2. The basic equations and some properties of partial differential eqn.

We will mostly concentrate on solving the two-dimensional incompressible flow problem in rectangular coordinates.

2.1. The basic equations

The fundamental equations for 2-dimensional incompressible flow are the Navier-Stokes equations and the continuity equation. In the absence of rotation, they are

$$(1) \quad \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \nu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)$$

$$(2) \quad \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -\frac{1}{\rho}\frac{\partial p}{\partial y} + \nu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right)$$

$$(3) \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$

u, v velocities   viscosity ν

p pressure

ρ density

We can obtain numerical solutions for this set of equations, but for simplicity, as a first step, we will use the vorticity - streamfunction approach.

If we define the vertical component of the vorticity as $\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$, by cross-differentiating (1) and (2), we obtain the vorticity equation

$$(4) \quad \frac{\partial \zeta}{\partial t} + u\frac{\partial \zeta}{\partial x} + v\frac{\partial \zeta}{\partial y} = \nu\left(\frac{\partial^2 \zeta}{\partial x^2} + \frac{\partial^2 \zeta}{\partial y^2}\right)$$

or

$$(5) \quad \boxed{\frac{\partial \zeta}{\partial t} + \vec{v}\cdot(\nabla \zeta) = \nu \nabla^2 \zeta = \frac{D\zeta}{Dt}}$$

The vorticity equation then consists of.

- an unsteady term $\frac{\partial \xi}{\partial t}$
- an advective term $\vec{v}.(\nabla \xi)$
- viscous term $\nu \nabla^2 \xi$

- Since $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$, we can define a streamfunction ψ such that $\frac{\partial \psi}{\partial x} = v$ and $\frac{\partial \psi}{\partial y} = -u$
The vorticity can then be expressed as

(6) $$\boxed{\nabla^2 \psi = \xi}$$ Poisson equation

- The vorticity equation is classified as <u>parabolic</u> which means that it is an initial value problem, wherein the solution is stepped out of some initial condition. On the other hand, the streamfunction equation (6) is <u>elliptic</u> or boundary - value problem which is usually solved by iterative methods

The vorticity equation can also be rewritten is what is called "conservative" form. By using the continuity equation $\nabla . \vec{v} = 0$, equation (5) can be rewritten as

(7) $$\frac{\partial \xi}{\partial t} + \underbrace{\nabla . (\vec{v} \xi)}_{\vec{v}.\nabla\xi + \xi\underbrace{(\nabla . \vec{v})}_{=0}} = \nu \nabla^2 \xi$$

the advantage of such a formulation will be discussed later.

Let's know perform a dimensional analysis of the vorticity equation which will then give us an idea of each terms importance

$(u, v) \to U$ $\quad (x, y) \to L$

$\zeta \to U/L$ $\quad t \to L/U$ advective time scale

then (7) can be rewritten as

(8) $$\frac{\partial \zeta'}{\partial t'} = - \nabla . (\vec{v}'\zeta') + \frac{1}{Re} \nabla^2 \zeta$$

with $Re = UL/\nu$, Reynolds number.

High Reynolds number, $Re \gg 1$ $\Rightarrow$ the advective term is dominant and L/U is the time which effectively characterise the flow. But for low Reynolds number $Re \ll 1$, a characteristic time depend on the diffusion is better

$$t \to \nu/L^2$$

which gives

(9) $$\frac{\partial \zeta'}{\partial t'} = - Re\, \nabla . (\vec{v}'\zeta') + \nabla^2 \zeta$$

As $Re \to 0$, the advective term drop out. The use of the appropriate time constant will minimize round off errors which is of importance.

We still have a complex set of equation and a lot can be learned from one-dimensional equation.

The One dimensional advection - diffusion equation is

(10) $$\frac{\partial \zeta}{\partial t} + \frac{\partial (u \zeta)}{\partial x} = \alpha \frac{\partial \zeta^2}{\partial x^2}$$

ζ is here the vorticity, but can also be any other advected or diffused flow property. u is in general a constant

Another transport equation is simply

(11) $$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \alpha \frac{\partial^2 u}{\partial x^2}$$ Burgers equation

(4) Insert

Basics of PDEs

Partial differential are used to model a wide variety of physical phenomena. A number of properties can be used to distinguish the different type of differential equations.

Example: $a u_{xx} + b u_{xy} + c u_{yy} + d u_x + e u_y$

$+ f = 0$

Order: The order of a PDE is the order of the highest occuring derivative. The order of the above exaple is 2. It is second order in x and y. Most equations derived from physical principles are usually 1st order in time and first or second order in space.

Linear: The PDE is linear if none of it's coefficient depend on the unknown function, i.e. $a, b, c, \ldots$ independent of u in the above exaple. Linear combinations of linear PDEs form another PDE.
$w = \alpha u + \beta v$ is a solution of a PDE where both u and v are solutions.
The Laplace equation $u_{xx} + u_{yy} = 0$ is linear.
The Burger equation $u_t + u u_x = 0$ is non linear.

The majority of numerical analysis results are valid for linear equations. Little is known or generalizable to non linear equations.

with the equivalent conservation form

(12) $$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) = \alpha \frac{\partial^2 u}{\partial x^2}$$

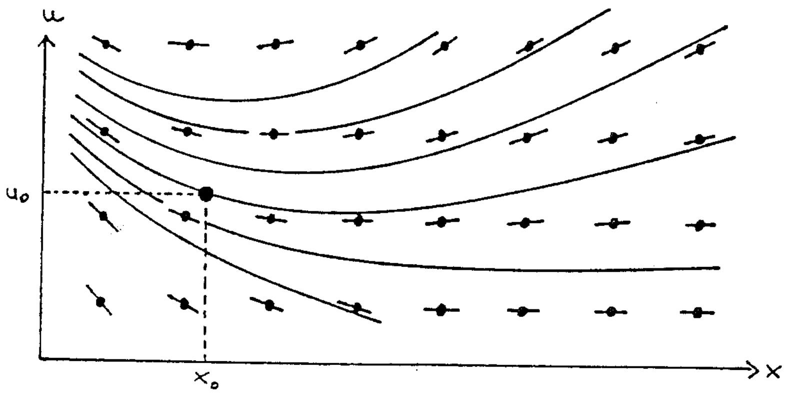
2.2 Some properties of partial differential equations

(*) Insert here

a) Ordinary differential equations of first order.

