26th September 2000

Differential Equations

For certain (*simple*) cases, DE's may be solved **analytically**, but usually **numerical solution methods** are necessary. Even when an *analytical* solution is possible, it is often easier to solve numerically rather than evaluate the *theoretical* expression.

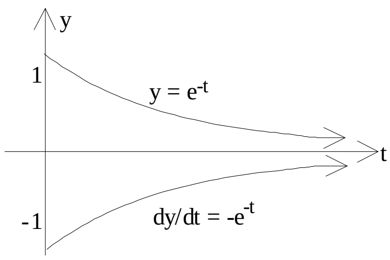
Ordinary Differential Equations

To start with, we shall concentrate on *first order ODE's* ($^{d}/_{dt}$, not $^{d^2}/_{dt^2}$, $^{d^3}/_{dt^3}$, etc.). *Initial value* problems may be expressed generally as $\underline{y}' = \underline{f}(t, \underline{y})$, where $t \geq t_0$, and $\underline{y}(t_0) = \underline{y}_0$. This *general* form gives a d-dimensional system of ODE's for a set of **dependent** variables $\underline{y} = (y_{(1)}\ y_{(2)} \ldots y_{(d)})^T$ which depend on the *independent* variable t.

Writing the system **fully** gives $dy_{(1)}/dt = f_1(t, \underline{y})$; .....; $dy_{(d)}/dt = f_d(t, \underline{y})$; the *initial condition* is $y_{(1)}(t_0) = y_{(1)0}, \ldots, y_{(d)}(t_0) = y_{(d)0}$; and the ODE holds *for all time t* after the initial time t_0. One Dimensional Systems of ODE's. In *general*, we have $y' = f(t, y)$, with $t \geq t_0$, and $y(t_0) = y_0$. *Examples*: $^{dy}/_{dt} = \sin t + \sin y$; $^{dy}/_{dt} = t^2+y^2$.

Two Dimensional Systems of ODE's. An example may be *derived* from a second order ODE: $^{d^2y}/_{dt^2} + \sin(t)^{dy}/_{dt} + \cos(t)y = 3$, with *initial* conditions $y(t_0) = 1$, and $(^{dy}/_{dt})_{t=t_0} = -1$. Put $z = {}^{dy}/_{dt}$, so that the above *second order ODE* may be written as $^{dz}/_{dt} = -\sin(t)z - \cos(t)y + 3$; $^{dy}/_{dt} = z$, with *initial conditions* $z(t_0) = -1$, and $y(t_0) = 1$. Therefore, $\underline{y} := (^{z}_{y})$.

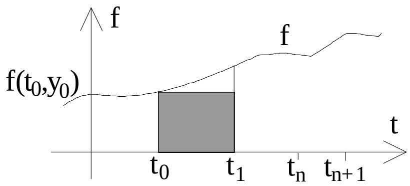
Example: A simple *one-dimensional* system is given by $^{dy}/_{dt} = -y$, with initial condition $y(0) = 1$. This may be solved *analytically:* $\int {}^{dy}/_{y} = -\int dt$, which *integrates* to give $\log y = -t + C$. Take the **exponential**, giving $y = e^{-t+C} = e^{-t}e^{C} = Ae^{-t}$. ($A = e^{C}$). The *constant* of integration may be found from the initial condition $y(0) = 1$ substitute this into the **solution** $y = Ae^{-t}$, giving $1 = Ae^{0}$. Therefore, $1 = A$; and the *solution* is $y = e^{-t}$.



27th September 2000

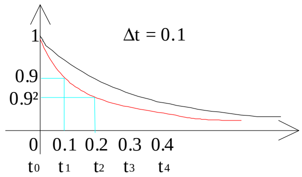
We have *solved* for a simple case. Consider how we approximate a solution in the **general** case for a 1-D system $y' = f(t,y)$, with $t \geq t_0$, and $y(t_0) = y_0$. We need to *integrate* $^{dy}/_{dt} = f(t,y)$ w.r.t. time. To calculate y at a *future* time $t_1 > t_0$, we need to *integrate* from t_0 to t_1, $\int_{t_0}^{t_1} {}^{dy}/_{dt}dt = \int_{t_0}^{t_1} f(t, y(t))dt$. Now $\int_{t_0}^{t_1} dy = \int_{t_0}^{t_1} f(t,y)dt$; $y(t_1)-y(t_0) = \int_{t_0}^{t_1} f(t,y)dt$; $y(t_1) = y(t_0)+ \int_{t_0}^{t_1} f(t,y)dt$. Note that we have *converted our DE* to an integral equation.

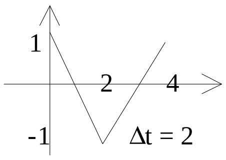
If we want an **approximate** value for $y(t_1)$, where t_1-t_0 is *not too large*, then we could approximate $f(t,y)$ within the interval $[t_0, t_1]$ by its value at t_0, i.e. $f(t_0, y_0)$. So $y(t_1) \approx y_0 + \int_{t_0}^{t_1} f(t_0,y_0)dt$; $y(t_1) \approx y_0+(\int_{t_0}^{t_1}dt)f(t_0,y_0)$; $y(t_1) \approx y_0+(t_1-t_0)f(t_0,y_0)$. We denote the *numerical* approximation of $y(t_1)$ as y_1, so that $y_1 = y_0+(t_1-t_0)f(t_0,y_0)$.



Having **found** an approximation y_1 at time t_1, we can then use it to find an *approximation* at time $t_2 > t_1$, where $y_2 = y_1 + (t_2-t_1)f(t_1, y_1)$. This procedure can be *continued* to produce approximates for t_3, t_4, etc., where, in **general**, we have $y_{n+1} = y_n + (t_{n+1}-t_n)f(t_n, y_n)$. Given a *sequence* t_0, $t_1 = t_0+\Delta t$, $t_2 = t_1+\Delta t$, ..., $t_{n+1} = t_n+\Delta t$, where Δt is the *time step*, we have $y_{n+1} = y_n+\Delta t.f(t_n, y_n)$, for n = 0,1,2,... This is known as the **<u>EULER METHOD</u>**, giving an *approximate solution* at the discrete times t_1, t_2,...

<u>Example</u>: Apply *Euler's* method to $y' = -y$, with $t \geq 0$, and $y(0) = 1$. Here, $y_{n+1} = y_n+\Delta t(-y_n)$; $y_{n+1} = (1-\Delta t)y_n$. First choose a *time step*, say $\Delta t = 0.1$, then step *forward* in time from t = 0: $y_0 = 1$ (we are *given* the initial value); $y_1 = (1-\Delta t)y_0 = 0.9y_0 = 0.9$; $y_2 = (0.9)y_1 = (0.9)^2$; ...; $y_n = (0.9)^n$. This seems to be a *decent* approximation to the **analytical** solution — it follows the <u>general</u> behaviour.

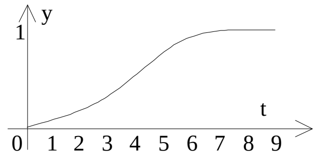




However, for **efficiency**, try a larger time step, say $\Delta t = 2$. Euler's scheme gives $t_{n+1} = (1-2)t_n = -t_n$. This approximant *oscillates* between 1 and -1. Trying $\Delta t = 3$, we get $y_{n+1} = -2y_n$. This oscillates *without* bound. Trying $\Delta t = 1$, **Euler's method** gives $y_{n+1} = 0y_n$, so that $y_0 = 1$; $y_1 = 0 = y_2 = ... = y_n$ (for $n \geq 1$).

28th September 2000

<u>Example</u>: Consider the scalar logistic equation $y' = y(1-y)$, with *initial condition* $y(0) = {}^1/_{10}$. Obtain an **approximate** numerical solution using Euler's method, with time step $\Delta t = 1$. Here, $y_{n+1} = y_n + \Delta t.f(t_n,y_n) = y_n+1(y_n(1-y_n)) = 2y_n-y_n^2$. So $y_0 = {}^1/_{10}$; $y_1 = {}^2/_{10}-{}^1/_{10}{}^2 = {}^{19}/_{100}$; $y_2 = {}^{38}/_{100}-{}^{19^2}/_{100^2} = {}^{3439}/_{10000}$; and so on: $y_3 = 0.5695$; $y_4 = 0.81479798$; $y_5 = 0.965$; $y_6 = 0.99883$; $y_7 = 0.9999986$; $y_8 = 0.99....$; etc. It is also possible to *solve* $y' = y(1-y)$ **analytically**: by using separation; by using partial fractions; and by using normal *log* integrations.

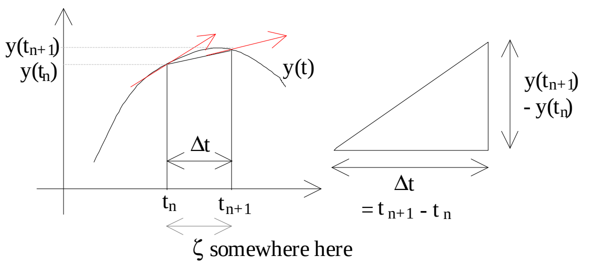


3rd October 2000

<u>Derivation of Euler's Method by Finite Difference Formulae</u>

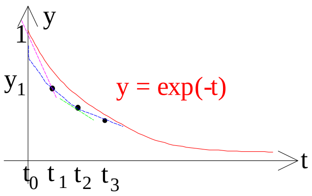
Previously, we have derived *Euler's method* by converting the differential equation to an integral equation, and then approximating the integral by using a <u>rectangle</u> rule. We now derive Euler's method by using a finite **difference** approximation to the derivative at discrete times t_0, $t_1 = t_0+\Delta t$, $t_2 = t_0+2\Delta t$, ..., $t_n = t_0+n\Delta t$, and so on.

The *derivative* $y' = {}^{dy}/_{dt}$ at time t_n is approximated by a *forward difference formula* ${}^{dy}/_{dt}|_{t_n} \approx y(t_{n+1})-y(t_n)/\Delta t$. This may be seen graphically *as shown* on the right. From this formula, we convert the **differential** equation into a **difference** equation $(y_{n+1}-y_n)/\Delta t = f(t_n,y_n)$. Rearranging, we obtain *Euler's method*, $y_{n+1} = y_n+\Delta t.f(t_n,y_n)$. The *finite difference approximation* is obtained by considering the <u>Taylor</u> series expansion of y(t) about $t = t_n$, *remembering* that $y(t_{n+1}) = y(t_n+\Delta t)$. So $y(t_n+\Delta t) = y(t_n)+\Delta t{}^{dy}/_{dt}|_{t_n} + {}^{\Delta t^2}/_{2!}{}^{d^2y}/_{dt^2}|_{t_n} + {}^{\Delta t^3}/_{3!}{}^{d^3y}/_{dt^3}|_{t_n} + ...$, which may be *written* as $\sum_{k=0}^{\infty} \frac{\Delta t^k}{k!}\frac{d^k y}{dt^k}\Big|_{t_n}$.



Another form (which we shall *immediately* use) is $y(t_n+\Delta t) = y(t_n) + \Delta t\,{}^{dy}/_{dt}|_{t_n} + {}^{\Delta t^2}/_{2!}\,{}^{dy}/_{dt}|_{t=\zeta}$, where $t_n \le \zeta \le t_{n+1}$. Truncating the Taylor series after the *second* term gives $y(t_{n+1}) \approx y(t_n)+\Delta t\,{}^{dy}/_{dt}|_{t_n}$, with an *error* of $O(\Delta t^2)$. **Rearranging**, we obtain ${}^{dy}/_{dt}|_{t_n} \approx y(t_{n+1})-y(t_n)/\Delta t$, with an *error* of $O(\Delta t)$, so the **approximation** is of first order. Higher order methods may be derived by retaining *more* terms in the Taylor series.

The Geometric Interpretation of Euler's Method



Consider the *graph* shown, which shows the true solution to an ODE, together with the **approximation** obtained with Euler's method. The ODE is $y' = -y$, with $t \geq 0$, and $y(0) = 1$; and the *difference* scheme used is given by $y_{n+1} = y_n - \Delta t y_n$.

4th October 2000

Existence, Uniqueness and Stability for First Order ODE Initial Value Problems

Up to now, we have assumed that we can *solve* the IVP $y' = f(t,y)$. The following theorems give conditions on the functional form of $f(t,y)$ so that a solution exists; so that the solution is **unique;** and so that it is stable to **perturbations** in the initial conditions. The main conditions on $f(t,y)$ are that it is *continuous* in t and y, and that $f(t,y)$ satisfies a Lipschitz condition in y. **Definition**: Lipschitz Condition: A function $f(t,y)$ is said to satisfy a *Lipschitz* condition on the variable y (on a set D) provided a constant $L > 0$ exists, with the property that $|f(t,y_1)-f(t,y_2)| < L|y_1-y_2|$, *where* (t,y_1) and $(t,y_2) \in D$. Here, D **could** be the set $a \le t \le b$, with $\alpha \le y \le \beta$.

Theorem: Existence Theorem: Suppose that D is the set $\{(t,y) \mid a \le t \le b, -\infty < y < \infty\}$; that $f(t,y)$ is a *continuous* function; and that it satisfies a *Lipschitz* condition on D in the variable y. It follows that the **initial** value problem $y' = f(t,y)$, with $a \le t \le b$, and where $y(a) = y_a$, has a *unique* solution $y(t)$ for $a \le t \le b$. The proof is *omitted*, where there are 2 main types of proof: one based on the Euler method, and the other based on the Picard iteration of the equivalent integral equation.

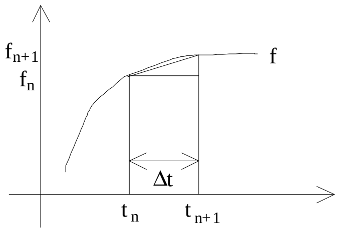
Theorem: Uniqueness Theorem: Under the same *assumptions* as above, the solution $y(t)$ for $a \le t \le b$ is unique. Proof *omitted*. It is important to know how the solution of the ODE behaves if there is a small **change** in the initial condition(s). The problem is said to be *stable* if the solution to a perturbed problem stays close to the unperturbed problem. Mathematically, there are *positive* constants ε and k, with the property that a unique solution $z(t)$ to the **problem** exists, (the problem is given by $z' = f(t,z)+\delta(t)$; with $a \le t \le b$; and where $z(a) = y_a+\varepsilon_0$), with $|z(t)-y(t)| < k_\varepsilon$ for all $a \le t \le b$ — *whenever* $|\varepsilon_0| < \varepsilon$, and whenever $|\delta(t)| < \varepsilon$. A *well-posed problem* is one which (a) has a **unique** solution, and (b) is *stable* to perturbation.

Notes: An easy method to see whether a function *satisfies* a Lipschitz condition is to use the following result: **Theorem**: If $f(t,y)$ satisfies $|{}^{\partial f}/_{\partial y}(t,y)| < L$ for all (t,y) on the *set*, then f satisfies a Lipschitz condition on the set in the **variable** y, with Lipschitz constant L. Proof *omitted*. Example: $y' = -y$. Here, $f(t,y) = -y$, and ${}^{\partial f}/_{\partial y} = -1$, so that $|{}^{\partial f}/_{\partial y}| = 1$. **Theorem**: The *Euler* method is convergent for functions $f(t,y)$ satisfying the *Lipschitz* condition.

5th October 2000

Multi-Step Methods

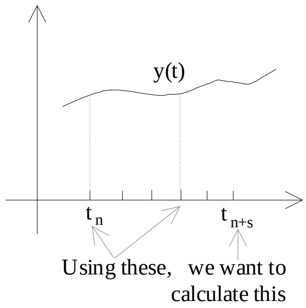
Recall that the **Euler** method, $y_{n+1} = y_n + \Delta t f(t_n, y_n)$, is not very *accurate* (because it is a first order method), and that it is not very stable. **Trapezoidal Euler Method** (the *modified* Euler method): this may be obtained by using a **trapezoidal** approximation to integrate the equation $y_{n+1}-y_n = \int_{t_n}^{t_{n+1}} f(t,y)dt$, giving $y_{n+1} = y_n+\Delta t\,{}^{1}/_{2}[f(t_n,y_n)+f(t_{n+1},y_{n+1})]$ Note: the *square* brackets contains the average of f(t,y) at (t_n, y_n) and (t_{n+1}, y_{n+1}).



Theta Methods. These methods *generalise* the trapezoidal rule by giving a weighted average of $f(t_n,y_n)$ and $f(t_{n+1},y_{n+1})$, to give $y_{n+1} = y_n+\Delta t[(1-\theta)f(t_n,y_n) + \theta f(t_{n+1},y_{n+1})]$, where $0 \le \theta \le 1$. *Notes*: (i) $\theta = 0$ gives the **Euler** method. (ii) $\theta = {}^{1}/_{2}$ gives the *Trapezoidal* method. (iii) $\theta = 1$ gives $y_{n+1} = y_n+\Delta t f(t_{n+1},y_{n+1})$, which is known as the *Backward Euler method*, used extensively with **stiff** equations.

For $\theta = 0$, the θ method is explicit. For $\theta \neq 0$, the θ method is implicit. The θ method is a *one-step method:* it involves **values** at t_n and t_{n+1}. They discard values calculated *previously*, i.e. y_{n-1}, y_{n-2}, etc. This does follow *logically* from the mathematics of **initial** value problems, where the solution depends on a single initial condition. If care is taken, previous values of y may be used to obtain more *accurate* methods — called multistep methods.

