|\leq 1$

- Truncation error: $O(\Delta t + \Delta x + \Delta y)$

Law Wendroff Scheme

$$U_{i,j}^{n+\frac{1}{2}} = U^{i,j-\frac{1}{2}\nu\Delta_{x0}U_{i,j}^n + \frac{1}{2}\nu_x^2\delta_x^2 U_{i,j}^n}$$

$$U_{i,j}^{n+1} = U_{i,j}^{n+\frac{1}{2}} - \nu_y \Delta_{y0} U_{i,j}^n + \frac{1}{2}\nu_y^2 \delta_y^2 U_{i,j}^n$$

- CFL condition: $\max\{|\nu_X|, |\nu_y|\} \leq 1$
- Stability (Fourier analysis) $|\nu_x| + |\nu_y| \leq 1$
- Truncation error: $O((\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2)$

13 ADI Schemes

13.1 Locally One Dimensional Scheme

$$(1 + \nu_x \Delta_{x0}) U_{i,j}^{n+\frac{1}{2}} = U_{i,j}^n$$

$$(1 + \nu_y \Delta_{y0}) U_{i,j}^{n+1} = U_{i,j}^{n+\frac{1}{2}}$$

- CFL condition: ?
- Stability: Unconditionally Stable. Fourier Analysis: we find $|\lambda| \leq 1$.
- Truncation error: $O(\Delta t + (\Delta x)^2 + (\Delta y)^2)$

13.2 Crank Nicolson Scheme

$$\frac{U_{i,j}^{n+1} - U_{i,j}^n}{\Delta t} + \frac{1}{2}\nu_x \Delta_{x0}(U_{i,j}^n + U_{i,j}^{n+1}) + \frac{1}{2}\nu_y \Delta_{y0}(U_{i,j}^n + U_{i,j}^{n+1}) = 0$$

- CFL condition: ?
- Stability: Unconditionally Stable. Unproven
- Truncation error: $O((\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2)$

Beam Wamming

$$\left(1 + \frac{1}{2}\nu_x \Delta_{x0}\right) U_{i,j}^* = \left(1 - \frac{1}{2}\nu_x \Delta_{x0}\right)\left(1 - \frac{1}{2}\nu_y \Delta_{y0}\right) U_{i,j}^n$$

$$\left(1 + \frac{1}{2}\nu_y \Delta_{x0}\right) U_{i,j}^{n+1} = U_{i,j}^*$$

- CFL condition: ?
- Stability: $|\lambda| = 1$
- Truncation error: ?

14 2.25.2014

Pierson Guthrey

Consistency, convergence, stability, and Lax Equivalence Theorem.
Consider the problem in the general form.

$$\begin{cases} \frac{\partial u}{\partial t} = Lu & \Omega \times [0, t_F] \\ g(u) = g_0 & u \in \partial\Omega \\ u(x,0) = u^0(x) & x \in \Omega, t = 0 \end{cases}$$

We assume that Ω is bounded, and L represents a differential operator such that $\frac{\partial u}{\partial t} = Lu$ is **well posed**:

- **Existence of solutions**: A solution exists for all data u^0 for which $\left\|u^0\right\|$ is bounded.
- **Continuous dependence on data**: There exists a constant K such that for any pair of solutions u and v, $\|u - v\| \le K\left\|u^0 - v^0\right\|$ for all $t \le t_F$.

Schemes for solutions

$$B_1\vec{\mathbf{U}}^{n+1} = B_0\vec{\mathbf{U}}^n + \vec{\mathbf{F}}^n$$

Assume B_1 exists. Then a solution to the difference scheme exists:

$$\vec{\mathbf{U}}^{n+1} = B_1^{\left(B_0\vec{\mathbf{U}}^n + \vec{\mathbf{F}}^n\right)}$$

The truncation error is defined by the equation

$$B_1\vec{\mathbf{u}}^{n+1} = B_0\vec{\mathbf{u}}^n + \vec{\mathbf{F}}^n + \vec{\mathbf{T}}^n$$

Thus subtracting the discrete PDE scheme by the continuous PDE scheme we get

$$\vec{\mathbf{U}}^{n+1} - \vec{\mathbf{u}}^{n+1} = B_1^B{}_0\left(\vec{\mathbf{U}}^n - \vec{\mathbf{u}}^n\right) - B_1^{\vec{\mathbf{T}}^n}$$