(13) General form : $F(u', u, x) = 0$ with $u(x)$

The problem of finding the solution $u(x)$ from (13) is equivalent to draw a stream line in the (x, u) plane through a field of velocity vectors whose directions u' are given by (13) $u, x \Rightarrow u'$



Each curve shown is obviously a solution of the differential equation.

Is one solution unique? Not necessarily

Example: $u = x u' + (u')^2$

$u = cx + c^2$ is a solution, but $u = -x^2/4$ is also a solution, but which cannot be obtained from a choice in the constant c.

We then need to know when a unique solution does exist.

Theorem

Given the first order differential equation $u' = F(x,u)$

If F satisfies

1. F is real, finite single valued and continuous for all x, u.

2. $\dfrac{\partial F(x,u)}{\partial u}$ is real, finite, single-valued and continuous.

then there is a unique $u = g(x)$ which passes through any given point of R. (<u>True of linear diff. eqn</u>)

Then to make the solution unique, we just have to prescribe a "boundary condition" or specify a point which solution curve is supposed to pass through.

Linear DE:

$$u^{(n)} + b_{n-1}(x)\, u^{(n-1)} + b_{n-2}(x)\, u^{(n-2)} + \dots + b_0(x)\, u = r(x)$$

b) Ordinary DE of 1st order in two independent variables

(14) <u>General Form</u>. $F(u_x, u_y, u, x, y) = 0$

This equation implies that at each point in (x, y, u) space, the partial derivatives u_x, u_y are related, but they are <u>not</u> individually fixed. The orientation of the surface solution is <u>not</u> prescribed as a function of x, y, u. This additional degree of freedom require boundary conditions at more than one point to assure a unique

solution (assuming, of course, a linear PE).
v has to be specified on a curve c.

System :
$$\begin{cases} a u_x + b u_y + c v_x + d v_y = 0 \\ A u_x + B u_y + C v_x + D v_y = 0 \end{cases}$$
The solution is uniquely determined if u and v are specified on a curve c.

References:

- Reiss, Callegari and Ahluwalia:
  Ordinary Differential Equations with applications.
  1976, Holt, Rinehart and Winston Eds

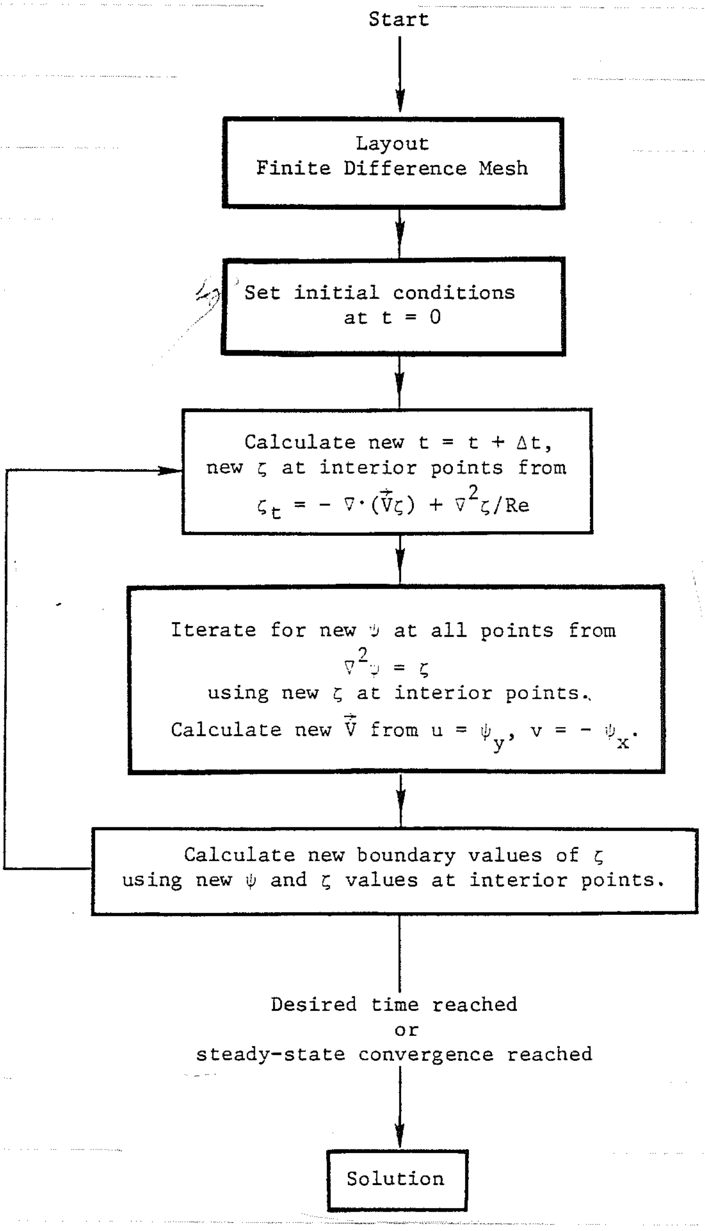
- Spiegel
  Applied Differential Equations.
  1967, Prentice-Hall Eds.

- Duff and Naylor
  Differential Equations of Applied Mathematics
  1966, John Wiley and Sons Eds.

3. Basic Finite Difference Concepts

We concentrate on the finite-difference approach. Other methods will be sketched later. Now, first the framework in which we proceed to solve the equations of Chapter 2.