A general s-step method may be written in the form $\Sigma^{s}_{m=0}\, a_m y_{n+m} = \Delta t \Sigma^{s}_{m=0}\, b_m f(t_{n+m},y_{n+m})$, *where* n = 0,1,2,..., and where a_m and b_m are given *constants* which are independent of Δt, n, and the differential equation being solved. It is usual to **normalise** a_s to be unity, i.e. to obtain $a_s = 1$. If $b_s = 0$, the scheme is *explicit*; otherwise the scheme is *implicit*.



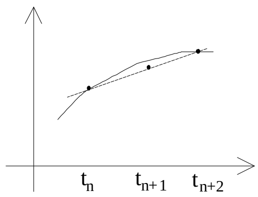
Forward Euler (an explicit) method: $y_{n+1} - y_n = \Delta t f(t_n,y_n)$. Here, s = 1, and $a_1 y_{n+1}+a_0 y_n = \Delta t b_0 f(t_n,y_n)$, so that $a_0 = -1$, $a_1 = 1$, $b_0 = 1$, and all *other* a_m and b_m are zero, with $b_1 = 0$. *Trapezoidal Rule:* $y_{n+1}-y_n = \Delta t({}^{1}/_{2}f(t_n,y_n)+{}^{1}/_{2}f(t_{n+1},y_{n+1}))$, so that s = 1, $a_1 = 1$, $a_0 = -1$, $b_0 = ½$, and $b_1 = {}^{1}/_{2}$. *Backward Euler:* s = 1, $a_1 = 1$, $a_0 = -1$, $b_0 = 0$, and $b_1 = 1$. *q-method*, for $0 \le \theta \le 1$: s = 1, $a_1 = 1$, $a_0 = -1$, $b_0 = (1-\theta)$, and $b_1 = \theta$.

The Adams Methods. An important *subclass* of the multistep methods are known as the Adams Methods. The explicit schemes are called Adams *Bashforth* methods, and the implicit schemes are known as the Adams Moulton *methods*. The coefficients are chosen so as to give **stable** and **accurate** results.

Adams Bashforth **s = 2** (*two* step scheme): $y_{n+2} = y_{n+1} + \Delta t\,{}^{1}/_{2}[3f(t_{n+1},y_{n+1})-f(t_n,y_n)]$. Adams Bashforth **s = 3** (*three* step scheme): $y_{n+3} = y_{n+2} + \Delta t\,{}^{1}/_{12}[23f(t_{n+2},y_{n+2}) - 16f(t_{n+1},y_{n+1}) + 5f(t_n,y_n)]$. Adams Bashforth **s = 4** (*four* step method): $y_{n+4} = y_{n+3} + \Delta t\,{}^{1}/_{24}[55f(t_{n+3},y_{n+3}) - 59f(t_{n+2},y_{n+2}) + 37f(t_{n+1},y_{n+1}) - 9f(t_n,y_n)]$. Adams Bashforth **s = 5**: $y_{n+5} = y_{n+4} + \Delta t/720[1901f(t_{n+4},y_{n+4}) - 2774f(t_{n+3},y_{n+3}) + 2616f(t_{n+2},y_{n+2}) - 1274f(t_{n+1},y_{n+1}) + 251f(t_n,y_n)]$.

Adams Moulton **s = 2:** $y_{n+2} = y_{n+1} + {}^{\Delta t}/_{12}[5f(t_{n+2},y_{n+2}) + 8f(t_{n+1},y_{n+1}) - f(t_n,y_n)]$. *Adams-Moulton* **s = 3:** $y_{n+3} = y_{n+2} + {}^{\Delta t}/_{24}[9f(t_{n+3},y_{n+3}) + 19f(t_{n+2},y_{n+2}) - 5f(t_{n+1},y_{n+1}) + f(t_n,y_n)]$. *Adams Moulton* **s = 4:** $y_{n+4} = y_{n+3} + {}^{\Delta t}/_{720}[251f(t_{n+4},y_{n+4}) + 646f(t_{n+3},y_{n+3}) - 246f(t_{n+2},y_{n+2}) + 106f(t_{n+1},y_{n+1}) - 19f(t_n,y_n)]$.

A *comparison* may be made between an s-step **Adams-Bashforth** method (an explicit method), and an (s-1) step **Adams-Moulton** scheme. Both require *s evaluations* of f at each time step, and both have the same **accuracy** (see the later section on *local truncation* errors).



__**Milne Methods.**__ Milne explicit s = 4 (*Milne’s* method): $y_{n+4} = y_n + \Delta t^4/_3[2f(t_{n+3},y_{n+3}) - f(t_{n+2},y_{n+2}) + 2f(t_{n+1},y_{n+1})]$. Milne implicit s = 2 (*Simpson’* method): $y_{n+2} = y_n + \Delta t^1/_3[f(t_{n+2},y_{n+2}) + 4f(t_{n+1},y_{n+1}) + f(t_n,y_n)]$. Although th *Milne methods* are more accurate than Adams’ methods, they have **limited** use because of problems with stability. __**Mid-point**__ method: $y_{n+2} = y_n + 2\Delta t f(t_{n+1},y_{n+1})$ (see the later material on *Taylor* series methods and Plunge-Kutta methods).

__*Backward Differentiation Formulae (BDF’s)*__. This class of *schemes* is used for certain types of ODE’s — **stiff** equations. s = 1 ⇒ *Backward* Euler formula: $y_{n+1} = y_n + \Delta t f(t_{n+1},y_{n+1})$. s = 2 ⇒ $y_{n+2} - {}^4/_3 y_{n+1} + {}^1/_3 y_n = {}^2/_3 \Delta t f(t_{n+2},y_{n+2})$. s = 3 ⇒ $y_{n+3} - {}^{18}/_{11} y_{n+2} + {}^9/_{11} y_{n+1} - {}^2/_{11} y_n = {}^6/_{11} \Delta t f(t_{n+3},y_{n+3})$.

Predictor-Corrector Methods

In general, *implicit* schemes have improved accuracy and stability as compared to a similar *explicit* scheme. But how can a solution to an implicit scheme be obtained? In general, they can lead to **non-linear** difference equations for y_{n+1} — which are *impossible* to solve. For example, consider $y' = e^y$, with $0 \le t$, and $y(0) = 1$. The *modified* Euler (trapezoidal) method is $y_{n+1} = y_n + {}^{\Delta t}/_2[f(t_n,y_n) + f(t_{n+1},y_{n+1})]$; and for $f(t,y) = e^y$, it *gives* $y_{n+1} = y_n + {}^{\Delta t}/_2[e^{yn} + e^{yn+1}]$. This **cannot** be solved *exactly* for y_{n+1}.

A **popular** strategy is to initially obtain a *first* estimate of y_{n+1}, denoted by y_{n+1}^*, from an **explicit** scheme, the predictor, and then to use it on the *right* hand side of an *implicit* scheme, the corrector. __Example__**:** An Euler / Modified *Trapezoidal* Euler predictor-corrector scheme consists of $y_{n+1}^* = y_n + \Delta t f(t_n,t_n)$ as the *predictor*; and $y_{n+1} = y_n + \Delta t^1/_2[f(t_n,y_n) + f(t_{n+1},y_{n+1}^*)]$ as the corrector.

The corrector may be *iterated* if required, although better accuracy is usually obtained by decreasing Δt. Usually, the **predictor** and the **corrector** are chosen to have the same order of accuracy, which enables good *error analysis*, and enables more intelligent code to approximate the error dynamically — and to adjust the *timestep* accordingly (in variable step size algorithms). **Suitable** predictor-corrector pairs are A-B3 with A-M2; A-B4 with A-M3; A-B5 with A-M4; and Milne explicit 4 with Milne implicit 2.

__*Starting a Multi-Step Scheme*__. When starting a multi-step method from $t = t_0$, a Runge-Kutta method (a single step method of *comparable* accuracy) is used to provide the values for $y_0, y_1, ..., y_{s-1}$ for use in the **s-step method**.

Sources of Errors

The various errors which occur as a *consequence* of replacing a DE by a numerical scheme are all largely referred to as **truncation** errors. When a DE is replaced by a difference equation, a *local truncation error* is made at each forward time step from t_n to t_{n+1}. The local errors at each timestep then **blend** together in some obscure way, producing the *global truncation error*, which is also known as the *accumulated truncation* error, or the *discretisation error*. It is usually difficult to **analyse** the discretisation error.

In addition to the truncation errors, **roundoff** errors occur at each computation — because of the *finite* arithmetic precision of computers/calculators. These are sometimes referred to as *computational errors* or *stability errors*. A necessary attribute of any numerical method is that of **convergence** to the exact solution — as the timestep or mesh is refined, the approximate solutions so obtained must *converge* to the exact solution.

Having obtained an approximate solution, it would be useful to know the discretisation error at time t_n: $e_n = y(t_n)-y_n$ (= *true* solution - *approximate* solution). This is difficult to calculate — we **don't** know y(t). The *relative* discretisation error, $e_n/y(t_n)$, is usually of more *importance* — since if the true solution grows, then perhaps **larger** errors may be tolerated; whereas if the true solution diminishes, then the error must do the same — or the true solution will be *swamped* by errors, and the computed solution will be **meaningless**.

A related (but different) concept is that of *stability*. A stable method is one that depends continuously on the initial data. If a small error or perturbation is introduced into a computation, then it should lead to a **correspondingly** small change in the subsequent calculations.

Local Truncation Errors

The local truncation error represents the *local* error made when approximating a DE by a difference equation. Consider the IVP $y' = f(t,y)$, with $y(t_0) = y_0$, and arbitrary *time-stepping* scheme denoted schematically by $\psi(\Delta t, f, y_{n+1}, y_n, ..., y_0) = 0$. Now $y_{n+1} = y_n+\Delta t f(t_n,y_n)$, or $y_{n+1}-y_n-\Delta t f(t_n,y_n) = 0$. The *local truncation* at time t_n is obtained by (1) **substituting** in the *exact* solution into the difference scheme; (2) using *Taylor* series expansions; and (3) dividing by Δt.

If the **first** term in the local truncation error *depends* on $(\Delta t)^p$ (for some integer p), then the *time-stepping scheme* is said to be of order p. We denote the l.t.e. by τ_{n+1}, so that $\tau_{n+1} = {}^1/_{\Delta t}\psi(\Delta t, f, y(t_{n+1}), y(t_n),..., y(t_0))$. If the l.t.e. is of *order* Δt^p, we write $\tau_{n+1} = O(\Delta t^p)$ as $\Delta t \to 0$.

Consistency. A *numerical scheme* is said to be consistent with the differential equation if the local truncation error tends to **zero** as $\Delta t \to 0$. *Example*: Find the local truncation error for the Euler method, using the DE $y' = f(t,y)$. The **Euler** method is given by $y_{n+1} = y_n+\Delta t f(t_n,y_n)$, which may be *rewritten* as ${}^1/_{\Delta t}(y_{n+1}-y_n)-f(t_n,y_n) = 0$. **Substituting** in the *exact* solution gives the truncation error, $\tau_{n+1} = -{}^1/_{\Delta t}(y(t_{n+1}) - y(t_n)) + f(t_n,y(t_n))$.

Now *expand* $y(t_{n+1}) = y(t_n+\Delta t)$ as a *Taylor* series: $y(t_{n+1}) = y(t_n) + \Delta t y'(t_n) + {}^{\Delta t^2}/_{2!}y''(\zeta_n)$, where $t_n \le \zeta_n \le t_{n+1}$. Now *substitute* into the expression for τ_{n+1}, which gives $\tau_{n+1} = -{}^1/_{\Delta t}[\Delta t y'(t_n)+O(\Delta t^2)] + f(t_n,y(t_n)) = -y'(t_n) + O(\Delta t) + f(t_n,y(t_n))$. Therefore, we conclude that $\tau_{n+1} = O(\Delta t)$.

12th October 2000

Tutorial

Q: Write down the *difference* equation obtained by using Euler's method to approximate $y' = ({}^y/_t)^2+({}^y/_t)$. A: *Euler's* method is $y_{n+1} = y_n+\Delta t f(t_n,y_n)$. Hence $y_{n+1} = y_n+\Delta t((y_n/t_n)^2+(y_n/t_n))$. Q: Use the *Modified* Euler (trapezoidal) method to obtain an **approximate** solution to $y' = -y$, where $t \geq 0$, and $y(0) = 1$. Investigate *timesteps* of $\Delta t = 0.1$, 1 and 2. Compare your *approximate* results with the exact solution.

A: $y_{n+1} = y_n + \Delta t{}^1/_2[f(t_n,y_n) + f(t_{n+1},y_{n+1})]$ is the *Modified* Euler method. In this question, $y_{n+1} = y_n+\Delta t{}^1/_2(-y_n-y_{n+1})$; $y_{n+1} = y_n(1-{}^{\Delta t}/_2)-{}^{\Delta t}/_2y_{n+1}$; $y_{n+1}(1+{}^{\Delta t}/_2) = y_n(1-{}^{\Delta t}/_2)$; $y_{n+1} = y_n[(1-{}^{\Delta t}/_2)/(1+{}^{\Delta t}/_2)]$. At $\Delta t = 0.1$, $y_1 = y_0[{}^{(1-0.05)}/_{(1+0.05)}] = {}^{0.95}/_{1.05} = 0.90476$; $y_2 = 0.90476\times0.90476 = 0.81859$; $y_3 = 0.74063$; $y_4 = 0.67009$; and $y_5 = 0.60627$.

At $\Delta t = 1$, $y_0 = 1$; $y_1 = 1{}^{(1/2)}/_{(3/2)} = {}^1/_3$; $y_2 = {}^1/_3\times{}^1/_3 = {}^1/_9$; $y_3 = 0.03703$; $y_4 = 0.012345$; and $y_5 = 4.115\times10^{-3}$. At $\Delta t = 2$, $y_0 = 1$; $y_1 = 1(0) = 0$; and $y_2 = y_3 = y_4 = y_5 = 0$. The *exact* solution is $y = e^{-t}$. $t = 0.1$ *gives* $y = 0.9048$; $t = 0.2$ gives $y = 0.8187$; $t = 0.3$ gives $y = 0.7408$; $t = 0.4$ gives $y = 0.6703$; and $t = 0.5$ gives $y = 0.6065$.

Q: Find the *local truncation error* for the **Modified** Euler method. A: The *Modified* Euler method is $y_{n+1} = y_n+\Delta t{}^1/_2[f(t_n,y_n) + f(t_{n+1},y_{n+1})]$, which may be *rewritten* as ${}^1/_{\Delta t}(y_{n+1}-y_n) - ½(f(t_n,y_n)+f(t_{n+1},y_{n+1})) = 0$ (---(1)). The above *constitutes* our **difference** scheme, with the IVP $y' = f(t,y)$, with $f(t_0) = y_0$.

If we substitute the *exact solution* into (1), the truncation error is given by $\tau_{n+1} = {}^1/_{\Delta t}(y(t_{n+1})-y(t_n)) - ½(f(t_n,y(t_n)) + f(t_{n+1},y(t_{n+1})))$. Now *expand* $y(t_{n+1}) = y(t_n+\Delta t)$ as a *Taylor* series: $y(t_{n+1}) = y(t_n) + \Delta t y'(t_n) + {}^{\Delta t^2}/_{2!}y''(t_n) + {}^{\Delta t^3}/_{3!}y'''(\zeta_n)$, where $t_n \le \zeta_n \le t_{n+1}$. Now *substitute* into the *expression* for τ_{n+1}, giving $\tau_{n+1} = {}^1/_{\Delta t}(y(t_n) + \Delta t y'(t_n) + {}^{\Delta t^2}/_2y''(t_n) + O(\Delta t^3) - y(t_n)) - ½(f(t_n,y(t_n)) + f(t_{n+1},y(t_{n+1}))) = y'(t_n) + {}^{\Delta t}/_2y''(t_n) + O(\Delta t^2) - ½(f(t_n,y(t_n)) + f(t_{n+1},y(t_{n+1})))$.

Now *we know that* $y' = f(t,y)$, or that $y'(t_n) = f(t_n,y(t_n))$, so that $y'(t_{n+1}) = f(t_{n+1}, y(t_{n+1}))$. Therefore, in τ_{n+1}, we **have** $\tau_{n+1} = y'(t_n) + {}^{\Delta t}/_2y''(t_n) + O(\Delta t^2) - ½(y'(t_n)+y'(t_{n+1}))$. Now *expand* $y'(t_{n+1})$ as a **Taylor series**: $y'(t_{n+1}) = y'(t_n+\Delta t) = y'(t_n) + \Delta t y''(t_n) + {}^{\Delta t^2}/_{2!}y'''(\zeta_n)$, where again, $t_n \le \zeta_n \le t_{n+1}$. So we *have* $\tau_{n+1} = y'(t_n) + {}^{\Delta t}/_2y''(t_n) + O(\Delta t^2) - ½(y'(t_n)+y'(t_n)+\Delta t y''(t_n)+O(\Delta t^2)) = y'(t_n) + {}^{\Delta t}/_2y''(t_n) + O(\Delta t^2) - ½(2y'(t_n)+\Delta t y''(t_n)+O(\Delta t^2)) = y'(t_n) + {}^{\Delta t}/_2y''(t_n) + O(\Delta t^2) - y'(t_n) - {}^{\Delta t}/_2y''(t_n) - {}^{O(\Delta t^2)}/_2$; $\tau_{n+1} = {}^{O(\Delta t^2)}/_2 = O(\Delta t^2)$. So the *local truncation error* is of order **2**.