Using the implied recursive relationship,

$$\vec{\mathbf{U}}^{n+1} - \vec{\mathbf{u}}^{n+1} = B_1^{\vec{\mathbf{T}}^{n-1} + B_1^B{}_0 B_1^{\vec{\mathbf{T}}^{n-2} + \dots + (B_1^B{}_0)^{n-1} B_1^{\vec{\mathbf{T}}^0}}}$$

So if $\left\|(B^{1^{B_0^n)}}\right\| \le K \;\forall\; n\Delta t \le t_F$ and $\left\|B^1\right\| \le K_1\Delta t$, then $\left\|(B_1^B{}_0)^m B_0^{)}\right\| \le K_1 K\Delta t \;\forall\; m \le n$ so $\left\|\vec{\mathbf{U}}^n - \vec{\mathbf{u}}^n\right\| \le K_1 K\Delta t \sum_{m=0}^{n-1}\left\|\vec{\mathbf{T}}^m\right\|$

- **Consistency**: $T_{i,j}^n \to 0$ as $\Delta t, \Delta x, \Delta y, ... \to 0$ for all i, j which implies $B_1\vec{\mathbf{u}}^{n+1} - \left(B_0\vec{\mathbf{u}}^n + \vec{\mathbf{F}}^n\right) \to \frac{\partial u}{\partial t} - Lu$
- **Accuracy**: If p, q are the largest positive numbers for which $T_{i,j}^n \le O\left((\Delta t)^p + h^q\right)$ as $\Delta t \to 0$and $h \to 0$ for sufficiently smooth u, where $h = \max \Delta x, \Delta y, ...$, the scheme is said to have **order of accuracy** p in Δt and q in h.
- **Stability**: The scheme is said to be **stable** if two solutions $\vec{\mathbf{U}}^n$ and $\vec{\mathbf{V}}^n$ of the scheme which have the same inhomogeneous terms $\vec{\mathbf{F}}^n$ but start form different initial data $\vec{\mathbf{U}}^0$ and $\vec{\mathbf{V}}^0$ satisfy

  $$\left\|\vec{\mathbf{U}}^n - \vec{\mathbf{V}}^n\right\| \le K\left\|\vec{\mathbf{U}}^0 - \vec{\mathbf{V}}^0\right\| \qquad \forall\; n\Delta t \le t_F$$

  for some constant K independent of the initial data and mesh sizes. Equivalently,

  $$\left\|\left(B_1^B{}_0\right)^n\right\| \le K \qquad \forall\; n\Delta t \le t_F$$

- If the PDE has constant coefficients, Fourier Analysis can be used
- A maximum principle is sometimes necessary for Parabolic

- **Convergence**: The scheme provides **convergent approximations** to the problem if $\left\|\vec{\mathbf{U}}^n - \vec{\mathbf{u}}^n\right\| \to 0$ as $\Delta t, h \to 0$, $n\Delta t \to t \in [0, t_F]$ for every u^0 for which the problem is well posed.

Lax Equivalence Theorem For a consistent difference approximation to a well posed linear evolutionary problem which is uniformly stable in the sense that $\|B\| \leq K\Delta t$ for some constant K, the stability of the scheme is necessary and sufficient for convergence.

14.1 Dissipation and Dispersion

The **Dissipation** of solutions of PDEs is when the Fourier modes do not grow with time and at least one mode decays. The PDE is **Non-dissipative** if the Fourier modes neither decay nor grow.
The **Dispersion** of of solutions of PDEs is when the Fourier modes of differing wave lengthd (or wave numbers) propagate at different speeds.

Von Neumann Condition A necessary condition for stability is that the exists a constant K such that

$$\left|\lambda(\vec{\mathbf{k}})\right| \leq 1 + K\Delta t, \ \forall\, \vec{\mathbf{K}}, n\Delta t \leq t_F$$

or as $\Delta t \to 0$ and $h \to 0$

$$\left|\lambda(\vec{\mathbf{k}})\right|^n \leq K$$

or

$$\left|\lambda(\vec{\mathbf{k}})\right|^n \leq (1 + K\Delta t)^n \approx (1 + K\Delta t n) + O((\Delta t)^2)$$