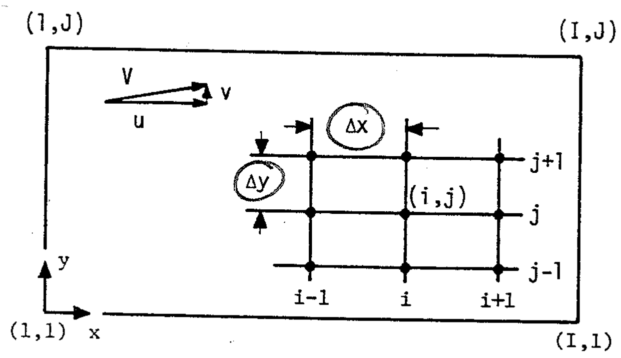
First a set of initial values ψ, ζ, u, v everywhere at time $t=0$. The computational cycle then starts with the use of a finite-difference equation for ζ to approximate $\frac{d\zeta}{dt}$. We then compute ζ at a new time level. Then we solve the Poisson equation for ψ which then gives us u, v. and so on as depicted by this figure


Start
Layout
Finite Difference Mesh
Set initial conditions
at t = 0
Calculate new t = t + Δt,
new ζ at interior points from
$\zeta_t = -\nabla\cdot(\vec{V}\zeta) + \nabla^2\zeta/Re$
Iterate for new ψ at all points from
$\nabla^2\psi = \zeta$
using new ζ at interior points.
Calculate new $\vec{V}$ from $u = \psi_y$, $v = -\psi_x$.
Calculate new boundary values of ζ
using new ψ and ζ values at interior points.
Desired time reached
or
steady-state convergence reached
Solution

3.1 Basic finite-difference forms.

a. Taylor series expansions

- Rectangular mesh



- Taylor series expansion in an interval about $x = a$

$$(1) \quad f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \ldots\ldots + \frac{f^{(n)}(a)(x-a)^n}{n!}$$

- then the uncentered first derivative form of $\frac{\partial f}{\partial x}$ can then be expressed as a function of

$f_{i,j}$, $f_{i+1,j}$, $f_{i-1,j}$

Taylor series expansion $\Longrightarrow$

$$(2) \quad f_{i+1,j} = f_{i,j} + \left.\frac{\partial f}{\partial x}\right|_{i,j} (x_{i+1,j} - x_{i,j}) + \frac{1}{2} \left.\frac{\partial^2 f}{\partial x^2}\right|_{i,j} (x_{i+1,j} - x_{i,j})^2 + \ldots$$

or

$$(3) \quad f_{i+1,j} = f_{i,j} + \left.\frac{\partial f}{\partial x}\right|_{i,j} \Delta x + \frac{1}{2} \left.\frac{\partial^2 f}{\partial x^2}\right|_{i,j} \Delta x^2 + O(\Delta x^3)$$

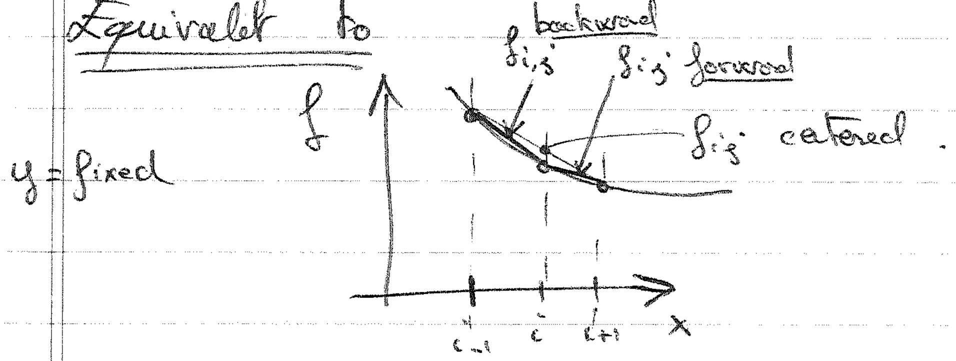
(4) $$\Rightarrow \boxed{\left.\frac{\partial f}{\partial x}\right|_{i,j} = \frac{f_{i+1,j} - f_{i,j}}{\Delta x} + O(\Delta x)}$$

Terms of order Δx
or first-order accuracy.

We can expand backward which then gives

(5) $$\boxed{\left(\frac{\partial f}{\partial x}\right)_{i,j} = \frac{f_{i,j} - f_{i-1,j}}{\Delta x}}$$

Equivalent to



The centered difference approximation $\frac{\partial f}{\partial x}$ is obtained by subtracting the forward and backward expansions:

(6)
$$\begin{cases} f_{i+1,j} = f_{i,j} + \left.\frac{\partial f}{\partial x}\right|_{i,j} \Delta x + \frac{1}{2} \left.\frac{\partial^2 f}{\partial x}\right|_{i,j} \Delta x^2 + \frac{1}{6} \left.\frac{\partial^3 f}{\partial x^3}\right|_{i,j} \Delta x^3 \\ \qquad + \frac{1}{24} \left.\frac{\partial^4 f}{\partial x^4}\right|_{i,j} \Delta x^4 + O(\Delta x^5) \\ f_{i-1,j} = f_{i,j} - \left.\frac{\partial f}{\partial x}\right|_{i,j} \Delta x + \frac{1}{2} \left.\frac{\partial^2 f}{\partial x}\right|_{i,j} \Delta x^2 - \frac{1}{6} \left.\frac{\partial^3 f}{\partial x^3}\right|_{i,j} \Delta x^3 \\ \qquad + \frac{1}{24} \left.\frac{\partial^4 f}{\partial x^4}\right|_{i,j} \Delta x^4 + O(\Delta x^5) \end{cases}$$

$$\Rightarrow \quad f_{i+1,j} - f_{i-1,j} = 2 \left.\frac{\partial f}{\partial x}\right|_{i,j} \Delta x + \frac{1}{3} \left.\frac{\partial^3 f}{\partial x^3}\right|_{i,j} \Delta x^3 + O(\Delta x^5)$$

or $$\left.\frac{\partial f}{\partial x}\right|_{i,j} = \frac{f_{i+1,j} - f_{i-1,j}}{2 \Delta x} - \frac{1}{6} \left.\frac{\partial^3 f}{\partial x^3}\right|_{i,j} \Delta x^2 + O(\Delta x^4)$$

(7) $$\boxed{\left.\frac{\partial f}{\partial x}\right|_{i,j} = \frac{f_{i+1,j} - f_{i-1,j}}{2 \Delta x} + O(\Delta x^2)}$$ Second-order accuracy

Analog expressions can be derived for y and t

(8) $$\left.\frac{\partial f}{\partial y}\right|_{i,j} = \frac{f_{i,j+1} - f_{i,j-1}}{2\Delta y} + O(\Delta y^2)$$

(9) $$\left.\frac{\partial f}{\partial t}\right|_{i,j}^{n} = \frac{f_{i,j}^{n+1} - f_{i,j}^{n-1}}{2\Delta t} + O(\Delta t^2).$$

We can also derive an expression for $\frac{\partial^2 f}{\partial x^2}$

(10) $$\boxed{\left.\frac{\partial^2 f}{\partial x^2}\right|_{i,j} = \frac{f_{i+1,j} + f_{i-1,j} - 2f_{i,j}}{\Delta x^2} + O(\Delta x^2)}$$

second order accurate

b. Polynomial fitting

Another method of obtaining finite-difference expressions is to fit an analytical function with free parameters to mesh-point values and then to analytically differentiate the function. Commonly, polynomials are used.