Taylor Series Methods

Recall that the *simple Euler method* uses a Taylor series truncated after the **first 2 terms.** The higher order Taylor series methods retain *more* terms, up to degree k, say. Now $y(t_{n+1}) = y(t_n+\Delta t)$, where $y(t_n+\Delta t) = y(t_n) + \Delta t y'(t_n) + {}^{\Delta t^2}/_{2!}y''(t_n) + {}^{\Delta t^3}/_{3!}y'''(t_n) + \ldots + (\Delta t^k/k!)y^{(k)}(t_n) + (\Delta t^{k+1}/(k+1)!)y^{(k+1)}(\zeta_n)$, where $t_n \le \zeta_n \le t_{n+1}$. To use this, we need *expressions* for **higher** derivatives of y(t).

But the *original* ODE gives that $y'(t) = f(t,y)$, so **differentiating** this w.r.t. t gives $y''(t) = f'(t,y)$; $y'''(t) = f''(t,y)$; ...; $y^{(k)}(t) = f^{(k-1)}(t,y)$. We know the *functional* form of f(t,y), so in **principle**, it can be differentiated — and so the derivatives can be substituted into the Taylor series expansion, giving $y(t_{n+1}) = y(t_n) + \Delta t f(t_n,y(t_n)) + {}^{\Delta t^2}/_{2!}f'(t_n,y(t_n)) + \ldots + (\Delta t^k/k!)f^{(k-1)}(t_n,y(t_n)) + (\Delta t^{k+1}/(k+1)!)f^{(k)}(\zeta_n, y(\zeta_n))$.

Using y_n to *approximate* the solution for $y(t_n)$, the corresponding *difference method* gives the Taylor series **method** of order k: $y_{n+1} = y_n + \Delta t f(t_n,y_n) + {}^{\Delta t^2}/_{2!}f'(t_n,y_n) + \ldots + (\Delta t^k/k!)f^{(k-1)}(t_n,y_n)$, with the *local truncation error* given by the expression $(\Delta t^{k+1}/(k+1)!)f^{(k)}(\zeta_n,y(\zeta_n))$. *Notes*: (1) The advantage of the **Taylor series methods** is the presence of high order accuracy.

(2) The *disadvantage* of the **Taylor series methods** is the requirement for the evaluation and the computation of the derivatives of f(t,y). In practice, this can be very *complicated* and *time consuming*. Consequently, they are *seldom* used in practice. (3) In general, the calculation of ${}^{df}/_{dt}$ requires the use of the **chain** rule, ${}^{df}/_{dt} = {}^{\partial f}/_{\partial t}+{}^{\partial f}/_{\partial y}{}^{dy}/_{dt}$.

Example: $f = f(t,y)$, so that ${}^{df}/_{dt} = ?$; ${}^{d^2f}/_{dt^2} = ?$; and so on. *Now* if $f(t,y) = t^2$, then $f' = 2t$, $f'' = 2$, $f''' = 0$, and $f^{(k)} = 0$ for $k > 2$. If $f(t,y) = y^2$, where $y' = y(t)$, then $f' = 2y{}^{dy}/_{dt} = 2yy' = 2yf$. For *higher* derivatives, must use the **operator** ${}^{d}/_{dt} = {}^{\partial}/_{\partial t}+f{}^{\partial}/_{\partial y}$, for example ${}^{d^2f}/_{dt^2} = ({}^{\partial}/_{\partial t}+f{}^{\partial}/_{\partial y})({}^{\partial}/_{\partial t}+f{}^{\partial}/_{\partial y})f$.

Runge-Kutta Methods

The *RK methods* use the high order accuracy of the **Taylor** series methods, while eliminating the *evaluation* of the derivatives of f(t,y). This is done by using approximations to f(t,y) at *intermediate* values within the time interval $[t_n, t_{n+1}]$, e.g. $t_n+{}^{\Delta t}/_2$, *remembering that* $t_{n+1} = t_n+\Delta t$. So although the *RK schemes* are **single-step**, they are **multi-stage**.

The *derivation* of general RK schemes involve the Taylor series expansions in two variables of f(t,y), and some involved *algebra*, so that we match the Taylor series methods up to the required order of **accuracy**. RK SCHEME OF ORDER 2: The *Modified* Euler method: $y_{n+1} = y_n + {}^{\Delta t}/_2[f(t_n,y_n) + f(t_{n+1},y_n+\Delta t f(t_n,y_n))]$. Note that this may be also *interpreted* as a modified Euler method solved with **one** iteration, using Euler's method as the *predictor*.

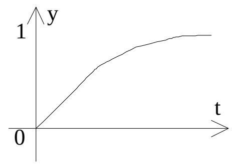
The mid-point method is $y_{n+1} = y_n+\Delta tf(t+{}^{\Delta t}/_2,\ y_n+{}^{\Delta t}/_2 f(t_n,y_n))$. The Heun method is $y_{n+1} = y_n+{}^{\Delta t}/_4[f(t_n,y_n) + 3f(t_n+{}^2/_3\Delta t,\ y_n+{}^2/_3\Delta tf(t_n,y_n)]$. *RK-4:* The most **popular** RK-4 scheme may be written using *auxiliary* variables k_1, k_2, k_3 and k_4 as follows: $k_1 = \Delta tf(t_n,y_n)$; $k_2 = \Delta tf(t_n+{}^{\Delta t}/_2,\ y_n+{}^{k_1}/_2)$; $k_3 = \Delta tf(t_n+{}^{\Delta t}/_2,\ y_n+{}^{k_2}/_2)$; and $k_4 = \Delta tf(t_{n+1},\ y_n+k_3)$, *remembering* that $t_{n+1} = t_n+\Delta t$. **Further,** $y_{n+1} = y_n+{}^1/_6(k_1+2k_2+2k_3+k_4)$, which has an *accuracy of order* 4. Note that the k_i's are introduced to **eliminate** the successive *nesting* of the function $f(t,y)$.

18th October 2000

Tutorial

Q: *Consider* the IVP $y' = -y+1$, with $0 \le t \le 1$, and $y(0) = 0$, which has *exact* solution $y = 1-e^{-t}$. (a) **Sketch** the exact solution; and (b) compare the values (at the *mesh points* 0.1, 0.2, 0.3, 0.4 and 0.5) for the exact solution and the **approximate** solution obtained by (i) the Runge-Kutta *fourth order* method, with $\Delta t = 0.1$; (ii) the *Euler Method*, with $\Delta t = 0.1$, and $\Delta t = 0.025$; and (iii) the *Modified Euler method*, with $\Delta t = 0.1$, and $\Delta t = 0.05$.

A: (a) **The sketch** is as shown on the right. (b) We must *compare* to the exact solutions as follows: 0.1 = 0.095163; 0.2 = 0.181269; 0.3 = 0.259182; 0.4 = 0.32968; and 0.5 = 0.393469. (i) For **RK-4**, we have $f(t_n,y_n) = -y_n+1$, and the *difference scheme is as follows:* $k_1 = (0.1)(-y_n+1)$; $k_2 = (0.1)(-(y_n+{}^{k_1}/_2)+1)$; $k_3 = (0.1)(-(y_n+{}^{k_2}/_2)+1)$; $k_4 = (0.1)(-(y_n+k_3)+1)$; and $y_{n+1} = y_n + {}^1/_6(k_1+2k_2+2k_3+k_4)$.



Now we *know that* $y_0 = 0$. Let us obtain $y_1 = y_0+{}^1/_6(k_1+2k_2+2k_3+k_4)$: $k_1 = (0.1)(-y_0+1) = (0.1)(0+1) = 0.1$; $k_2 = (0.1)(-(y_0+{}^{k_1}/_2)+1) = (0.1)(-(0+{}^{0.1}/_2)+1) = 0.095$; $k_3 = (0.1)(-(y_0+{}^{0.095}/_2)+1) = 0.09525$; and $k_4 = (0.1)(-(y_0+0.09525)+1) = 0.090475$. Therefore, $y_1 = 0 + {}^1/_6(0.1 + (2\times0.095) + (2\times0.09525) + 0.090475) = 0.0951625$. *Similarly* for y_2, y_3, y_4 and y_5. Comparison: the values *obtained* for RK-4 are very **close** to the actual ones.

(ii) **Euler** method. The *difference scheme* is $y_{n+1} = y_n+\Delta tf(t_n,y_n)$; $y_{n+1} = y_n+\Delta t(-y_n+1)$. When $\Delta t = 0.1$, $y_0 = 0$; $y_1 = y_0+0.1(-y_0+1) = 0+0.1(-0+1) = 0.1$; $y_2 = 0.1+0.1(-0.1+1) = 0.19$; $y_3 = 0.19+0.1(-0.19+1) = 0.271$; $y_4 = 0.3439$; and $y_5 = 0.40951$. The *values are fairly close*. When $\Delta t = 0.025$, $y_{n+1} = y_n+0.025(-y_n+1)$, and we have $y_0 = 0$; $y_1 = 0.025$; $y_2 = 0.049375$; $y_3 = 0.073140625$; and $y_4 = 0.09631210937$. ***This*** is what we compare to the *mesh point* 0.1. Going on, $y_5 = ...$ **Conclusion**: the values for $\Delta t = 0.025$ are more *accurate* than the values for $\Delta t = 0.1$.

(iii) **Modified Euler** method. The *difference scheme* is $y_{n+1} = y_n+\Delta t½[f(t_n,y_n)+f(t_{n+1},y_{n+1})] = y_n+\Delta t{}^1/_2[(-y_n+1)+(-y_{n+1}+1)] = y_n + \Delta t - {}^{\Delta t}/_2 y_n - {}^{\Delta t}/_2 y_{n+1}$; $y_{n+1}(1+{}^{\Delta t}/_2) = y_n(1-{}^{\Delta t}/_2) + \Delta t$; $y_{n+1} = [y_n(1-{}^{\Delta t}/_2)+\Delta t]/[1+{}^{\Delta t}/_2]$. At $\Delta t = 0.1$, $y_{n+1} = (y_n(0.95)+0.1)/1.05$. So $y_0 = 0$; $y_1 = {}^{0.1}/_{1.05} = 0.095238$; $y_2 = 0.1814$; $y_3 = 0.2594$; $y_4 = 0.3299$; and $y_5 = 0.3937$. The values compare *well*.

When $\Delta t = 0.05$, $y_{n+1} = (y_n(0.975)+0.05)/1.025$. Therefore, $y_0 = 0$; $y_1 = {}^{0.05}/_{1.025} = 0.04878$; $y_2 = 0.09518$ *(compare to mesh point 0.1);* $y_3 = 0.1393$; $y_4 = 0.1813$ *(compare to mesh point 0.2); and* $y_5 = ...$ **Conclusion**: As with the *Euler* method, smaller Δt values mean better results. And when $\Delta t = 0.1$, the *modified Euler method* returns the **better** results.

19th October 2000

Partial Differential Equations

Some standard pde's. (1) $\partial^2\phi/\partial t^2$ - $c^2\nabla^2\phi$, the *wave* equation. Here, ϕ is a **variable**, such as the *displacement of a string, or an electric field,* which depends on **time** and **position** in space: in 1-D, x; in 2-D, x and y; and in 3-D, x, y and z. In 3-D, $\phi = \phi(x,y,z,t)$. In 1-D, $\nabla^2\phi = \partial^2\phi/\partial x^2$. In 2-D, $\nabla^2\phi = \partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2$. In 3-D, $\nabla^2\phi = \partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2 + \partial^2\phi/\partial z^2$.

(2) $\nabla^2\phi = \phi(x,y,z)$, *Poisson's* equation. (3) $\partial\phi/\partial t = \kappa\nabla^2\phi$, the *diffusion* equation. (4) $i\hbar\,\partial\phi/\partial t = -\hbar^2/2m\,\nabla^2\phi + v\phi$, *Schrödinger's* equation. (5) $\partial V/\partial t + \frac{1}{2}\sigma^2 S^2 \partial^2 V/\partial S^2 + rS\,\partial V/\partial S - rV = 0$, *Black-Scholes' equation in financial mathematics of derivatives* — options and futures. For reference, V = option value; S = current value; t = time; σ = volatility; and r = interest rate.

Classification of PDE's

The general *second order pde* may be written in **two** variables as $a\,\partial^2\phi/\partial x^2 + b\,\partial^2\phi/\partial x\partial y + c\,\partial^2\phi/\partial y^2 + d\,\partial\phi/\partial x + e\,\partial\phi/\partial y + f\phi + g = 0$, where *a, b, c, d, e, f and g* are in general **functions** of x, y, ϕ, $\partial\phi/\partial x$ and $\partial\phi/\partial y$, but **not** functions of $\partial^2\phi/\partial x^2$, $\partial^2\phi/\partial x\partial y$ and $\partial^2\phi/\partial y^2$. The pde is *classified* as **elliptic** when $b^2-4ac < 0$; as **parabolic** when $b^2-4ac = 0$; and as **hyperbolic** when $b^2-4ac > 0$.

Examples: (1) $\partial^2\phi/\partial t^2 - c^2\partial^2\phi/\partial x^2 = 0$. *Generally*, we have $a\,\partial^2\phi/\partial t^2 + b\,\partial^2\phi/\partial t\partial x + c\,\partial^2\phi/\partial x^2$. Here, we have *a = 1, b = 0, and c = -c²*. So b^2-4ac is $0+4c^2$. This is **always** > 0, so the wave equation is <u>hyperbolic</u>. (2) $\partial^2\phi/\partial x^2 + \partial^2\phi/\partial y^2 = f$. Here, *a = 1, b = 0, and c = 1,* so that $b^2-4ac = -4$. This is **always** -ve, so that Poisson's equation is <u>elliptic</u>. (3) Set $\kappa = 1$, so that $\partial^2\phi/\partial x^2 - \partial\phi/\partial t = 0$. Here, *a = 1, and b = c = 0,* so that $b^2-4ac = 0$, and thus the *diffusion* equation is <u>parabolic</u>.

Boundary Conditions

There are *four types of boundary conditions* that are employed: (1) <u>Dirichlet</u>: ϕ is given on the *boundary*. (2) <u>Neumann</u>: **The Gradient** of ϕ is given at the *boundary*. (3) <u>Cauchy</u>: ϕ **and** its slope are given at a *boundary*. (4) <u>Robbins</u> (<u>mixed</u>): a **linear combination** of ϕ and its slope is given at the *boundary*. The table shown *summarises* the **types** of boundary conditions appropriate for the three *classes* of PDE.

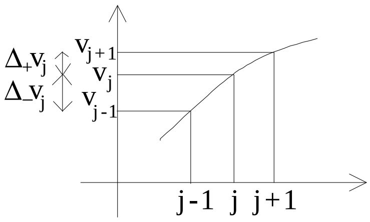
Equation Class	**Boundary**	**Boundary Condition**
Hyperbolic	open	Cauchy
Parabolic	open	Dirichlet or Neumann
Elliptic	closed	Dirichlet or Neumann

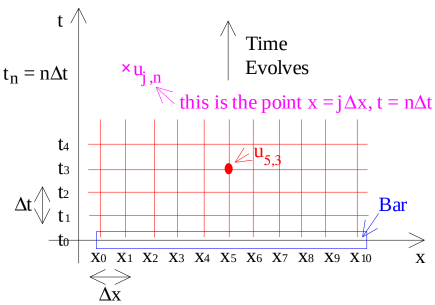
24th October 2000

Discretisation

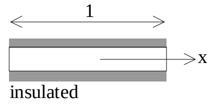
<u>Finite Difference Method:</u> (1) Construct a *mesh;* (2) Approximate the *differential equation* with a *difference equation;* (3) **Solve** the difference method; (4) **Justify** the method (with convergence and stability arguments). **<u>Notation</u>**: We shall use U for the *exact solution of the pde;* u for the *exact solution of the difference equation;* and $\bar{u}$ for the *approximate solution of the difference equation*. What we **want** is U, but what we actually **get** is $\bar{u}$.

Difference operators. *Forward Difference*: $\Delta_+ v_j = v_{j+1} - v_j$. *Backward Difference*: $\Delta_- v_j = v_j - v_{j-1}$. *Central Difference*: $\delta v_j = v_{j+½} - v_{j-½}$. *Double Interval Central Difference*: $\Delta_0 v_j = ½(\Delta_+ + \Delta_-)v_j = ½(v_{j+1} - v_{j-1})$. *Second Order Central Difference*. $\delta^2 v_j = v_{j+1} - 2v_j + v_{j-1}$. The *difference operators* may be used to **approximate** differential operators, e.g. ${}^{dv}/_{dt} = {}^1/_{\Delta t}(v_{j+1} - v_j) + O(\Delta t)$; ${}^{dv}/_{dt} = {}^1/_{\Delta t}(v_j - v_{j-1}) + O(\Delta t)$; and ${}^{d^2v}/_{dt^2} = {}^1/_{\Delta t^2}(v_{j+1} - 2v_j + v_{j-1}) + O(\Delta t^2)$. **Proofs** of the above formulae use *Taylor Series Approximations*. **<u>Difference Notation for 2 independent variables:</u>** $v = v(x,t)$; ${}^{\partial v}/_{\partial x} = [(v_{i+1,j} - v_{i,j})/(\Delta x)] + O(\Delta x)$; ${}^{\partial v}/_{\partial t} = [(v_{i,j+1} - v_{i,j})/(\Delta t)] + O(\Delta t)$; and ${}^{\partial^2 v}/_{\partial x^2} = [(v_{i+1,j} - 2v_{i,j} + v_{i-1,j})/(\Delta x^2)] + O(\Delta x^2)$.

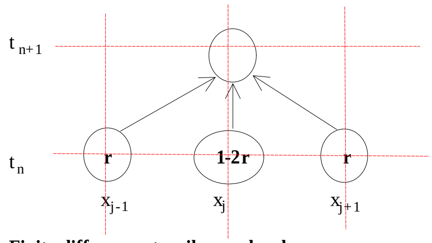




<u>Parabolic Equations</u>. We shall consider the *1-D diffusion/heat equation*, ${}^{\partial U}/_{\partial t} = {}^{\partial^2 U}/_{\partial x^2}$, *where* $U = U(x,t)$. <u>Physically,</u> this models the heat flow along a **bar**, as shown on the right. $U(x,t)$ is the *temperature* at **position** x and at **time** t. The finite difference mesh consists of the *graph* as shown on the left. **<u>Notation</u>**: $U(x,t)$ (at $x_j = j\Delta x$, and at $t_n = n\Delta t$) will be *denoted* by $U_{j,n}$. The **finite difference approximation** to $U_{j,n}$ will be *denoted* by $u_{j,n}$.



<u>Simple Explicit Method</u>. This uses *forward difference* for ${}^{\partial U}/_{\partial t}$, and *central difference* for ${}^{\partial^2 U}/_{\partial x^2}$, to **give** $(u_{j,n+1} - u_{j,n})/(\Delta t) = (u_{j+1,n} - 2u_{j,n} + u_{j-1,n})/(\Delta x^2)$. This may be *rewritten* as $u_{j,n+1} = ru_{j+1,n} + (1-2r)u_{j,n} + ru_{j-1,n}$, where $r = {}^{\Delta t}/_{\Delta x^2}$. This allows us to *march forward* in time from t_n to t_{n+1}, and the process can be **pictured** as shown on the right.



<u>Finite difference stencil or molecule</u>

25th October 2000

<u>Example:</u> Derive *four* steps in the solution of ${}^{\partial U}/_{\partial t} = {}^{\partial^2 U}/_{\partial x^2}$ (in $0 \le x \le 1$) using the *simple explicit scheme;* the **initial** condition $U(x,0) = 1$; and the **boundary** conditions $U(0,t) = 0$ for $t > 0$, and $U(1,t) = 0$ for $t > 0$. Consider *three* cases, and do four steps: (a) $\Delta x = ¼$, $\Delta t = {}^1/_{32}$ ($r = ½$); (b) $\Delta x = {}^1/_8$, $\Delta t = {}^1/_{128}$ ($r = ½$); and (c) $\Delta x = ¼$, $\Delta t = {}^1/_{16}$ ($r = 1$).

n=4	0	1/4	1/4	1/4	0
n=3	0	1/4	1/2	1/4	0
n=2	0	1/2	1/2	1/2	0
n=1	0	1/2	1	1/2	0
n=0	0	1	1	1	0
(a)	j=0	j=1	j=2	j=3	j=4

n=4	0	3/8	3/4	7/8	1	7/8	3/4	3/8	0
n=3	0	1/2	3/4	1	1	1	3/4	1/2	0
n=2	0	1/2	1	1	1	1	1	1/2	0
n=1	0	1	1	1	1	1	1	1	0
(n=0)	1	1	1	1	1	1	1	1	1
(b)	j=0	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8

n=4	0	-2	3	-2	0
n=3	0	1	-1	1	0
n=2	0	0	1	0	0
n=1	0	1	1	1	0
n=0	1	1	1	1	1
(c)	j=0	j=1	j=2	j=3	j=4

A: For (a) and (b,) we have **calculations** of the *type* $u_{1,1} = ½u_{2,0} + ½u_{0,0}$, etc., *because* $r = ½$. For (c), we *have* $r = 1 \Rightarrow u_{j,n+1} = u_{j+1,n} - u_{j,n} + u_{j-1,n}$ in $u_{j,n+1} = ru_{j+1,n} + (1-2r)u_{j,n} + ru_{j-1,n}$. The *latter* scheme (with $r = 1$) is unstable, but **those** with $r = ½$ *were* stable, and gave the correct type of solution. Note that the <u>simple explicit scheme</u> is (1) *single / one step*; and (2) *explicit* (i.e. a calculation of a <u>single</u> value at the forward time step involves only values known at the *previous* time step(s)).

26th October 2000

Local Truncation Errors

$\partial U/\partial t = \partial^2 U/\partial x^2$; $u_{j,n+1}-u_{j,n}/\Delta t = u_{j+1,n}-2u_{j,n}+u_{j+1,n}/\Delta x^2$. This represents the *discrepancy* between the PDE and the **difference** equation. To find it, (1) *Replace* u by U in the difference equation; (2) Expand in terms of Taylor series; and (3) **Cancel** off the terms that satisfy the original PDE. What remains is the *local truncation error* at point x_i and t_n, which we denote by $T_{j,n}$.

Q: Find the *local truncation error* of the simple explicit scheme ($u_{j,n+1}-u_{j,n}/\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$). A: First *expand U* as a Taylor series in t. **Note** that $U_{j,n+1} = U(j\Delta x, (n+1)\Delta t) = U(x_j, t_{n+1})$, and that $U_{j,n} = U(j\Delta x, n\Delta t) = U(x_j,t_n)$, and so the *Taylor series* in t about $t_n = n\Delta t$ is given by $U_{j,n+1} = U_{j,n} + (\partial U/\partial t)_{j,n}\Delta t + (\partial^2 U/\partial t^2)_{j,\tau n}(\Delta t^2/2)$, *where* $t_n \leq \tau_n \leq t_{n+1}$.

Therefore, $U_{j,n+1}-U_{j,n}/\Delta t = (\partial U/\partial t)_{j,n} + O(\Delta t)$. Now for the *Taylor series in x*: we express $U_{j+1,n}$ and $U_{j-1,n}$ in terms of their *values* at x_j and t_n: $U_{j+1,n} = U_{j,n}+(\partial U/\partial x)_{j,n}\Delta x + (\partial^2 U/\partial x^2)_{j,n}{}^{\Delta x^2}/2 + (\partial^3 U/\partial x^3)_{j,n}{}^{\Delta x^3}/6 + O(\Delta x^4)$; and $U_{j-1,n} = U_{j,n}-(\partial U/\partial x)_{j,n}\Delta x + (\partial^2 U/\partial x^2)_{j,n}{}^{\Delta x^2}/2 - (\partial^3 U/\partial x^3)_{j,n}{}^{\Delta x^3}/6 + O(\Delta x^4)$; so that $U_{j+1,n}-2U_{j,n}+U_{j-1,n}/\Delta x^2 = [(\partial^2 U/\partial x^2)_{j,n}\Delta x^2+O(\Delta x^4)]/\Delta x^2 = (\partial^2 U/\partial x^2)_{j,n} + O(\Delta x^2)$. Therefore, the *difference equation* is $[U_{j,n+1}-U_{j,n}/\Delta t] - [U_{j+1,n}-2U_{j,n}+U_{j-1,n}/\Delta x^2] = O(\Delta t)+O(\Delta x^2)$. This is the *truncation error* T_{ij}. Note that this scheme is **first** order accurate in **time**, and **second** order accurate in **space**.

Consistency, Total Error, Convergence, Stability

Consistency. We say that a *finite difference scheme* is consistent with the PDE if the local truncation errors tend to **zero** (for each fixed point in the domain) as $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$ *independently*. Example: $T_{i,j} = {}^{\Delta t}/_{\Delta x} + \Delta x^2$. Here, *ensure* that $\Delta t = O(\Delta x^2)$. **Total Error**. The total error at point (x_j, t_n) is *given* by $U_{j,n}-\bar{u}_{j,n} = (U_{j,n}-u_{j,n}) + (u_{j,n}-\bar{u}_{j,n})$ *(= global discretisation error + roundoff error / computational stability).*

Convergence. If, at each point (x_i, t_n), the discretisation error ($e_{j,n} = U_{j,n}-u_{j,n}$) tends to **zero** as $\Delta x, \Delta t \rightarrow 0$, we say that the *finite difference scheme* is convergent. **Stability**. Suppose that in a computation *involving* a difference scheme, an error $\in_0$ is *introduced* at time t_0, and suppose that **no** further errors occur.

Let $\in_n$ denote the error at *time* t_n which has resulted from the **single** error at t_0. Then the scheme is **stable** if $|\in_n|$ remains *bounded* as $n \rightarrow \infty$. Note: stability does **not** in general imply that the scheme has *small truncation error*. There are **two** main methods for testing whether a scheme is stable: (1) **The Fourier Method** (*the von Neumann* method), in which a *finite* Fourier series representation of the error is used; and (2) **The Matrix Method**, in which the scheme is expressed in *matrix form*, and the **eigenvalues** of the associated matrix are examined.

31st October 2000

Stability

Which method do we use? The Fourier method is *easier*, but ignores the effects of boundary conditions. The matrix method is more *rigorous*, but is more difficult to calculate.

The Fourier Method (The von Neumann Method)

Let us **write** $\bar{u}_{j,n} = u_{j,n} + \epsilon_{j,n}$ in a *difference scheme*. If the scheme is **linear**, then the error will satisfy the same *equation* as $u_{j,n}$. We now express the error as a *finite Fourier series*: $\epsilon_{j,n} = \Sigma_k \hat{\epsilon}_{k,n}(t)e^{ikx_j}$ (x_j in the exponential — not xj), and, because the scheme is *linear*, we only need consider the effect of a **single** term, $\epsilon_{j,n} = \hat{\epsilon}_n e^{ikx_j}$. We need to obtain the *condition* so that $|\hat{\epsilon}_n| \leq 1$ — so that the error will **not** grow.

<u>Example:</u> Investigate the stability of the *simple explicit scheme*, $u_{j,n+1} = ru_{j+1,n} + (1\text{-}2r)u_{j,n} + ru_{j-1,n}$, where $r = {}^{\Delta t}/_{\Delta x^2}$. *Substituting* $\epsilon_{j,n} = \hat{\epsilon}_n e^{ikx_j}$ into the *difference* scheme, we obtain the following expression: $\hat{\epsilon}_{n+1}e^{ikx_j} = r\hat{\epsilon}_n e^{ikx_{(j+1)}} + (1\text{-}2r)\hat{\epsilon}_n e^{ikx_j} + r\hat{\epsilon}_n e^{ikx_{(j-1)}} = r\hat{\epsilon}_n e^{ik(x_j+\Delta x)} + (1\text{-}2r)\hat{\epsilon}_n e^{ikx_j} + r\hat{\epsilon}_n e^{ik(xj-\Delta x)} = r\hat{\epsilon}_n e^{ikx_j}e^{ik\Delta x} + (1\text{-}2r)\hat{\epsilon}_n e^{ikx_j} + r\hat{\epsilon}_n e^{ikx_j}e^{-ik\Delta x}$. Therefore, *cancelling*, $\hat{\epsilon}_{n+1} = [1\text{-}2r+r(e^{ik\Delta x}+e^{-ik\Delta x})]\hat{\epsilon}_n$.

Now **use** $e^{\pm i\theta} = \cos\theta \pm i\sin\theta$, so that $e^{i\theta}+e^{-i\theta} = 2\cos\theta$. Therefore, $\hat{\epsilon}_{n+1} = [1\text{-}2r+r2\cos(k\Delta x)]\hat{\epsilon}_n = [1\text{-}2r(1\text{-}\cos k\Delta x)]\hat{\epsilon}_n$; or $\hat{\epsilon}_{n+1} = g\hat{\epsilon}_n$, where $g = 1\text{-}2r(1\text{-}\cos k\Delta x)$ is the **amplification** (or **growth**) factor. For *stability*, we require $|g| \leq 1$, which is known as the <u>von Neumann stability condition</u>.

For $g = 1\text{-}2r(1\text{-}\cos k\Delta x)$, we *require* $|1\text{-}2r(1\text{-}\cos k\Delta x)| \leq 1$. *Now* $-1 \leq \cos\theta \leq 1$, so that $0 < {}^{\Delta t}/_{\Delta x^2} = r$; and so $0 \leq 1\text{-}\cos\theta \leq 2$. We require $|\text{-}4r| \leq 1$, which means that $-1 \leq 1\text{-}4r \leq 1$, where the red inequality is *redundant*. The **left** inequality gives $4r \leq 2$, or $r \leq ½$. Note that for a *given spatial mesh* size Δx, the stability condition gives a **restriction** on the timestep of Δt: $\Delta t \leq {}^{\Delta x^2}/_2$.

This is *extremely restrictive* — the timestep becomes **extremely** small for fine spatial meshes. We say that the *simple explicit scheme* is <u>*conditionally stable*</u>. Perhaps an implicit scheme would give **unconditional** stability?

<u>Exercises 1</u>

Q: Investigate whether the *following functions* satisfy a Lipschitz condition in the variable y, and, if so, find a **Lipschitz** constant. (a) $f(t,y) = t|y|$ for $\{(t,y) \mid 1\leq t\leq 2, -3\leq y\leq 4\}$; (b) $f(t,y) = -y+t+1$ for $\{(t,y) \mid 0\leq t\leq 1, -\infty\leq y\leq\infty\}$; and (c) $f(t,y) = 1+t\sin(ty)$ for $\{(t,y) \mid 0\leq t\leq 2, -\infty\leq y\leq\infty\}$.

A: We want a *constant* $L > 0$, with $|f(t,y_1)-f(t,y_2)| < L|y_1-y_2|$; or we use the following <u>*theorem*</u>: if $f(t,y)$ satisfies $|{}^{\partial f}/_{\partial y}(t,y)| < L$ for all (t,y) on the set, then *f satisfies a Lipschitz condition* on the set in the variable y, with Lipschitz **constant** L. (a) $f(t,y) = t|y|$; ${}^{\partial f}/_{\partial y}(t,y) = t|1| = t$. So if $t < L$, then we *choose* $L > t$. So any $L > 2$ will do.

Note that $f(t,y)$ is *not* differentiable at $y = 0$. We could <u>also</u> use $|t|y_2|-t|y_1|| = |t||y_2|-|y_1|| \leq |t||y_2-y_1| \leq 2|y_2-y_1|$. The **first** inequality may be *rewritten as* $(|a|-|b|)^2 = |a|^2 - 2|a||b|+|b|^2 = a^2-2|a||b|+b^2 \leq a^2-2ab+b^2 =$ (*since* $|a||b| \geq ab$) $= (a-b)^2$. Hence $-|a||b| \leq -ab$. Taking the **+ve square roots** gives $||a|-|b|| \leq |a-b|$.

(b) $f(t,y) = -y+t+1$; ${}^{\partial f}/_{\partial y}(t,y) = -1$; $|-1| = 1$ — so *any* $L > 1$ will do, *say* $L = 1$. (c) $f(t,y) = 1+t\sin(ty)$; $|{}^{\partial f}/_{\partial y}(t,y)| = |t^2\cos(ty)| = t^2|\cos(ty)| \leq 4$ (since $t \in [0,2] \Rightarrow t^2 \leq 4$, and $|\cos\theta| \leq 1$ for any θ). So the *Lipschitz condition* is satisfied, with **constant** $L = 4$.

Q: Write down the *difference equations* obtained by using Euler's method to **approximate** the solution for each of the following *initial value problems:* (a) $y' = \sin t+e^{-t}$; (b) $y' = {}^2/_t y+t^2e^t$. A: (a) $y_{n+1} = y_n+\Delta t(\sin t_n+e^{-t_n})$; (b) $y_{n+1} = y_n+\Delta t((2/t_n)y_n + t_n^2e^{t_n})$.

Q: Show that the *Adams-Bashforth two-step method* is **consistent**, and has a **local truncation error** of $O(\Delta t^2)$. A: We have $y_{n+2} = y_{n+1} + {}^{\Delta t}/_2(3f(t_{n+1},y_{n+1}) - f(t_n,y_n))$; and we rewrite this as ${}^1/_{\Delta t}(y_{n+2}-y_{n+1}) = ½(3f(t_{n+1},y_{n+1}) - f(t_n,y_n))$. Substitute in the *exact solution*, giving ${}^1/_{\Delta t}(y(t_{n+2})-y(t_{n+1})) - ½(3f(t_{n+1},y(t_{n+1})) - f(t_n,y(t_n)) = 0$, or ${}^1/_{\Delta t}(y(t_{n+1})-y(t_n)) - ½(3f(t_n,y(t_n)) - f(t_{n-1},y(t_{n-1}))) = 0$.