Parabolic fit: Data at i, $i+1$, $i-1$ for f

For convenience, $x = 0$ is at the location i

$$f(x) = a + bx + cx^2 \qquad \left\{ \begin{array}{l} f_{i-1} = a - b\Delta x + c\Delta x^2 \\ f_i = a \\ f_{i+1} = a + b\Delta x + c\Delta x^2 \end{array} \right.$$

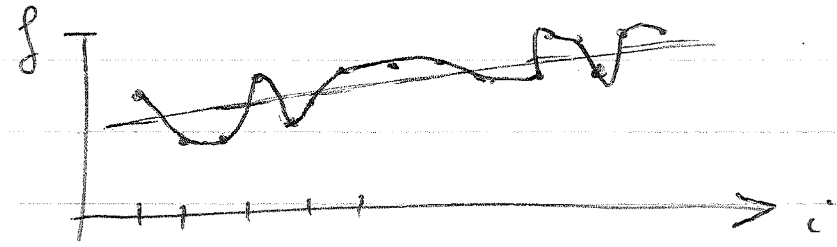
$$\Rightarrow c = \frac{f_{i+1} + f_{i-1} - 2f_i}{2\Delta x^2}$$

$$b = \frac{f_{i+1} - f_{i-1}}{2\Delta x}$$

(11) $$\Longrightarrow \boxed{\left.\frac{\partial f}{\partial x}\right|_i = b \quad \text{and} \quad \left.\frac{\partial^2 f}{\partial x^2}\right| = 2c}$$

which are obviously equivalent to the second order FD obtained in the previous section.

If we just use $y = ax + b$, then we obtained a first-order accuracy (forward and backward of the previous section). Higher polynomials give higher order. Beware of too high



In general, a cubic spline (polynomial) is often used ~~[illegible]~~ since they indicate the presence of an inflexion point:

c) <u>Integral method</u>

In the integral method, we satisfy the governing equation in an integral sense, rather than a differential sense. We write the model equation in conservation form

(12) $$\frac{\partial \zeta}{\partial t} = -\frac{\partial (u\zeta)}{\partial x} + \alpha \frac{\partial^2 \zeta}{\partial x^2}$$

Integration from t to $t+\Delta t$ and $x - \frac{\Delta x}{2}$ to $x + \frac{\Delta x}{2}$

(13) $$\int_{x-\Delta x/2}^{x+\Delta x/2} \left(\int_{t}^{t+\Delta t} \frac{\partial \zeta}{\partial t}\, dt \right) dx = -\int_{t}^{t+\Delta t} \left(\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \frac{\partial (u\zeta)}{\partial x}\, dx + \kappa \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \frac{\partial^2 \zeta}{\partial x^2}\, dx \right) dt$$

$\Longrightarrow$

(14) $$\int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \left(\zeta^{t+\Delta t} - \zeta^{t}\right) dx = -\int_{t}^{t+\Delta t} \left((u\zeta)_{x+\frac{\Delta x}{2}} - (u\zeta)_{x-\frac{\Delta x}{2}} \right) dt + \alpha \int_{t}^{t+\Delta t} \left[\frac{\partial \zeta}{\partial x}\Big|_{x+\Delta x/2} - \frac{\partial \zeta}{\partial x}\Big|_{x-\frac{\Delta x}{2}} \right] dt$$

Theorem: Mean value - theorem

$$\underbrace{\int_{z_1}^{z_1+\Delta z} f(z)\,dz = f(\bar{z})\cdot\Delta z \;\; \cancel{\text{with}} \;\; \underset{\text{such that}}{\bar{z} \in [z_1, z_1+\Delta z]}}_{\text{Convergence is assured for } \Delta z \to 0}$$

Using $\bar{z}$ at the lower integration limit (Euler's integration) then (14) can be rewritten as

(15) $$\left[\zeta_x^{t+\Delta t} - \zeta_x^{\Delta t}\right]\Delta x = -\left[(u\zeta)^t_{x+\frac{\Delta x}{2}} - (u\zeta)^t_{x-\Delta x/2}\right]\Delta t + \alpha\left[\frac{\partial\zeta}{\partial x}\Big|^t_{x+\frac{\Delta x}{2}} - \frac{\partial\zeta}{\partial x}\Big|^t_{x-\frac{\Delta x}{2}}\right]\Delta t$$

The first derivatives can be evaluated as

$$\zeta^t_{x+\Delta x} = \zeta^t_x + \int_x^{x+\Delta x} \frac{\partial\zeta}{\partial x}\,dx$$

or

$$\frac{\partial\zeta}{\partial x}\Big|^t_{x+\Delta x/2} \simeq \frac{\zeta^t_{x+\Delta x} - \zeta^t_x}{\Delta x}$$

$$(u\zeta)^t_{x+\Delta x/2} = \frac{1}{2}\left[(u\zeta)^t_x + (u\zeta)^t_{x+\Delta x}\right] \Longrightarrow$$

(16) $$\frac{\zeta_x^{t+\Delta t} - \zeta_x^{t}}{\Delta t} = -\frac{(u\zeta)^t_{x+\Delta x} - (u\zeta)^t_{x-\Delta x}}{2\Delta x} + \alpha\,\frac{\zeta^t_{x+\Delta x} + \zeta^t_{x-\Delta x} - 2\zeta^t_x}{\Delta x^2}$$

Integration from $t-\Delta t$ to $t+\Delta t$ will give centered in time

Advantage of this method is better appreciated in non rectangular coordinate systems and because of the conservation property.