Now $y(t_{n+1}) = y(t_n+\Delta t) = y(t_n) + \Delta ty'(t_n) + {}^{\Delta t^2}/_2y''(t_n) + {}^{\Delta t^3}/_{3!}y'''(t_n) + ...$; so that ${}^1/_{\Delta t}(\Delta ty'(t_n) + {}^{\Delta t^2}/_2y''(t_n) + {}^{\Delta t^3}/_{3!}y'''(t_n)+...) - ½(...) = 0$. *Using* $y' = f(t,y)$, or $y'(t_n) = f(t_n,y(t_n))$, we *have* $y'(t_n) + {}^{\Delta t}/_2y''(t_n) + {}^{\Delta t^2}/_{3!}y'''(t_n) + ... - ½(3y'(t_n)-y'(t_{n-1})) = 0$. Now $y'(t_{n-1}) = y'(t_n) - \Delta ty''(t_n) + {}^{\Delta t^2}/_{2!}y'''(t_n) - ....$; so that ... $- ½(3y'(t_n) - y'(t_n) + \Delta ty''(t_n) - ...) = 0$. In *full*, $y'(t_n) + {}^{\Delta t}/_2y''(t_n) + {}^{\Delta t^2}/_{3!}y'''(t_n) + ... - ½(2y'(t_n) + \Delta ty''(t_n) - {}^{\Delta t^2}/_{2!}y'''(t_n) + ...) = 0$. Therefore, ${}^{\Delta t^2}/_{3!}y'''(t_n) + ... - ½(-{}^{\Delta t^2}/_{2!}y'''(t_n)+...) = 0 = \tau_{n+1}$. Note that τ_{n+1} is the *truncation error*, where $\tau_{n+1} = O(\Delta t^2)$. It **is** consistent, because as $\Delta t\rightarrow 0$, we have $\tau_{n+1}\rightarrow 0$ — no terms are *not dependent* on Δt.

Q: Show that the **theta** method is *consistent;* is of order <u>two</u> for $\theta = ½$; and is otherwise of order <u>one</u>. A: *Theta method*: $y_{n+1} = y_n+\Delta t[(1-\theta)f(t_n,y_n)+\theta f(t_{n+1},y_{n+1})]$, where $0 \leq \theta \leq 1$, or $y_{n+1}-y_n/\Delta t = (1-\theta)f(t_n,y_n) + \theta f(t_{n+1},y_{n+1})$. Substituting in the **exact** solution, ${}^1/_{\Delta t}(y(t_{n+1})-y(t_n)) = (1-\theta)f(t_n,y(t_n))+\theta f(t_{n+1},y(t_{n+1}))$; ${}^1/_{\Delta t}(\Delta ty'(t_n) + {}^{\Delta t^2}/_2y''(t_n) + {}^{\Delta t^3}/_{3!}y'''(t_n)+...) = ...$ And **now** ${}^1/_{\Delta t}(...) = (1-\theta)y'(t_n) + \theta y'(t_{n+1})$.

So ${}^1/_{\Delta t}(\Delta ty'(t_n)+{}^{\Delta t^2}/_2y''(t_n)+{}^{\Delta t^3}/_{3!}y'''(t_n)+...) - (1-\theta)y'(t_n)-\theta y'(t_{n+1}) = 0$; ${}^1/_{\Delta t}(...) - (1-\theta)y'(t_n) - \theta(y'(t_n)+\Delta ty''(t_n)+{}^{\Delta t^2}/_2y'''(t_n)+...) = 0$. *Now* $y'(t_n) + {}^{\Delta t}/_2y''(t_n) + {}^{\Delta t^2}/_{3!}y'''(t_n) + ... + y'(t_n)(\theta-1-\theta) - \theta(\Delta ty''(t_n)+{}^{\Delta t^2}/_2y'''(t_n)+...) = 0$; $y''(t_n)({}^{\Delta t}/_2-\theta\Delta t) + y'''(t_n)({}^{\Delta t^2}/_3-\theta{}^{\Delta t^2}/_2)+... = 0$. When $\theta = ½$, the first term *disappears*, and the l.t.e. is of **order** Δt^2. *Otherwise*, it is of order Δt. **QED**.

Q: In a circuit with *impressed voltage E; resistance R; inductance L; and capacitance C in parallel*, the <u>current</u> I satisfies the differential equation ${}^{dI}/_{dt} = C{}^{d^2E}/_{dt^2}+{}^1/_R{}^{dE}/_{dt}+{}^1/_LE$. Suppose that *C = 0.3 farads; R = 1.4 ohms; L = 1.7 henries;* and the *voltage* is given by $E(t) = e^{-0.06\pi t}\sin(2t-\pi)$. If $I(0) = 0$, find the *current* I for the values $t = 0.1j$, for $j = 0, 1, ..., 100$, using **Euler's** method.

A: **Euler's** method is $y_{n+1} = y_n+\Delta tf(t_n,y_n)$. Here, $y_{n+1} = y_n+0.1(c{}^{d^2E}/_{dt^2}+{}^1/_R{}^{dE}/_{dt}+{}^1/_LE)$; $E'(t) = -0.06\pi E(t)+2e^{-0.06\pi t}\cos(2t-\pi)$; and $E''(t) = -0.06\pi E'(t) - 4E(t)$. Now $y_0 = 0$; $C = 0.3$; $R = 1.4$; $L = 1.7$; $E(0) = 0$; $E'(0) = -2$; and $E''(0) = -0.06\pi(-2)$. Therefore, $y_1 = y_0 + 0.1(0.3(-0.06\pi(-2)) + {}^1/_{1.4}(-2) + {}^1/_{1.7}(0))$; $y_1 = 0+0.1(...) = -0.01315$. Further, $y_2 = y_1+0.1(0.3E''(y_1)+{}^1/_{1.4}E'(y_1)+{}^1/_{1.7}E(y_1)) = ...$; and *so on* for y_3, y_4, etc.

Exercises 3

Q: Classify the *following* **partial** differential equations into (i) *linear, quasi-linear, or non-linear;* (ii) *hyperbolic, parabolic, or elliptic;* and (iii) state what sort of *boundary conditions* would be appropriate in each case. (a) $\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0$; (b) $\frac{\partial^2 U}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} = 0$, where c is a *constant;* (c) $\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$; (d) $c^2\frac{\partial^2 \phi}{\partial x^2} = \frac{\partial^2 \phi}{\partial t^2} + p\frac{\partial \phi}{\partial t} + q\phi$, where c, p and q are *constants*; (e) $\frac{\partial^2 \phi}{\partial x^2} + x\frac{\partial^2 \phi}{\partial y^2} = 0$; and (f) $\frac{\partial^2 \phi}{\partial x^2} + (1-x^2-y^2)\frac{\partial^2 \phi}{\partial y^2} = x^2-y^2$.

A: (a) *Linear*. In general, we have $a\frac{\partial^2 \phi}{\partial x^2}+b\frac{\partial^2 \phi}{\partial x\partial y}+c\frac{\partial^2 \phi}{\partial y^2}+d.... = 0$. Here, $a = 1$, $b = 0$, and $c = 1$; thus $b^2-4ac = -ve$, and so the equation is **elliptic**. Boundary condition: <u>closed; Dirichlet or Neumann</u>. (b) *Linear*. $a = 1$, $c = -\frac{1}{c^2}$, and $b = 0$; so that $b^2-4ac = -4(1)(-\frac{1}{c^2}) = \frac{4}{c^2} > 0$, and thus the equation is **hyperbolic**. Boundary condition: <u>open; Cauchy.</u> (c) *Linear*. $a = 0$, $b = 0$, and $c = 1$; so that $b^2-4ac = 0$, and thus the equation is **parabolic**. Boundary condition: <u>open; Dirichlet or Neumann</u>. (d) *Linear*. $a = c^2$, $b = 0$, and $c = -1$; so that $b^2-4ac = -4(c^2)(-1) = 4c^2 > 0$, and thus the equation is **hyperbolic**. Boundary condition: <u>open; Cauchy</u>. (e) *Linear*. $a = 1$, $b = 0$, $c = x$; so that $b^2-4ac = -4x$. The equation is **elliptic** when $x > 0$; **parabolic** when $x = 0$; and **hyperbolic** when $x < 0$. Boundary conditions as <u>appropriate</u>. (f) *Linear*. $a = 1$, $b = 0$, and $c = (1-x^2-y^2)$; so that $b^2-4ac = x^2+y^2-1$. When $x^2+y^2 > 1$, the equation is **hyperbolic**; $x^2+y^2 = 1 \Rightarrow$ **parabolic**; $x^2+y^2 < 1 \Rightarrow$ **elliptic** (*on, below, or above* a circle of radius 1 in $\mathbf{R}^2$). Boundary conditions as <u>appropriate</u>.

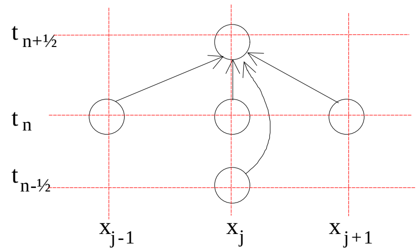
Q: By using *Taylor Series expansions*, prove that (a) $\frac{\partial v}{\partial x}(x_j) = (v_{j+1}-v_j/\Delta x) + O(\Delta x)$; (b) $\frac{\partial u}{\partial x}(x_j) = (u_j-u_{j-1}/\Delta x) + O(\Delta x)$; (c) $\frac{\partial \phi}{\partial x}(x_j) = (\phi_{j+1}-\phi_{j-1}/2\Delta x) + O(\Delta x^2)$; (d) $\frac{\partial^2 v}{\partial x^2}(x_j) = (v_{j+1}-2v_j+v_{j-1}/\Delta x^2) + O(\Delta x^2)$; (e) $\frac{\partial v}{\partial x}(x_j) = (-3v_j+4v_{j+1}-v_{j+2}/2\Delta x) + O(\Delta x^2)$; and (f) $\frac{\partial^2 v}{\partial x^2}(x_j) = (2v_j-5v_{j+1}+4v_{j+2}-v_{j+3}/\Delta x^2) + O(\Delta x^2)$. <u>**Note**</u>: In an exam, use $u'_j = u'(x_j) = \frac{du}{dx}\big|_{x=x_j}$, *etc*.

A: (a) RHS: $v_{j+1} = v_j + (\frac{\partial v}{\partial x})_j\Delta x + (\frac{\partial^2 v}{\partial x^2})\frac{\Delta x^2}{2} + ...$ Therefore, $v_{j+1}-v_j = (\frac{\partial v}{\partial x})_j\Delta x + (\frac{\partial^2 v}{\partial x^2})\frac{\Delta x^2}{2} + ...$; $v_{j+1}-v_j/\Delta x = (\frac{\partial v}{\partial x})_j+(\frac{\partial^2 v}{\partial x^2})\frac{\Delta x}{2!}+... = (\frac{\partial v}{\partial x})_j+O(\Delta x)$. <u>**QED**</u>. (b) $u_j-u_{j-1} = u_j - (u_j - (\frac{\partial u}{\partial x})_j\Delta x + (\frac{\partial^2 u}{\partial x^2})\frac{\Delta x^2}{2!} + ...) = (\frac{\partial u}{\partial x})_j\Delta x - (\frac{\partial^2 u}{\partial x^2})\frac{\Delta x^2}{2!} + - + ...$; $u_j-u_{j-1}/\Delta x = (\frac{\partial u}{\partial x})_j - (\frac{\partial^2 u}{\partial x^2})\frac{\Delta x}{2!} + - ... = (\frac{\partial u}{\partial x})_j + O(\Delta x)$. <u>**QED**</u>. (c) $\phi_{j+1} - \phi_{j-1} = \phi_j + (\frac{\partial \phi}{\partial x})_j\Delta x + (\frac{\partial^2 \phi}{\partial x^2})\frac{\Delta x^2}{2!} + ... - \phi_j + (\frac{\partial \phi}{\partial x})_j\Delta x - (\frac{\partial^2 \phi}{\partial x^2})\frac{\Delta x^2}{2} + ... = 2((\frac{\partial \phi}{\partial x})_j\Delta x + (\frac{\partial^3 \phi}{\partial x^3})\frac{\Delta x^3}{3!} + ...)$. Now $\phi_{j+1}-\phi_{j-1}/2\Delta x = (\frac{\partial \phi}{\partial x})_j + (\frac{\partial^3 \phi}{\partial x^3})_j\frac{\Delta x^2}{3!} + ... = (\frac{\partial \phi}{\partial x})_j + O(\Delta x^2)$. <u>**QED**</u>. (d) $v_{j+1}-2v_j+v_{j-1} = v_j + (\frac{\partial v}{\partial x})_j\Delta x + (\frac{\partial^2 v}{\partial x^2})\frac{\Delta x^2}{2!} + (\frac{\partial^3 v}{\partial x^3})\frac{\Delta x^3}{3!} +... (-2v_j) + v_j - (\frac{\partial v}{\partial x})\Delta x + (\frac{\partial^2 v}{\partial x^2})\frac{\Delta x^2}{2!} - ().... = 2((\frac{\partial^2 v}{\partial x^2})_j\frac{\Delta x^2}{2!} + (\partial^4 v/\partial x^4)_j(\Delta x^4/4!) + ...)$. So $v_{j+1}-2v_j+v_{j-1}/\Delta x^2 = (\frac{\partial^2 v}{\partial x^2})_j + O(\Delta x^2)$. <u>**QED**</u>. (e) $v_{j+2} = v_j + \frac{\partial v}{\partial x}2\Delta x + \frac{\partial^2 v}{\partial x^2}\frac{4\Delta x^2}{2!} + \frac{\partial^3 v}{\partial x^3}\frac{8\Delta x^3}{3!} + ...$ So $-3v_j+4v_{j+1} - v_{j+2} = -3v_j + 4(v_j + \frac{\partial v}{\partial x}\Delta x + \frac{\partial^2 v}{\partial x^2}\frac{\Delta x^2}{2!} + \frac{\partial^3 v}{\partial x^3}\frac{\Delta x^3}{3!} + ...) - (v_j + \frac{\partial v}{\partial x}2\Delta x + \frac{\partial^2 v}{\partial x^2}\frac{4\Delta x^2}{2!} + \frac{\partial^3 v}{\partial x^3}\frac{8\Delta x^3}{3!} + ...) = \frac{\partial v}{\partial x}2\Delta x + \frac{\partial^3 v}{\partial x^3}(-\frac{4\Delta x^3}{3!}) + ...$ So $-3v_j+4v_{j+1}-v_{j+2}/2\Delta x = \frac{\partial v}{\partial x} + \frac{\partial^3 v}{\partial x^3}(-\frac{2\Delta x^2}{3!}) + ... = \frac{\partial v}{\partial x} + O(\Delta x^2)$. <u>**QED**</u>.

(f) $2v_j-5v_{j+1}+4v_{j+2}-v_{j+3} = 2v_j-5(v_j+\frac{\partial v}{\partial x}\Delta x+\frac{\partial^2 v}{\partial x^2}\frac{\Delta x^2}{2!}+\frac{\partial^3 v}{\partial x^3}\frac{\Delta x^3}{3!}+...) + 4(v_j+\frac{\partial v}{\partial x}2\Delta x+\frac{\partial^2 v}{\partial x^2}\frac{4\Delta x^2}{2!}+\frac{\partial^3 v}{\partial x^3}\frac{8\Delta x^3}{3!}+...) - (v_j+\frac{\partial v}{\partial x}3\Delta x+\frac{\partial^2 v}{\partial x^2}\frac{9\Delta x^2}{2!}+\frac{\partial^3 v}{\partial x^3}\frac{3^3\Delta x^3}{3!}+...) = v_j(2-5+4-1) + \frac{\partial v}{\partial x}\Delta x(-5+8-3) + \frac{\partial^2 v}{\partial x^2}\frac{\Delta x^2}{2!}(-5+16-9) + \frac{\partial^3 v}{\partial x^3}\frac{\Delta x^3}{3!}(-5+32-3^3) + (\partial^4 v/\partial x^4)(\Delta x^4/4!)(-5+64-81) = \frac{\partial^2 v}{\partial x^2}\Delta x^2 + (-\frac{22}{24})\Delta x^4(\partial^4 v/\partial x^4)$. Therefore, $2v_j-5v_{j+1}+4v_{j+2}-v_{j+3}/\Delta x^2 = \frac{\partial^2 v}{\partial x^2}+\Delta x^2(-\frac{22}{24}(\partial^4 v/\partial x^2) = \frac{\partial^2 v}{\partial x^2}+O(\Delta x^2)$. <u>**QED**</u>. Look at *sample solutions*, and remember that formulae (e) and (f) may be useful to obtain **second order accurate approximations** at *boundaries*, i.e. when v_{j-1} and v_{j-2} are not available.

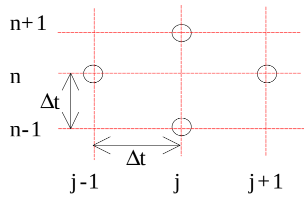
Q: Use *central differences* to derive a **difference formula** for $\partial^2u/\partial x\partial y$, and give its *order of accuracy*. A: Let $\delta x = u_{p+½,q} + u_{p-½q}$; and let $\delta y = u_{p,q+½} + u_{p,q-½}$. *Now* $u_{p\pm½,q} = u_{p,q} \pm \frac{\partial u}{\partial x}\frac{\Delta x}{2} + \frac{\partial^2 u}{\partial x^2}\frac{\Delta x^2}{4(2!)} + \frac{\partial^3 u}{\partial x^3}\frac{\Delta x^3}{8(3!)} + \ldots$; and $u_{p,q\pm½} = u_{p,q} \pm \frac{\partial u}{\partial x}\frac{\Delta y}{2} + \frac{\partial^2 u}{\partial y^2}\frac{\Delta y^2}{4(2!)} \pm \frac{\partial^3 u}{\partial y^3}\frac{\Delta y^3}{8(3!)} + \ldots$ Let us **calculate** $(u_{p+½,q}-u_{p,q})(u_{p,q+½}-u_{p,q}) = \frac{\partial u}{\partial x}\frac{\Delta x}{2}\left(\frac{\partial u}{\partial y}\frac{\Delta y}{2}+\frac{\partial^2 u}{\partial y^2}\frac{\Delta y^2}{4(2!)}+\ldots\right) + \frac{\partial^2 u}{\partial x^2}\frac{\Delta x^2}{4(2!)}(\ldots) = \frac{\partial^2 u}{\partial x\partial y}\frac{\Delta x\Delta y}{4} + \frac{\partial^3 u}{\partial x\partial y^2}\frac{\Delta y^2\Delta x}{4(2!)2} + \ldots + \ldots.$ Therefore, $\frac{4}{\Delta x\Delta y}(u_{p+½,q}-u_{p,q})(u_{p,q+½}-u_{p,q}) = \frac{\partial^2 u}{\partial x\partial y} + a\Delta y + b\Delta y^2 + \ldots + c\Delta c + d\Delta x\Delta y + \ldots + \ldots. = \frac{\partial^2 u}{\partial x\partial y} + O(\Delta y) + O(\Delta x)$.

Exercises 4



Q: Construct a *difference scheme* for $\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$ using <u>central differences</u> only. What *problems* does this pose in starting the solution? A: Central difference: $\delta v_j = v_{j+½}-v_{j-½}$. <u>**Notation**</u>: Denote $U(x,t)$ (at $x_j = j\Delta x$ and $t_n = n\Delta t$) by $u_{j,n}$. It follows that $[u_{j,n+½}-u_{j,n-½}]/\Delta$ $= \frac{\partial}{\partial x}([u_{j+½,n}-u_{j-½,n}]/\Delta x)$; $[u_{j,n+½}-u_{j,n-½}]/\Delta t = (\frac{1}{\Delta x})[u_{j+1,n}-u_{j,n}-u_{j,n}+u_{j-1,n}/\Delta x]$ $u_{j,n+½}-u_{j,n-½}/\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$. <u>*Problem*</u>: we need **four** values to get the <u>next</u> value.

Q: The **leap frog** scheme for the diffusion equation $\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} = 0$ uses *central differences* for both <u>time</u> and <u>space</u>, and results in the following three level difference equation: $u_{j,n+1} = u_{j,n-1} + 2r(u_{j+1,n} - 2u_{j,n} + u_{j-1,n})$, where $r = \frac{\Delta t}{\Delta x^2}$. (a) Prove that the *local truncation error* of this scheme is $O(\Delta t^2) + O(\Delta x^2)$. (b) Use the *von Neumann* method to show that the amplification factor g satisfies $g^2+8r\sin^2(\frac{k\Delta x}{2})g-1 = 0$, for any integer k, and hence deduce that the leap frog scheme is *unconditionally unstable*.



A: *Schematically*, we think of the scheme as shown on the **right**. (a) Consider the Taylor series expansion about $x = x_j$ and $t = t_n$. *Taylor* series in t: $u_{j,n\pm1} = U(x_j, t_n\pm\Delta t) = u_{j,n} \pm (\frac{\partial U}{\partial t})_{j,n}\Delta t + (\frac{\partial^2 U}{\partial t^2})_{j,n}\frac{\Delta t^2}{2!} \pm (\frac{\partial^3 U}{\partial t^3})_{j,n}\frac{\Delta t^3}{3!} + (\partial^4U/\partial t^4)_{j,n}(\Delta t^4/4!) + O(\Delta t^5)$. Note: we could *write* for the **final** term $\pm(\partial^5U/\partial t^5)_{j,\tau_n}(\Delta t^5/5!)$, where $t_n \le \tau_n \le t_{n+1}$. **Taylor** series in x: $U_{j\pm1,n} = U(x_j \pm \Delta x, t_n)$ = exactly as above, but **replace** the t's with x's.

The above comes from the *Taylor series expansion* of f(x) about x = a: $f(a+h) = f(a) + \frac{df}{dx}(a)h + \frac{d^2f}{dx^2}(a)\frac{h^2}{2!} + \ldots + (d^5f/dx^5)(a+\zeta)(h^5/5!)$, where $a \le \zeta \le h$. Now the *leapfrog* scheme can be **rewritten** as $u_{j,n+1}-u_{j,n-1}/2\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$. <u>Substitution</u>. *RHS:* $(\frac{1}{\Delta x^2})(2(\frac{\partial^2 U}{\partial x^2})_{j,n}\frac{\Delta x^2}{2} + 2(\partial^4U/\partial x^4)_{j,n}(\Delta x^4/4!) + \ldots)$. *LHS*: $(\frac{1}{2\Delta t})(2(\frac{\partial U}{\partial t})_{j,n}\Delta t + 2(\frac{\partial^3 U}{\partial t^3})\frac{\Delta t^3}{3!} + \ldots.)$. So *we have* $(\frac{1}{\Delta t})((\frac{\partial U}{\partial t})\Delta t + (\frac{\partial^3 U}{\partial t^3})\frac{\Delta t^3}{3!}+\ldots) = (\frac{2}{\Delta x^2})((\frac{\partial^2 U}{\partial x^2})\frac{\Delta x^2}{2} + (\partial^4U/\partial x^4)(\Delta x^4/4!)+\ldots)$.

So obtaining the *difference error*, $\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} - [(\frac{1}{\Delta t})((\frac{\partial U}{\partial t})\Delta t + (\frac{\partial^3 U}{\partial t^3})\frac{\Delta t^3}{3!} + \ldots) - (\frac{2}{\Delta x^2})((\frac{\partial^2 U}{\partial x^2})\frac{\Delta x^2}{2} + (\partial^4U/\partial x^4)(\Delta x^4/4!))] = 0$; $\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} - \frac{\partial U}{\partial t} - \frac{\partial^3 U}{\partial t^3}\frac{\Delta t^2}{3!} - \ldots + \frac{\partial^2 U}{\partial x^2} + (\partial^4U/\partial x^4)\Delta x^2(\frac{2}{4!}) + \ldots = 0$. **Therefore,** $-(\frac{\partial^3 U}{\partial t^3}\frac{\Delta t^2}{3!} + (\partial^5U/\partial t^5)(\Delta t^4/5!) + \ldots) + ((\partial^4U/\partial x^4)\Delta x^2(\frac{2}{4!}) + \ldots) = 0$. So we have an *error* $O(\Delta t^2) + O(\Delta x^2)$. **QED**. (b) Let $\epsilon_{j,n} = \hat{\epsilon}_n e^{ikx_j}$. Therefore, we obtain the following: $\hat{\epsilon}_{n+1}e^{ikx_j} = \hat{\epsilon}_{n-1}e^{ikx_j} + 2r[\hat{\epsilon}_n e^{ikx_{j+1}} - 2\hat{\epsilon}_n e^{ikx_j} + \hat{\epsilon}_n e^{ikx_{j-1}}] = \hat{\epsilon}_{n-1}e^{ikx_j} + 2r\hat{\epsilon}_n[e^{ikx_j}e^{ik\Delta x} - 2e^{ikx_j} + e^{ikx_j}e^{-ik\Delta x}]$. It follows that $\hat{\epsilon}_{n+1} = \hat{\epsilon}_{n-1} + 2r\hat{\epsilon}_n[e^{ik\Delta x}-2+e^{-ik\Delta x}] = \hat{\epsilon}_{n-1} + 2r\hat{\epsilon}_n[2\cos(k\Delta x)-2]$. Now write $\hat{\epsilon}_n = g\hat{\epsilon}_{n-1}$, so that $\hat{\epsilon}_{n-1} = \frac{1}{g}\hat{\epsilon}_n$. Therefore, $\hat{\epsilon}_{n+1} = \frac{1}{g}\hat{\epsilon}_n + 2r\hat{\epsilon}_n[2\cos(k\Delta x)-2]$; $\hat{\epsilon}_{n+1} = (\frac{1}{g}+2r(2\cos(k\Delta x)-2))\hat{\epsilon}_n$.

Now $\cos(2\theta) = 1-2\sin^2\theta$; $2\cos(2\theta) = 2-4\sin^2\theta$; $2\cos(k\Delta x) = 2-4\sin^2(^{k\Delta x}/_2)$. So $\hat{\epsilon}_{n+1} = (^1/_g+2r(2-4\sin^2(^{k\Delta x}/_2)-2))\hat{\epsilon}_n = (^1/_g-8r\sin^2(^{k\Delta x}/_2))\hat{\epsilon}_n$. **Therefore,** $g = {}^1/_g-8r\sin^2(^{k\Delta x}/_2)$; $g^2 = 1-8r\sin^2(^{k\Delta x}/_2)g$; $g^2+8r\sin^2(^{k\Delta x}/_2)g-1 = 0$. **QED**. Now deduce that the scheme is *unconditionally unstable* — we have to have prove that $|g| \leq 1$ for **stability**. Consider that g has *roots* as follows:

$(g-g_1)(g-g_2) = 0$; $g^2-gg_1-gg_2+g_1g_2 = 0$; $g^2-g(g_1+g_2)+g_1g_2 = 0$. We *have* to obey $|g_1| \leq 1$, and $|g_2| \leq 1$. *Comparing* to $g^2+8r\sin^2(^{k\Delta x}/_2)g-1 = 0$, we have $g_1g_2 = -1$, and $g_1+g_2 = -8r\sin^2(^{k\Delta x}/_2)$. From the *first equation*, if $g_1 \leq 1$, then we must have $g_2 \geq 1$ — which is **invalid** for $g_2 > 1$. Similarly, if $g_1 \geq 1$, then we must have $g_2 \leq 1$ — which is **invalid** for $g_1 > 1$.

Therefore, for stability, we must have $g_1 = 1$, and $g_2 = -1$ — or vice versa. But then, *equation* 2 $\Rightarrow 0 = -8r\sin^2(^{k\Delta x}/_2)$, which is *not allowed*. **Conclusion**: the scheme is unconditionally unstable — we cannot find values that give $|g| \leq 1$.

Q: The *Dufort Frankell scheme* is a **modification** of the leap frog scheme, in that it replaces $2u_{j,n}$ in the spatial difference by its time average, $(u_{j,n+1}+u_{j,n-1})$, which results in the following *three level difference scheme* for the diffusion equation $^{\partial U}/_{\partial t}-^{\partial^2 U}/_{\partial x^2} = 0$: $u_{j,n+1} = u_{j,n-1} + 2r\{u_{j+1,n} - (u_{j,n+1}+u_{j,n-1}) + u_{j-1,n}\}$, where $r = {}^{\Delta t}/_{\Delta x^2}$.

(a) Use *Taylor Series expansions* to show that the local truncation error is $T_{j,n} = {}^{\Delta x^2}/_{12}(\partial^4U/\partial x^4)_{j,n} - {}^{\Delta t^2}/_6(^{\partial^3 U}/_{\partial t^3})_{j,n} + {}^{\Delta t^2}/_{\Delta x^2}(^{\partial^2 U}/_{\partial t^2})_{j,n}+\ldots$ (b) Deduce that the *scheme is consistent* **only** if $\Delta t\rightarrow 0$ faster than Δx. (c) If $^{\Delta t}/_{\Delta x} = C$ as Δt, $\Delta x\rightarrow 0$, where $C > 0$ is a constant, show that the scheme is *inconsistent*, and simulates $^{\partial U}/_{\partial t}-^{\partial^2 U}/_{\partial x^2} = C^2\,{}^{\partial^2 U}/_{\partial t^2}$. (d) Use the *Fourier Method* to show that the **amplification** factor is given by $g = [2r\cos(k\Delta x) \pm (\sqrt{1-4r^2\sin^2(k\Delta x)})]/1+2r$, and *deduce* that the scheme is unconditionally stable.

A: Using *Taylor Series expansions*, $u_{j,n\pm 1} = u_{j,n} \pm (^{\partial U}/_{\partial t})\Delta t + (^{\partial^2 U}/_{\partial t^2})^{\Delta t^2}/_{2!} \pm (^{\partial^3 U}/_{\partial t^3})^{\Delta t^3}/_{3!} + \ldots$; *and* $u_{j\pm 1,n} = u_{j,n} \pm (^{\partial U}/_{\partial x})\Delta x + (^{\partial^2 U}/_{\partial x^2})^{\Delta x^2}/_{2!} \pm (^{\partial^3 U}/_{\partial x^3})^{\Delta x^3}/_{3!} + \ldots$ Now the *difference scheme* is $u_{j,n+1} = u_{j,n-1} + 2\,{}^{\Delta t}/_{\Delta x^2}\{\ldots\}$, or $u_{j,n+1}-u_{j,n-1}/\Delta t = 2u_{j+1,n}+2u_{j-1,n}/\Delta x^2 - {}^2/_{\Delta x^2}(u_{j,n+1}+u_{j,n-1})$. **Substituting** in the Taylor series', $^2/_{\Delta t}((^{\partial U}/_{\partial t})\Delta t + (^{\partial^3 U}/_{\partial t^3})^{\Delta t^3}/_{3!} + \ldots) = {}^2/_{\Delta x^2}(2u_{j,n} + 2(^{\partial^2 U}/_{\partial x^2})^{\Delta x^2}/_{2!} + 2\ldots) - {}^2/_{\Delta x^2}(2(u_{j,n}+(^{\partial^2 U}/_{\partial t^2})^{\Delta t^2}/_{2!} + (\partial^4U/\partial t^4)\Delta t^4/4! + \ldots))$.

Taking the *difference* to get the truncation error, $^{\partial U}/_{\partial t} - {}^{\partial^2 U}/_{\partial x^2} - ((^{\partial U}/_{\partial t}) + (^{\partial^3 U}/_{\partial t^3})^{\Delta t^2}/_{3!} + (\partial^5U/\partial t^5)\Delta t^4/5! + \ldots) + {}^1/_{\Delta x^2}(2(u_{j,n} + (^{\partial^2 U}/_{\partial x^2})^{\Delta x^2}/_{2!} + \ldots)) - {}^2/_{\Delta x^2}(u_{j,n} + (^{\partial^2 U}/_{\partial t^2})^{\Delta t^2}/_{2!} + \ldots) = 0$. So *we have* $-((^{\partial^3 U}/_{\partial t^3})^{\Delta t^2}/_{3!} + (\partial^5U/\partial t^5)\Delta t^4/5! + \ldots) + {}^2/_{\Delta x^2}((\partial^4U/\partial x^4)\Delta x^4/4! + (\partial^6U/\partial x^6)\Delta x^6/6!+\ldots) - {}^2/_{\Delta x^2}((^{\partial^2 U}/_{\partial t^2})^{\Delta t^2}/_{2!} + \ldots)$; or $-(^{\partial^3 U}/_{\partial t^3})^{\Delta t^2}/_6 + (\partial^4U/\partial x^4)^{\Delta x^2}/_{12} - {}^{\Delta t^2}/_{\Delta x^2}(^{\partial^2 U}/_{\partial t^2}) + \ldots$. **QED**.

(b) If we *let* $\Delta t\rightarrow 0$, and let $\Delta x\rightarrow 0$, then the first **two** series disappear with no problem. It is the $^{\Delta t^2}/_{\Delta x^2}$ terms that are *troublesome*. Letting just $\Delta x^2\rightarrow 0$, then the *third* series "blows up". So we indeed need Δt^2 to tend to zero *faster* than Δx^2 for the l.t.e. to tend to zero — and not go off to ∞. (c) $^{\Delta t}/_{\Delta x} = C$, so that $^{\Delta t^2}/_{\Delta x^2} = C^2$. Again, when Δx, $\Delta t\rightarrow 0$, the first two series disappear.

We are *left* with the $C^2(\partial^2U/\partial t^2)_{j,n}$ term only — which does **not** tend to zero (as $C > 0$). Therefore, the scheme is *inconsistent*. Note that it simulates $\partial U/\partial t - \partial^2U/\partial x^2 = C^2\partial^2U/\partial t^2$, because when taking the *difference on the previous page*, the blue terms will cancel, and so, in this *case*, the l.t.e. will tend to zero. (d) $\hat{\epsilon}_{n+1}e^{ikx_j} = \hat{\epsilon}_{n-1}e^{ikx_j} + 2r\{\hat{\epsilon}_n e^{ikx_{j+1}} - \hat{\epsilon}_{n+1}e^{ikx_j} - \hat{\epsilon}_{n-1}e^{ikx_j} + \hat{\epsilon}_n e^{ikx_{j-1}}\} = \hat{\epsilon}_{n-1}e^{ikx_j} + 2r\{\hat{\epsilon}_n e^{ikx_j}e^{ik\Delta x} - \hat{\epsilon}_{n+1}e^{ikx_j} - \hat{\epsilon}_{n-1}e^{ikx_j} + \hat{\epsilon}_n e^{ikx_j}e^{-ik\Delta x}\}$. **Cancelling**, we obtain $\hat{\epsilon}_{n+1} = \hat{\epsilon}_{n-1} + 2r\{\hat{\epsilon}_n e^{ik\Delta x} - \hat{\epsilon}_{n+1} - \hat{\epsilon}_{n-1} + \hat{\epsilon}_n e^{-ik\Delta x}\}$.