3.2 Truncation errors, consistency, stability, and convergence

Suppose $u(x,t)$ is the exact solution to the initial value problem

$$\frac{\partial u}{\partial t} = L(x,t) \qquad (17)$$

and ~~U~~ $(n\Delta t, j\Delta x)$ ~~be~~ $= U_j^n$ is the solution to the FD approximation of (17). This approximation must be consistent, stable and must converge to be useful in physical problems.

Consistency: A FD approximation is consistent with a differential equation is the FD equation converges to the correct differential equation as the space and time grid spacing → 0

Stability: If U_j^n is the numerical solution and u_j^n the exact solution at $t = n\Delta t$ and $x = j\Delta x$, then the FD approximation is stable if $z_j^n = U_j^n - u_j^n$ remains bounded as n tends to infinity for fixed Δt.

Convergence: If the difference between the theorical solutions of the FD and Differential equations at a fixed point (x,t) tends to zero as $\Delta t \to 0$ and $\Delta x \to 0$ and $n, j \to \infty$, then the finite difference approximation converges to the continuous equation.

Truncation error: The local difference between the FD approximation and the Taylor series representation of the continuous problem as a fixed point is the truncation error.

Theorem (Lax and Richtmyer)

Given a properly posed linear initial value problem and a finite difference approximation to it that satisfies the consistency condition, stability (as Δx and $\Delta t \to 0$) is the necessary and sufficient condition for convergence.

Example:

Let's consider the one-dimensional advection equation with constant speed c

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \qquad (18)$$

The Taylor series for second order derivatives are

$$u_j^{n+1} - u_j^{n-1} = 2\Delta t\left(\frac{\partial u}{\partial t}\right)_j^n + \frac{\Delta t^3}{3}\left(\frac{\partial^3 u}{\partial t^3}\right)_j^n + \cdots$$

$$u_{j+1}^n - u_{j-1}^n = 2\Delta x\left(\frac{\partial u}{\partial x}\right)_j^n + \frac{\Delta x^3}{3}\left(\frac{\partial^3 u}{\partial x^3}\right)\Bigg|_j^n + \cdots$$

Combining we obtain

addition

$$(19)\quad \left\{\begin{aligned} &\overbrace{\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t}}^{\text{FD}} = \overbrace{-c/_{2\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right)}^{\text{Approximation}} \\ &+ \cancel{\left(\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x}\right)_j^n} = 0 \text{ by definition (Eqn. 18)} \\ &+ \underbrace{\frac{\Delta t^2}{12}\left[\left(\frac{\partial^3 u}{\partial t^3}\right)_j^{\eta_1} + \left(\frac{\partial^3 u}{\partial t^3}\right)_j^{\eta_2}\right] + \frac{c\Delta x^2}{12}\left[\left(\frac{\partial^3 u}{\partial x^3}\right)_{\xi_1}^n + \left(\frac{\partial^3 u}{\partial x^3}\right)_{\xi_2}^n\right]}_{\text{Truncation error} = \varepsilon_j^n} \end{aligned}\right.$$

Addition

or expanding into ~~a~~ finite Taylor series (mean-value theorem)

$$u_{j\pm1}^n = u_j^n \pm \Delta x \left(\frac{\partial u}{\partial x}\right)_j + \frac{\Delta x^2}{2}\left(\frac{\partial^2 u}{\partial x^2}\right)_j \pm \frac{\Delta x^3}{6}\left(\frac{\partial^3 u}{\partial x^3}\right)^n_{\xi}$$

where ξ is located in $[x_j, x_{j+1}]$
$[x_{j-1}, x_j]$

Substraction + division by $2\Delta x$ gives

$$\frac{u_{j+1} - u_{j-1}}{2\Delta x} = \left(\frac{\partial u}{\partial x}\right)^n_j + \frac{\Delta x^2}{12}\left(\left(\frac{\partial^3 u}{\partial x^3}\right)^n_{\xi_1} + \left(\frac{\partial^3 u}{\partial x^3}\right)_{\xi_2}\right)$$

with $\xi_1 \in [x_j, x_{j+1}]$
$\xi_2 \in [x_{j-1}, x_j]$

$\Rightarrow$ the error introduced by this approx is of order Δx^2 provided that the third derivative of solution remain bounded

In time,

$$u_j^{n+1} = u_j^n + \Delta t \left(\frac{\partial u}{\partial t}\right)^n_j + \frac{\Delta t^2}{2}\left(\frac{\partial^2 u}{\partial t^2}\right)^n_j$$

$$t_n \leq \eta \leq t_{n+1}$$

This FD approximation is consistent if the truncation error which is $O(\Delta t^2 + \Delta x^2)$ goes to zero as $\Delta t, \Delta x \to 0$

From (19)

$$\left|\varepsilon_j^n\right| \leq \frac{\Delta t^2}{12\cancel{6}} M_1 + |c| \frac{\Delta x^2}{\cancel{6}12} M_2$$

where M_1 and M_2 are the bounds for $\left|\frac{\partial^3 u}{\partial t^3}\right|$ and $\left|\frac{\partial^3 u}{\partial x^3}\right|$ respectively. Note that these bounds hold for the true solution, i.e. they are independent of the numerical treatment of the equation. Therefore $\varepsilon_j^n \to 0$ as $\Delta x, \Delta t \to 0$

If we consider only finite-difference forward in time, then

$$\left|\varepsilon_j^n\right| \leq \frac{\Delta t}{2} M_3 + |c| \frac{\Delta x^2}{12} M_4$$

where M_3 and M_4 are the bounds for $\left|\frac{\partial^2 u}{\partial t^2}\right|$ and $\left|\frac{\partial^3 u}{\partial x^3}\right|$ respectively

We are now interested in the accumulated error of the FD solution. If we consider the latter (FD forward in time)

$$(20) \quad U_j^{n+1} = U_j^n - \frac{\lambda}{2}\left(U_{j+1}^n - U_{j-1}^n\right) \quad \text{with } \boxed{\lambda = \frac{c\Delta t}{\Delta x}}$$

$$(21) \quad u_j^{n+1} = u_j^n - \frac{\lambda}{2}\left(u_{j+1}^n - u_{j-1}^n\right) + \Delta t\, \varepsilon_j^n$$