Now $\hat{\epsilon}_{n+1}(1+2r) = \hat{\epsilon}_{n-1}(1-2r) + 2r\{\hat{\epsilon}_n(e^{ik\Delta x}+e^{-ik\Delta x})\} = \hat{\epsilon}_{n-1}(1-2r) + 2r\hat{\epsilon}_n\cos(k\Delta x)(2)$. It follows that $\hat{\epsilon}_n = g\hat{\epsilon}_{n-1}$, so that $\hat{\epsilon}_{n-1} = {}^1/_g\hat{\epsilon}_n$. Therefore, $\hat{\epsilon}_{n+1}(1+2r) = {}^1/_g\hat{\epsilon}_n(1-2r) + 4r\hat{\epsilon}_n\cos(k\Delta x)$; $\hat{\epsilon}_{n+1}(1+2r) = \hat{\epsilon}_n({}^{1-2r}/_g + 4r\cos(k\Delta x))$. So $g = ({}^1/_{1+2r})({}^{1-2r}/_g + 4r\cos(k\Delta x))$; and thus $g^2(1+2r) = (1-2r) + 4rg\cos(k\Delta x)$.

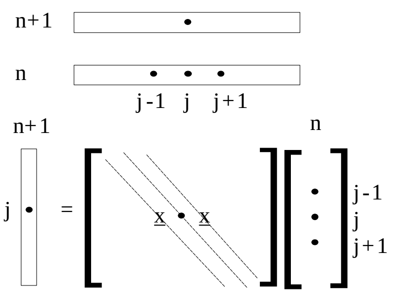
Applying the *quadratic formula*, $g = {}^{+4r\cos(k\Delta x)\ \pm\ (\sqrt{16r^2\cos^2(k\Delta x)+4(1+2r)(1-2r)})}/_{2(1+2r)} = {}^{4r\cos(k\Delta x)\ \pm\ 2(\sqrt{4r^2\cos^2(k\Delta x)+(1+2r)(1-2r)})}/_{2(1+2r)} = {}^{2r\cos(k\Delta x)\ \pm\ (\sqrt{4r^2\cos^2(k\Delta x)+(1-4r^2)})}/_{(1+2r)}$. Now concentrate on the *square root:* $4r^2\cos^2(k\Delta x) + (1-4r^2) = 4r^2(\cos^2(k\Delta x)-1) + 1 = -4r^2\sin^2(k\Delta x) + 1$. Therefore, $g = {}^{2r\cos(k\Delta x)\ \pm\ (\sqrt{-4r^2\sin^2(k\Delta x)+1})}/_{(1+2r)}$. **QED**.

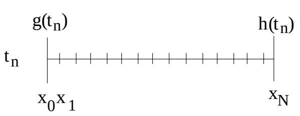
Now we must have $|g| \le 1$, or $-1 \le g \le 1$; $-1 \le {}^{2r\cos(k\Delta x)\ \pm\ (\sqrt{-4r^2\sin^2(k\Delta x)+1})}/_{(1+2r)} \le 1$; $-1-2r \le 2r\cos(k\Delta x) \pm \sqrt{(1-4r^2\sin^2(k\Delta x))} \le 1+2r$. Now for the red term, $-1 \le \cos(k\Delta x) \le 1$; $-2r \le 2r\cos(k\Delta x) \le 2r$. For the blue term, $0 \le \sin^2(k\Delta x) \le 1$; $0 \ge -4r^2\sin^2(k\Delta x) \ge -4r^2$, **or** $-4r^2 \le -4r^2\sin^2(k\Delta x) \le 0$; $1-4r^2 \le 1-4r^2\sin^2(k\Delta x) \le 1$; $\sqrt{(1-4r^2)} \le \sqrt{(1-4r^2\sin^2(k\Delta x))} \le 1$. **But** $\sqrt{(1-4r^2)} \ge 0$, so $0 \le \sqrt{(1-4r^2\sin^2(k\Delta x))} \le 1$. So we **have** $-1-2r \le A \pm B \le 1+2r$, *where* $-2r \le A \le 2r$, *and* $0 \le B \le 1$. Clearly, **any** combination of $A \pm B$ will not stray <u>outside</u> the bounds of our inequality — so we can say that the scheme *is* unconditionally stable, as $|g| \le 1$.

2nd November 2000

<u>Matrix Method</u>

In this method, we express the *difference scheme* (over the **whole** mesh space) for a single timestep in the form of a <u>matrix equation</u>. The *properties* of the matrix control whether the scheme is **stable,** e.g. $\underline{u}_{n+1} = A\underline{u}_n$; $||\underline{u}_{n+1}|| \le ||\underline{u}_n||$; and $||A|| \le 1$. The *largest eigenvalue* in magnitude is a measure of the "<u>norm</u>" of the matrix. Consider $\partial U/\partial t = \partial^2U/\partial x^2$ on $0 \le x \le 1$, with *boundary* condition $U(0,t) = g(t)$, with $t > 0$; and $U(1,t) = h(t)$, with $t > 0$.





The **mesh** looks like as shown on the left. At $t = t_n$, $x_0 = g(t_n)$, and $x_N = h(t_n)$. The *difference scheme* is $u_{j,n+1} = ru_{j+1,n} + (1-2r)u_{j,n} + ru_{j-1,n}$. Applying the <u>simple explicit scheme,</u> we get the shown *matrix* equation for the difference scheme, relating the values at t_{n+1} to those at t_n. It **can** be expressed as $\underline{u}_{n+1} = A\underline{u}_n + \underline{c}_n$, where A is a *tridiagonal* matrix.

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \cdots \\ \cdots \\ u_{N-2} \\ u_{N-1} \end{pmatrix}_{n+1} = \begin{pmatrix} 1-2r & r & & & 0 \\ r & 1-2r & r & & \\ & \cdots & \cdots & \cdots & \\ & & \cdots & \cdots & \\ & & & \cdots & \\ & & & \cdots & r \\ 0 & & & r & 1-2r \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \\ \\ u_{N-2} \\ u_{N-1} \end{pmatrix} + \begin{pmatrix} rg \\ 0 \\ \\ \cdots \\ \\ 0 \\ rh \end{pmatrix}_n$$

We now suppose that **roundoff** errors enter along the *first line* at $t = t_0$, and if no **further** errors are made, then $\underline{u}_1 = A\underline{u}_0+\underline{c}_0$ (*exact*), and $\bar{\mathbf{u}}_1 = A\bar{\mathbf{u}}_0 + \bar{\mathbf{c}}_0$ (note that *the over bars* denote **‘rounded** values’). Subtracting gives the *error* at the first time step, $\underline{e}_1 = A\underline{e}_0 + \underline{b}_0$. This process is continued to give, at the $(n+1)^{th}$ time step, $\underline{e}_{n+1} = A\underline{e}_n+\underline{b}_n$, where $\underline{e}_n = \underline{u}_n-\bar{\mathbf{u}}_n$, and $\underline{b}_n = \underline{c}_n-\bar{\mathbf{c}}_n$. *No new errors* are introduced, so that $\underline{b}_n = 0$ for $n > 0$.

At the **second** step, $\underline{e}_2 = A\underline{e}_1 = A(A\underline{e}_0+\underline{b}_0) = A^2\underline{e}_0 + A\underline{b}_0$, and, at the $(n+1)^{th}$ time step, $\underline{e}_n = A^n\underline{e}_0+A^{n-1}\underline{b}_0$. *Assuming* that A has N-1 distinct eigenvalues, we **expand** $\underline{e}_0$ and $\underline{b}_0$ in terms of the *eigenvectors* of A. If we denote the eigenvectors of A by $\underline{\phi}_1, \underline{\phi}_2, \ldots, \underline{\phi}_{N-1}$, *then* $\underline{e}_0 = \alpha_1\underline{\phi}_1 + \alpha_2\underline{\phi}_2 + \ldots + \alpha_{N-1}\underline{\phi}_{N-1}$, and $\underline{b}_0 = \beta_1\underline{\phi}_1 + \beta_2\underline{\phi}_2 + \ldots + \beta_{N-1}\underline{\phi}_{N-1}$. If λ_k is the eigenvalue of A that has *associated* eigenvector $\underline{\phi}_k$, then $A\underline{\phi}_k = \lambda_k\underline{\phi}_k$; $A^2\underline{\phi}_k = \lambda^2_k\underline{\phi}_k$; …; $A^n\underline{\phi}_k = \lambda^n_k\underline{\phi}_k$, for $1 \le k \le N-1$.

It follows that $\underline{e}_n = A^n\underline{e}_0 + A^{n-1}\underline{b}_0$; $\underline{e}_n = \alpha_1\lambda^n_1\underline{\phi}_1 + \alpha_2\lambda^n_2\underline{\phi}_2 + \ldots + \alpha_{N-1}\lambda^n_{N-1}\underline{\phi}_{N-1} + \beta_1\lambda^{n-1}_1\underline{\phi}_1 + \beta_2\lambda_2^{n-1}\underline{\phi}_2 + \ldots + \beta_{N-1}\lambda^{n-1}_{N-1}\underline{\phi}_{N-1}$. For the error to remain *bounded*, the magnitude of the **largest** eigenvalue of A must have *modulus* ≤ 1. This is called the spectral radius of A, and is denoted by $\rho(A)$. We *need* $\rho(A) \le 1$.

7th November 2000

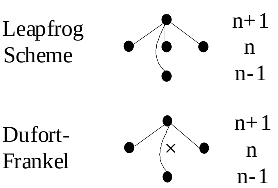
$$\begin{pmatrix} a & b & & & \\ c & a & b & & \\ & c & a & b & \\ & & \ddots & \ddots & \ddots \\ & & \ddots & \ddots & b \\ & & & c & a \end{pmatrix}$$

It can be *shown* (see a book) that the Eigenvalues of an M×M tridiagonal matrix, as shown on the right, are $\lambda_k = a+2\sqrt{(bc)}\cos(^{k\pi}/_{M+1})$, where $k = 1,2,\ldots,M$. In *our case*, $M = N-1$; $a = 1-2r$; $b = r$; and $c = r$. So $\lambda_k = 1-2r+2r\cos(^{k\pi}/_N)$, with $k = 1,2,\ldots,N-1$. Now $\lambda_k = 1-4r\sin^2(^{k\pi}/_{2N})$ (*use* $\cos 2\theta = 1-2\sin^2\theta$). For $|\lambda_k| \le 1$, we need $|1-4r\sin^2(^{k\pi}/_{2N})| \le 1$, *or* $-1 \le 1-4r\sin^2(^{k\pi}/_{2N}) \le 1$. The *right hand inequality* is always true, and the left hand inequality **gives** $4r\sin^2(^{k\pi}/_{2N}) \le 2$. Now $\sin^2(^{k\pi}/_{2N}) \le 1$ for every k. So we require $4r \le r$, i.e. $r \le ½$ ($r = {}^{\Delta t}/_{\Delta x^2}$). This is the *same result* as given by the Fourier method. Taking account of the *boundary conditions* does not usually affect the stability condition — **stability** is primarily a property of the difference scheme.

Lax Equivalence Theorem

The following theorem is important as it links the three ideas of *consistency*, *stability and convergence*. (**Consistency** and **stability** are usually far easier to prove for than convergence). **Theorem**: Given a *properly-posed linear initial value problem*, and a *finite difference approximation* to it that is consistent, then the scheme is convergent if and only if it is stable.

Computational Stencils / Molecules



Other Explicit Schemes. *Leap Frog* (Exercises 4, Q3): $u_{j,n+1}-u_{j,n-1}/2\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$. This scheme is **consistent**, with local truncation error $O(\Delta t^2)+O(\Delta x^2)$, but it is *unconditionally* unstable. *Dufort-Frankel* (Exercise 4, Q4): this scheme *replaces* $u_{j,n}$ by its time average in the *leapfrog* scheme: $u_{j,n+1}-u_{j,n-1}/2\Delta x = u_{j+1,n}-2(^1/_2)(u_{j,n+1} + u_{j,n-1})+u_{j-1,n}/\Delta x^2$. The *Dufort-Frankel* scheme is **consistent** if $\Delta t\to 0$ faster than Δx; and is *unconditionally* stable.

Parabolic Equations: Implicit Methods

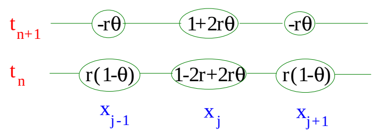
Explicit methods usually require small timesteps so as to satisfy a **stability** condition, e.g. $r = {}^{\Delta t}/_{\Delta x^2} \leq \frac{1}{2}$. This problem is *exacerbated* in 2 and 3 dimensions. Implicit methods allow a larger timestep, but require the solution of a *matrix equation* at each timestep.

8th November 2000

Implicit Methods for the 1-D Heat Equation

Recall that we had ${}^{\partial U}/_{\partial t} = {}^{\partial^2 U}/_{\partial x^2}$, and $u_{j,n+1}-u_{j,n}/\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$. Instead of *evaluating* the **spatial derivative** at time t_n, we use a *weighted* average at times t_n and t_{n+1}: $u_{j,n+1}-u_{j,n}/\Delta t = (1-\theta)[u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2] + \theta[u_{j+1,n+1}-2u_{j,n+1}+u_{j-1,n+1}/\Delta x^2]$, where $0 \leq \theta \leq 1$. Note that $\theta = 0$ gives the *simple explicit scheme* (**Forward** Euler); $\theta = \frac{1}{2}$ gives the *Crank-Nicolson* scheme; and $\theta = 1$ gives the *fully implicit* (**Backward** Euler) scheme.

Rewriting the difference scheme gives $-r\theta u_{j-1,n+1} + (1+2r\theta)u_{j,n+1} - r\theta u_{j+1,n+1} = r(1-\theta)u_{j-1,n} + (1-2r+2r\theta)u_{j,n} + r(1-\theta)u_{j+1,n}$. The *computational molecule / stencil* is as shown on the right, so in the implicit scheme, **three** unknowns at time t_{n+1} are related to three values at **time** t_n.



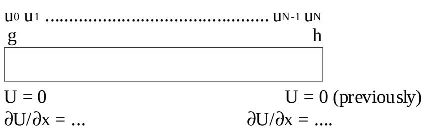
We obtain a matrix equation to *solve* at each timestep. The form of the matrix equation is as follows: $A\underline{u}_{n+1} = B\underline{u}_n + \underline{b}_n$, where $\underline{u}_{n+1}$ and $\underline{u}_n$ are the vectors of the values of the *dependent variable* at the mesh points; A and B are given by the **difference** scheme; and $\underline{b}_n$ gives the *boundary conditions.*

Consider the Crank-Nicolson scheme, where $\theta = \frac{1}{2}$: $-r\frac{1}{2}u_{j-1,n+1}+(1+r)u_{j,n+1}-r\frac{1}{2}u_{j+1,n+1} = r\frac{1}{2}u_{j-1,n}+(1-r)u_{j,n}+r\frac{1}{2}u_{j+1,n}$. The matrix equation *looks like as shown on the right*.

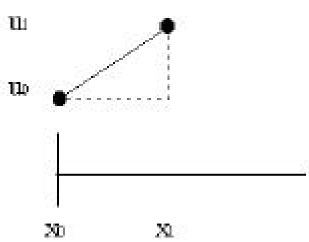
$$\begin{pmatrix} (1+r) & -r/2 & & & 0 \\ -r/2 & \ddots & \ddots & & \\ & -r/2 & (1+r) & -r/2 & \\ & & \ddots & \ddots & -r/2 \\ 0 & & & -r/2 & (1+r) \end{pmatrix} \begin{pmatrix} u_1 \\ .. \\ u_{j-1} \\ u_j \\ u_{j+1} \\ .. \\ u_{N-1} \end{pmatrix}_{n+1} = \begin{pmatrix} (1-r) & r/2 & & & 0 \\ r/2 & \ddots & \ddots & & \\ & r/2 & (1-r) & r/2 & \\ & & \ddots & \ddots & r/2 \\ 0 & & & r/2 & (1-r) \end{pmatrix} \begin{pmatrix} u_1 \\ .. \\ u_{j-1} \\ u_j \\ u_{j+1} \\ .. \\ u_{N-1} \end{pmatrix}_{n} + \underline{b}_n$$

Stability of the Implicit Scheme. If we apply the Fourier method to the implicit scheme, we get the amplification factor $g = {}^{[1-4r(1-\theta)\sin^2(k\Delta x/2)]}/_{[1+4r\theta\sin^2(k\Delta x/2)]}$. Further, the **stability** criterion turns out to be: for $\theta \geq \frac{1}{2}$, *unconditionally* stable; and for $\theta < \frac{1}{2}$, stable if $r \leq {}^1/_{2(1-2\theta)}$. Note: remember that $r = {}^{\Delta t}/_{\Delta x^2}$.

Derivative Boundary Conditions



So far, we have looked at *Dirichlet type boundary conditions*, where the dependent variable is given at the boundary. For Neumann and Cauchy type boundary conditions, we have to calculate the *boundary* values as well. There are **two** standard methods of treating derivatives. (1): Use *forward or backward spatial derivatives*. Example: Consider that we are given the boundary condition ${}^{\partial U}/_{\partial x} = kU$ at $x = 0$. The finite *difference* version of this is $u_1-u_0/\Delta x = ku_0$, which *gives* $u_0 = u_1/1+k\Delta x$. Note that this is only *first* order accurate.



(2): Introduce **guard points**. Guard points are *ficticious* points outside the domain. They allow *second order accurate* central difference approximations at the boundary, and are subsequently **eliminated** from the calculations.