The accumulated error is $e_j^n = u_j^n - U_j^n$

By subtraction of (21) to (20),

$$(22) \quad e_j^{n+1} = e_j^n - \frac{\lambda}{2}\left(e_{j+1}^n - e_{j-1}^n\right) + \Delta t\, \varepsilon_j^n$$

By defining $\mathcal{E}^{n} = \max_{j} |e_j^n|$ and $\varepsilon = \max_{j,n} |\varepsilon_j^n|$

then

(23) $$\boxed{\mathcal{E}^{n+1} \leqslant (1+|\lambda|)\,\mathcal{E}^{n} + \Delta t\,\varepsilon}$$

Successive use of this recursion formula does not lead to a finite bound for $\mathcal{E}$

$$\mathcal{E}^{n+1} \leq (1+|\lambda|)\left[(1+|\lambda|)\,\mathcal{E}^{n-1} + \Delta t\,\varepsilon\right] + \Delta t\,\varepsilon$$

$$\leq - - - -$$

$$\leq \left[1 + (1+|\lambda|) + (1+|\lambda|)^2 + \cdots + (1+|\lambda|)^n\right]\Delta t\,\varepsilon$$

if $\mathcal{E}^0 = 0$

$$\leqslant \frac{(1+|\lambda|)^n - 1}{|\lambda|}\,\Delta t\,\varepsilon$$

$$\leqslant \frac{\varepsilon\,\Delta x}{|c|}\left[\left(1 + \frac{|c|\,t}{n\,\Delta x}\right)^n - 1\right]\frac{\varepsilon\,\Delta x}{|c|} \qquad \left(\text{with } \Delta t = \frac{t}{n}\right)$$

$$\leqslant \frac{\varepsilon\,\Delta x}{|c|}\left(e^{\frac{|c|\,t}{\Delta x}} - 1\right)$$

which goes to ∞ as $\left(\begin{array}{l}\Delta x \to 0 \\ n \to \infty\end{array}\right.$

Failure to find an upper limit for the error does not imply that this error will grow indefinitely. This can be done only by a practical test.

For this case, it turns out that an upper limit can be found if we replace U_j^n of (20) by $\frac{1}{2}\left(U_{j-1}^n + U_{j+1}^n\right)$

Then, instead of (22), we have

(24) $$e_j^{n+1} = \left(\tfrac{1}{2} + \tfrac{\lambda}{2}\right) e_{j-1}^{n} + \left(\tfrac{1}{2} - \tfrac{\lambda}{2}\right) e_{j+1}^{n} + \Delta t\,\varepsilon_j^n$$

or

(25) $$\mathcal{E}^{n+1} \leq \left(\left|\tfrac{1}{2} + \tfrac{\lambda}{2}\right| + \left|\tfrac{1}{2} - \tfrac{\lambda}{2}\right|\right)\mathcal{E}^{n} + \Delta t\,\varepsilon$$

As long as $|\lambda| \leq 1$ (CFL criteria)

$$E^{n+1} \leq E^{n} + \Delta t \varepsilon$$

$$\leq n \Delta t \varepsilon = t \varepsilon$$

The accumulated error at a fixed time is then proportional to the truncation error ε.

From Taylor series expansion

$$\tfrac{1}{2}\left(u^{n}_{j-1} + u^{n}_{j+1}\right) = u^{n}_{j} + \frac{\Delta x^2}{4}\left(\left(\frac{\partial^2 u}{\partial x^2}\right)^{n}_{j} + \left(\frac{\partial^2 u}{\partial x^2}\right)^{n}_{j}\right)$$

The overall truncation error can be bounded by

$$\left|\varepsilon^{n}_{j}\right| \leq \Delta t \frac{M_1}{2} + \Delta x \frac{|c| M_2}{2\lambda} + \Delta x^2 \frac{|c| M_3}{6}$$

where M_1, M_2 and M_3 are upper bounds for $\frac{\partial^2 u}{\partial t^2}$, $\frac{\partial^2 u}{\partial x^2}$, $\frac{\partial^3 u}{\partial x^3}$ respectively.

Thus the scheme is <u>consistent</u> first $(\lambda \neq 0)$ and E, the accumulated error vanishes as the mesh width goes to zero

$$\left\{\begin{array}{l} \lim\limits_{\substack{\Delta x \to 0 \\ \Delta t \to 0 \\ \lambda < 0}} u^{n}_{j} = u(x,t) \end{array}\right.$$

This FD scheme is then <u>convergent</u>

3.3 Normes and numerical stability analysis

a. <u>Vector and matrix norms and stability definition.</u>

Stability is associated with the property of a numerical solution which remain finite at all points in the (x,t) domain. (Unstable ⇔ blow up of the solution)

R

~~A~~ vector norm is defined as a measure of a vector in real-number space. The norm must satisfy

- $\|\vec{x}\| \geqslant 0$, $\|\vec{x}\| = 0 \iff \vec{x} = \vec{0}$
- $\|\alpha \vec{x}\| = |\alpha| \, \|\vec{x}\|$ for any scalar α
- $\|\vec{x} + \vec{y}\| \leqslant \|\vec{x}\| + \|\vec{y}\|$ for any $\vec{x}, \vec{y}$

A frequently used norm is the L_p norm.

$$\|\vec{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$

where $\vec{x} = (x_j)$ is an n-dimensional vector. Most usual ore (a) Euclidian norm, $p=2$ (L_2)

(L_∞) (b) "maximum" norm, $p=\infty$ $\|x\|_\infty = \max_j |x_j|$

(c) (L_1) norm, $p=1$ $\|x\|_1 = \sum_j |x_j|$

If we define $\vec{U}^n$ by $\vec{U}^n = (U_j^n)$, then a numerical scheme is stable if there exists a number M such that $\|\vec{U}^n\| \leq M \|\vec{U}^0\|$ (M can be a function of time + since solutions do grow in time)

By analogy with the definition of vector norms, we define the matrix norm as a measure of a matrix in real-number space. The following conditions must be satisfied

- $\|A| \geqslant 0$, $\|A\| = 0 \iff A = (0)$
- $\|\kappa A\| = |\alpha| \, \|A\|$ for any scalar α
- $\|A + B\| \leqslant \|A\| + \|B\|$
  $\|A \cdot B\| \leqslant \|A\| \, \|B\|$ for any A, B

The most common norms are as before the L_1, L_2, and L_∞ norms

$\|A\|_1 = \max_j \sum_i |a_{ij}|$ (Sum of all colums.)

$\|A\|_\infty = \max_i \sum_j |a_{ij}|$ (Sum of all rows)

$\|A\|_2 = \sqrt{\lambda(A^T A)}$ where λ is the absolute largest eigenvalue of the ~~[illegible]~~ matrix $A^T A$

Feb 4, 1992

b) The Lax - Richtmyer Theorem

get definition for before

Theorem: (Repeat) Numerical stability and consistency of a finite difference scheme imply convergence.

This theorem is important because it enables us to prove convergence of a numerical solution without explicit knowledge of the exact solution. The FD equation can be rewritten as

(26) $$\vec{U}^{n+1} = L\,\vec{U}^{n} + \vec{R}^{n}\ .$$

where L is a linear operator (expressed in a matrix form) and $\vec{R}^{n}$ is the inhomogeneous part of the equation such as forcing

Another definition of the stability, slightly more restrictive, but for most purposes, equivalent is the following. A finite FD scheme of the type of (26) is stable for any time T and any $\delta > 0$, there exists two values η, ρ such that

$$\|(L)^{n}\| \leq M \qquad \text{for all } \Delta x < \delta\ ,\ \Delta t < \eta \Delta x \text{ and } \eta$$

n → power

provided that $n\Delta t \leq t$

Since $$\|\vec{U}^{n}\| \leq \|(L)^{n}\|\ \|\vec{U}^{0}\| + \|(L)^{n-1}\|\,\|\vec{R}^{0}\| + \ldots + \|(L)^{0}\|\,\|\vec{R}^{n-1}\|$$

and since we can reasonably assume the total forcing $\sum_k \|\vec{R}^{k}\|$ to be be finite, this definition does imply the one given before.

Proof of the LR Theorem:

$$\begin{cases} \vec{U}^{n+1} = L\,\vec{U}^{n} + \vec{R}^{n} \\ \vec{u}^{n+1} = L\,\vec{u}^{n} + \vec{R}^{n} + \Delta t\ \vec{\varepsilon}^{n} \end{cases}$$

The accumulated error vector $\vec{e}^{\,n+1}$ is then

$$\vec{e}^{\,n+1} = L\,\vec{e}^{\,n} + \Delta t\,\vec{\varepsilon}^{\,n}$$

$$= L\left(L e^{n-1} + \Delta t\,\vec{\varepsilon}^{\,n-1}\right) + \Delta t\,\vec{\varepsilon}^{\,n}$$

$$\vdots$$

$$= \left((L)^n\,\vec{\varepsilon}^{\,0} + (L)^{n-1}\,\vec{\varepsilon}^{\,1} + \dots (L)^0\,\vec{\varepsilon}^{\,n}\right)\Delta t$$

$$\Longrightarrow \quad \|\vec{e}^{\,n}\| \leq \Delta t\left(\|(L)^{n-1}\|\,\|\vec{\varepsilon}^{\,0}\| + \dots + \|(L)^0\|\,\|\varepsilon^{n-1}\|\right)$$

Since the scheme is consistent, for every $\varepsilon > 0$, there exist two values δ, η such that $\|\vec{\varepsilon}^{\,k}\| < \varepsilon$ for all $\Delta x < \delta$, $\Delta t < \eta\,\Delta x$

Since the scheme is furthermore stable, we have $\|(L)^k\| \leq M$ for all k, $k\Delta t \leq t$, then

(27) $$\|\vec{e}^{\,n}\| \leq n\Delta t\,\varepsilon M = t\,\varepsilon M$$

Since ε is arbitrarily small, the theorem is proven.

The LR theorem also holds in the opposite direction convergence and consistency $\Longrightarrow$ stability.

c) Stability analysis

The previous theorem allows us to concentrate on the stability of the numerical scheme rather than it's convergence, once you admit consistency.

In section 3.2, we were not able to prove convergence of the scheme

$$u_j^{n+1} = u_j^n - \lambda/2\left(u_{j+1}^n - u_{j-1}^n\right)$$

1 – Using matrix norms

The linear operator applicable in this case is

$$L = \begin{pmatrix} 1 & -\lambda/2 & & & \lambda/2 \\ \lambda/2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -\lambda/2 \\ -\lambda/2 & & & \lambda/2 & 1 \end{pmatrix}$$

Amplification matrix

Since $\|L^n\| \leq \|L\|^n$, stability is assured if $\|L\| \leq 1$
Actually, $\|L\| \leq 1 + O(\Delta t)$ is sufficient since

$$\lim_{n\to\infty} \|L\|^n \leq \lim \left(1 + \frac{O(t)}{n}\right)^n = e^{O(t)}$$

which is compatible with the previous definition
This criteria is named after Von Neuman.

~~Take fro ad of (17)~~

We find that $\|L_1\| = \|L_\infty\| = 1 + |\lambda|$
Since $\lambda = \dfrac{c\Delta t}{\Delta x}$, the assumption $|\lambda| = O(\Delta t)$ would imply $\Delta x =$ constant. This is incompatible with the limit process $\Delta x, \Delta t \to 0$. Hence neither L_1 or L_∞ can be used. The L_2 norm requires knowledge of the eigenvalues of $L^T L$

* The linear operator for the diffusive scheme (24) is

$$L = \begin{pmatrix} 0 & (\tfrac{1}{2}-\tfrac{\lambda}{2}) & & & (\tfrac{1}{2}+\tfrac{\lambda}{2}) \\ (\tfrac{1}{2}+\tfrac{\lambda}{2}) & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & (\tfrac{1}{2}-\tfrac{\lambda}{2}) \\ (\tfrac{1}{2}-\tfrac{\lambda}{2}) & & & (\tfrac{1}{2}+\tfrac{\lambda}{2}) & 0 \end{pmatrix}$$

U_j^n is replaced by $\frac{1}{2}\left(U_{j-1}^n + U_{j+1}^n\right)$.

We find that

$$\|L_3\|_1 = \|L\|_\infty = \begin{cases} 1 & \text{if } |\lambda| \leq 1 \\ |\lambda| & \text{if } |\lambda| > 1 \end{cases}$$

Hence, in this case, stability is assured as long as $|\lambda| \leq 1$ (Same as for convergence) a addition

Let's know consider the following parabolic differential equation

(1) $$\frac{\partial F}{\partial t} = K \frac{\partial^2 F}{\partial x^2}$$

$$\frac{F_j^{n+1} - F_j^n}{\Delta t} = K \frac{F_{j+1}^n - 2F_j^n + F_{j-1}^n}{\Delta x^2}$$

or

$$F_j^{n+1} = \lambda F_{j-1}^n + (1-2\lambda) F_j^n + \lambda F_{j+1}^n$$

with $\lambda = \frac{K \Delta t}{\Delta x^2}$

If the boundary values $F_0^n = F_J^n = 0$, then

$$\begin{pmatrix} F_1^{n+1} \\ \vdots \\ F_j^{n+1} \\ \vdots \\ F_{J-1}^{n+1} \end{pmatrix} = \begin{pmatrix} 1-2\lambda & \lambda & & 0 \\ \lambda & 1-2\lambda & \lambda & \\ & \lambda & \ddots & \ddots \\ & & \ddots & \ddots \end{pmatrix} \begin{pmatrix} F_1^n \\ \vdots \\ \vdots \\ F_{J-1}^n \end{pmatrix}$$

(2) $$\mathbf{F}_n = \underbrace{\mathbf{A}}_{\text{amplification matrix}} \mathbf{F}_{n-1} = \mathbf{A}^n \mathbf{F}_0$$

The eigenvalues μ_i of L are the roots of

$$|L - \mu I| = 0$$ where I is the identity matrix. Determinant of order $J-1$ $\Rightarrow$ $J-1$ eigenvalues. Associated with each eigenvalues is an eigenvector v_i which satisfies $L v_i = \mu_i v_i$ $i = 1, 2$ ---

Eigenvectors $\Leftrightarrow$ base $\Rightarrow$ $F_0 = \sum_i c_i v_i$

using (2)

$$F_n = \sum_i c_i L^n v_i = \sum_i c_i L^{n-1} \underbrace{L v_i}_{\mu_i v_i}$$

$$= \dots = \sum_i c_i \mu_i^n v_i$$

stable if $|\mu_i| \le 1$ for all i

Can be allowed for some growth namely

$$|\mu_i| \le 1 + O(\Delta t)$$

(Spectral radius).

Remember that this scheme was not perfectly consistent and $|\lambda|$ is bounded away from 0. Both $|\lambda| < 1$ and $|\lambda| \geq \lambda_0 > 0$ must be satisfied for convergence

Addition

2- Using Fourier Methods (or Von Neuman analysis)

The previous method is attractive, but often difficult to put in practice in more complicated situations. A less general, but simpler method is based on a Fourier decomposition of solution U_j^n:

(28) $$U_j^n = \sum_{k=-J}^{J} A_k^n e^{ikx_j}$$

The exact solution is

(29) $$u(x,t) = \sum_{k=-n}^{n} B_k(t) e^{ikx}$$

We can determine the amplitudes $B_k(t)$ term by term. Each $B_k(t)$ has then to satisfy

$$\left(\sum_{-n}^{n}\left(\frac{\partial B_k}{\partial t} + cikB_k\right)e^{ikx} = 0\right)$$

(30) $$\frac{\partial B_k}{\partial t} = -ikcB_k$$

or

(31) $$B_k = a_k e^{-ikct}$$ where $a_k = B_k(0)$ represents the initial conditions.

Let's throw (28) in (22) $\quad x_{j+1} = x_j + \Delta x$

(32) $$U_j^{n+1} = \sum A_k^n e^{ikx_j} - \lambda/2 \left[\sum A_k^n \underbrace{\left(e^{ikx_{j+1}} - e^{ikx_{j-1}}\right)}_{2i\sin(k\Delta x)\, e^{ikx_j}}\right]$$

$$= \sum A_k^n \left(1 - i\lambda \sin(k\Delta x)\right) e^{ikx_j}$$

$$= \sum A_k^{n+1} e^{ikx_j}$$

or $$A_k^{n+1} = A_k^n \left(1 - i\lambda \sin (k\Delta x)\right)$$

The ratio A_k^{n+1}/A_k^n is called the amplification factor G

(33) $$G = 1 - i\lambda \sin (k\Delta x) \quad ; \quad A_k^{n+1} = G A_k^n$$

If solutions are to remain bounded, then we have $|G| \le 1$ (Von Neuman).

$$|G|^2 = (1 - i\lambda \sin k\Delta x)(1 + i\lambda \sin k\Delta x)$$
$$= 1 + \lambda^2 \sin^2 k\Delta x$$

which shows that (22) is unstable for all Δt.

Exercise: Same for diffusion equation
— for both together.

Useful

$\boxed{\text{Von Neuman condition}}$ (more restricted) than before.

Sufficient

Two-time level

$$\underbrace{A_k^{n+1}}_{\substack{\text{vector}\\ \text{for a system}\\ \text{of partial dif.}}} = \underbrace{G}_{\text{amplification matrix of the scheme.}} A_k^n$$

$$= G^n A_k^0$$

The scheme is stable if

spectral radii of the matrix G $= S_R(G)$

$$|\mu_i| \le 1 + O(\Delta t) \quad \text{for all } i$$

where μ_i are the eigenvalues of the amplification matrix G since we have

$$(S_R^{(G)})^n \le \|G^n\| \le \|G\|^n.$$

See (Richtmyer, ~~[illegible]~~ for details)

Models to Cover

Liz Smith

NODS ; JPL

Ocean Atlas 1.0.

Met

Low-order models

GFDL Global Spectral model (Linda)

Dynamical core only

Mid- April (2d Week)

PO

QG — ix - Holland
2 layers + spectral in the vertical

LBF

BB

SPEM

PRINCETON (Perry)

GFDL (Doc ?)