9th November 2000

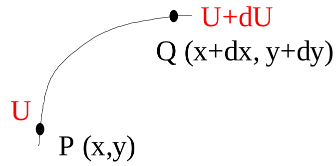
Example: If the *boundary* condition at x = 0 is $\partial U/\partial x = kU$, we can form a *second order accurate* difference as $u_1 - u_{-1}/2\Delta x = ku_0$. *Rearranging* gives $u_{-1} = u_1 - 2\Delta x k u_0$, and so u_{-1} may be **eliminated** when the system is solved. See the *exercise sheet* for a full example on this.

14th November 2000

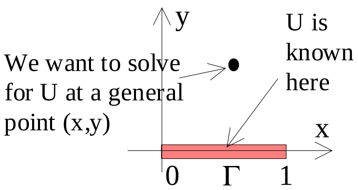
First Order Hyperbolic Equations

Consider the *first order quasi-linear partial differential equation* $a\,\partial U/\partial x + b\,\partial U/\partial y = c$, where a, b and c are in general **functions of x, y and U**, but *not* of $\partial U/\partial x$ and of $\partial U/\partial y$. We now show how the above PDE *transforms* into two ODE's if we choose to **integrate** along a particular family of curves *related* to the equation — these are the **characteristic curves** of the PDE.

Suppose that we move from P (at position (x,y)) to Q (at *position* (x+dx, y+dy)) along the curve. Let the change in the value of U between P and Q be dU. **Then** $dU = \partial U/\partial x\,dx + \partial U/\partial y\,dy$. *Compare* this with the PDE $c = a\,\partial U/\partial x + b\,\partial U/\partial y$. From the PDE, we have $\partial U/\partial x = (c - b\,\partial U/\partial y)/a$, and **substituting** this into the total *derivative, we obtain* $dU = [(c - b\,\partial U/\partial y)/a]dx + \partial U/\partial y\,dy$, or $\partial U/\partial y(a\,dy - b\,dx) + (c\,dx - a\,dU) = 0$.

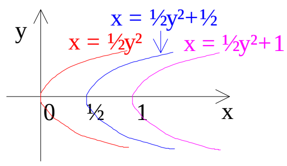


This does **not** involve $\partial U/\partial x$ at all, since a, b and c are **not** functions of $\partial U/\partial x$. Now choose the *curve on the x-y plane* such that $a\,dy - b\,dx = 0$, or $dy/dx = b/a$. This defines the *characteristics of the PDE*. The solution for U along **such** a curve must then satisfy $c\,dx - a\,dU = 0$. Therefore, $dU/dx = c/a$; and the set of *equations* can conveniently be written as $dx/a = dy/b = dU/c$.



The PDE **can** thus be solved by first finding the *characteristic curves* (by solving $dy/dx = b/a$), and then *integrating* U along such a curve (by solving $dU/dx = c/a$, or $dU/dy = c/b$). **Example**: Solve $y\,\partial U/\partial x + \partial U/\partial y = 2$, given U on y = 0, and $0 \le x \le 1$. A: The *curve* on which the values of U are specified is called the **initial curve,** Γ.

The PDE becomes $dx/y = dy/1 = dU/2$. The *characteristics* are given by $dy/dx = 1/y$. Now $\int y\,dy = \int dx$ + constant; $\frac{1}{2}y^2 = x + \text{constant}$. This *may be written as* $x = \frac{1}{2}y^2 + A$, where A is a **constant** of integration. This defines a *family of parabolas*, with intercept on the x-axis at A. Let x_j be some *value* of x in $0 \le x \le 1$. Then the *characteristic* through $(x_j, 0)$ is $x - x_j = \frac{1}{2}y^2$, or $y^2 = 2(x - x_j)$.



15th November 2000

Along *any of the characteristic curves* $y^2 = 2(x - x_j)$, we may **solve** for U by using dU = 2dy — therefore U = 2y+C by *integration* (C is a constant). This constant C is given by the value of U at y = 0, which is known because y = 0 is the initial line/curve in this example. Suppose that $U = U_j$ at $x = x_j$ on y = 0. Then $U = 2y + U_j$ **along** the characteristic $y^2 = 2(x - x_j)$.

Notes: (1) *Giving* initial values on $0 \le x \le 1$ defines the solution **only** in this region, bounded by the characteristics *through* $x = 0$ and $x = 1$ (the end points of the initial curve). (2) **Outside** this region, the solution is *undefined*. (3) A solution is only defined if the **initial** curve is not a *characteristic* curve. (4) If the **initial** values of U along the initial curve are *discontinuous*, then the solution U is discontinuous. (5) If the **characteristics cross**, then the solution becomes *multi-valued*, e.g. *shock waves* in fluid flow.

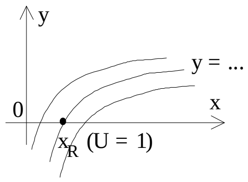
Second Order Hyperbolic Equations

The *general second order* quasi-linear PDE may be **written** as $a\,{}^{\partial^2 U}/_{\partial x^2} + b\,{}^{\partial^2 U}/_{\partial x \partial y} + c\,{}^{\partial^2 U}/_{\partial y^2} + e = 0$. We shall *use* the following *identities*: $d({}^{\partial U}/_{\partial x}) = {}^{\partial^2 U}/_{\partial x^2}dx + {}^{\partial^2 U}/_{\partial x \partial y}dy$, and $d({}^{\partial U}/_{\partial y}) = {}^{\partial^2 U}/_{\partial x \partial y}dx + {}^{\partial^2 U}/_{\partial y^2}dy$ (**Justification:** $dU = {}^{\partial U}/_{\partial x}dx + {}^{\partial U}/_{\partial y}dy$, so that $d({}^{\partial V}/_{\partial x}) = {}^{\partial}/_{\partial x}\,{}^{\partial V}/_{\partial x}dx + {}^{\partial}/_{\partial y}\,{}^{\partial V}/_{\partial x}dy$). We use these identities to *eliminate* ${}^{\partial^2 U}/_{\partial x^2}$ and ${}^{\partial^2 U}/_{\partial y^2}$ from the PDE.

This *gives* $a[{}^{d}/_{dx}({}^{\partial U}/_{\partial x}) - {}^{\partial^2 U}/_{\partial x \partial y}\,{}^{dy}/_{dx}] + b\,{}^{\partial^2 U}/_{\partial x \partial y} + c[{}^{d}/_{dy}\,{}^{\partial U}/_{\partial y} - {}^{\partial^2 U}/_{\partial x \partial y}\,{}^{dx}/_{dy}] + e = 0$. **Rearranging** gives ${}^{\partial^2 U}/_{\partial x \partial y}[-a({}^{dy}/_{dx})^2 + b({}^{dy}/_{dx}) - c] + a\,{}^{d}/_{dx}({}^{\partial U}/_{\partial x}){}^{dy}/_{dx} + c\,{}^{d}/_{dx}({}^{\partial U}/_{\partial y}) + e\,{}^{dy}/_{dx} = 0$. We try to find curves *such that* $a({}^{dy}/_{dx})^2 - b({}^{dy}/_{dx}) + c = 0$, then, along the *curves* obtained by solving the quadratic for ${}^{dy}/_{dx}$, the PDE **reduces** to $a\,{}^{d}/_{dx}({}^{\partial U}/_{\partial x}){}^{dy}/_{dx} + c\,{}^{d}/_{dx}({}^{\partial U}/_{\partial y}) + e\,{}^{dy}/_{dx} = 0$.

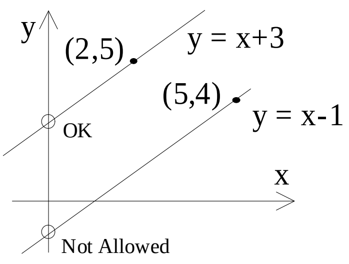
16th October 2000

Tutorial



Q: The function U *satisfies the equation* $x^2U\,{}^{\partial U}/_{\partial x} + e^{-y}\,{}^{\partial U}/_{\partial y} = -U^2$, with the condition $U = 1$ on $y = 0$, for $0 < x < \infty$. Calculate the *Cartesian* equation of the characteristic through the **point** R, $(x_R,0)$, with $x_R > 0$, and the solution *along this characteristic*. A: The PDE **becomes** ${}^{dx}/_{x^2U} = dy/e^{-y} = {}^{dU}/_{-U^2}$. Note that we cannot yet *attack* ${}^{dy}/_{dx} = ...$ because the dx part **depends** on U. So try $dy/e^{-y} = {}^{dU}/_{-U^2}$. Now $\int e^y dy = \int -U^{-2}dU$; $e^y = {}^{1}/_{U} +$ constant. Use the *initial condition* $e^0 = 1+$constant, so that the *constant is 0*, and thus $e^y = {}^{1}/_{U}$, or $\mathbf{U = e^{-y}}$. Now we *can* attack ${}^{dy}/_{dx} = e^{-y}/Ux^2$: $\int e^y dy = \int {}^{dx}/_{Ux^2}$; $\int e^y dy = \int dx.e^y/x^2$; $y = -{}^{1}/_{x} +$ constant. To *get* the constant, we know that **when** $y = 0$, $x = x_R$. So $0 = (-1/x_R) +$ constant, and thus *constant* $= 1/x_R$. Therefore, $y = (1/x_R) - {}^{1}/_{x}$ is the *characteristic curve*, with $U = e^{-y}$ being a **solution** along the characteristic.

Q: ${}^{\partial U}/_{\partial x} + {}^{x}/_{\sqrt{U}}\,{}^{\partial U}/_{\partial y} = 2x$, with $U = 0$ on $x = 0$ and $y \ge 0$; *and* $U = 0$ on $y = 0$ and $x > 0$. Calculate the **analytical solution** at the points (2,5) and (5,4); and sketch the characteristics *through* these two points. A: We have $dx = {}^{dy\sqrt{(U)}}/_{x} = {}^{dU}/_{2x}$; so that ${}^{dy}/_{dx} = {}^{x}/_{\sqrt{U}}$; $dy\sqrt{U} = xdx$ (---(1)) — *oops*! — and therefore we use ${}^{dy\sqrt{(U)}}/_{x} = {}^{dU}/_{2x}$; $dy = {}^{dU}/_{2\sqrt{(U)}}$; $y = \sqrt{(U)} +$ constant. Further, we also **use** $2xdx = dU$; $x^2 = U +$ constant. So we have $U = x^2+A$, and $y = B+\sqrt{(U)}$, where A and B are constants.

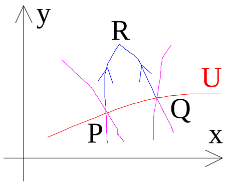


When we are at the *point* $R(x_R,0)$, $U = 0$, so that $A = -x_R^2$, and $B = 0$. Therefore, $U = x^2 - x_R^2$ (---(2)), and $U = y^2$ (---(3)). Now (3) in (2) $\Rightarrow$ $x^2 - y^2 = x_R^2$. So the solution along the *characteristic* $x^2 - y^2 = x_R^2$ from $(x_R,0)$ is $U = y^2$. When we are at the **point** $S(0,y_S)$, $A = 0$, and $B = y_S$. So $U = x^2$ (---(4)), and $y = y_S + \sqrt{(U)}$ (---(5)). Putting (4) into (5) $\Rightarrow$ $y = y_S + x$; $y - y_S = x$. So the *solution* along the characteristic $y - y_S = x$ is $U = x^2$.

At (5,4), the *first* characteristic gives $25-16 = x_R^2$; $9 = x_R^2$; $3 = x_R$. The *second* characteristic gives $4-y_S = 5$; $y_S = -1$ — not allowed, because $y_S < 0$. So we use the **first** characteristic, and so $U(5,4) = 4^2 = 16$. At (2,5), the first *characteristic* gives $4-25 = x_R^2$, not allowed. The second *characteristic* gives $5-y_S = 2$; $y_S = 3$ — which *is* allowed. So we use the **second** characteristic, and so $U(2,5) = 2^2 = 4$.

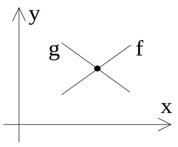
Q: $\partial U/\partial x - (x-y-1)\partial U/\partial y = x(y-x)$. A: *The solution is* $dx/1 = dy/-(x-y-1) = dU/x(y-x)$. Therefore, $dy/dx = -(x-y-1)$; $dy/dx - y = 1-x$. **Recall** that if $dy/dx + P(x)y = Q(x)$, the *general solution* is $e^{\int Pdx}y = \int Qe^{\int Pdx}dx + c$. Here, $\int Pdx = \int -1dx = -x$; so that $e^{-x}y = \int(1-x)e^{-x}dx+c$; $e^{-x}y = \int e^{-x}dx - \int xe^{-x}dx+c$. **Let** $u = x$, so that $du/dx = 1$; and let $dv/dx = e^{-x}$, so that $v = -e^{-x}$. So we *have* $e^{-x}y = -e^{-x}-[-xe^{-x}-\int e^{-x}]+c$; $e^{-x}y = -e^{-x}+xe^{-x}-e^{-x}+c$; $y = -1 + x-1 + c/e^{-x}$; $y + 2 = x + (c/e^{-x})$.

21st November 2000

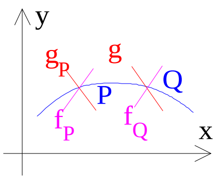


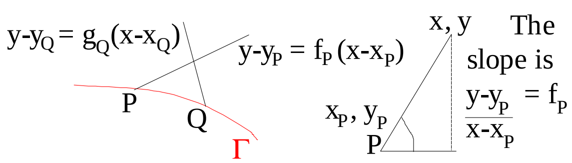
The **quadratic** has two *real distinct solutions* for dy/dx (the *characteristic* direction) if $b^2-4ac > 0$, so that the **discriminant** implies that we have a hyperbolic equation. It has *two equal solutions* if $b^2-4ac = 0$ (a parabolic equation); and it has *two complex conjugate roots* if $b^2-4ac < 0$ (an **elliptic** equation). Note: $dy/dx = b\pm\sqrt{(b^2-4ac)}/2a = b\pm i\sqrt{(4ac-b^2)}/2a$.

The Numerical Solution of Hyperbolic Equations by Characteristics



At any **point**, there are two characteristic directions given by the solution of the *quadratic* for dy/dx, and we shall denote them as f and g. We may use the *above theory* to solve initial value problems, where the unknown and the **first** derivatives ($\partial U/\partial x$ and $\partial U/\partial y$) are given on an *initial curve*, Γ. (i.e. we are given a **Cauchy** boundary condition). **Suppose** that $P(x_P, y_P)$ and $Q(x_Q, y_Q)$ are *two neighbouring points* on Γ, and that f_P, g_P; f_Q and g_Q are the *values of the characteristic slopes* at P and Q. We want to obtain the **solution** for U at a point R *off* the initial line. Let R be the **intersection** of the straight line through P (with slope f_P) with the *straight line through Q* (with slope g_Q). We are **assuming** here that if P and Q are close enough, then the straight lines will be a *fairly good approximation* to the true characteristic curves.





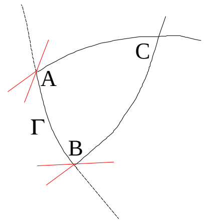
There are **several** steps to this procedure. (1) Find the *co-ordinate* of R, (x_R, y_R). For *intersection*, $y_R-y_P = f_P(x_R-x_P)$; and $y_R-y_Q = g_Q(x_R-x_Q)$, which gives us a *set of simultaneous equations* for x_R and y_R. (2) Use the **equations** for $\partial U/\partial x$ and $\partial U/\partial y$ along *both characteristics* to find $\partial U/\partial x$ and $\partial U/\partial y$ at R. (On PR) $a\,d/dx(\partial U/\partial x)f_P + c\,d/dx(\partial U/\partial y) + ef_P = 0$. (On QR) $a\,d/dx(\partial U/\partial x)g_Q + c\,d/dx(\partial U/\partial y) + eg_Q = 0$. The *simplest way to difference these* is along PR: $a_Pf_P[(\partial U/\partial x)_R-(\partial U/\partial x)_P] + c_P[(\partial U/\partial y)_R-(\partial U/\partial y)_P] + e_P[y_R-y_P] = 0$. (The Red bits are *unknown*). *Along QR*, $a_Qg_Q[(\partial U/\partial x)_R-(\partial U/\partial x)_Q] + c_Q[(\partial U/\partial y)_R-(\partial U/\partial y)_Q] + e_Q[y_R-y_Q] = 0$.

(**Note**: If we had some *approximate values available* at R, then it would be more **accurate** to replace a by $a_P+a_R/2$, etc. This *would* be available in an **iterating** scheme). By solving the above *two simultaneous equations*, we obtain approximations for $(\partial U/\partial x)_R$ and $(\partial U/\partial y)_R$. (3) Now find U at R, by *using* $dU = \partial U/\partial x\,dx + \partial U/\partial y\,dy$.

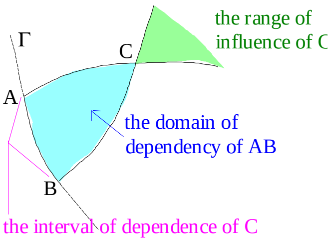
Along PR, $U_R-U_P = \frac{1}{2}[(\partial U/\partial x)_R+(\partial U/\partial x)_P](x_R-x_P) + \frac{1}{2}[(\partial U/\partial y)_R+(\partial U/\partial y)_P](y_R-y_P)$. **Along QR,** $U_R-U_Q = \frac{1}{2}[(\partial U/\partial x)_R+(\partial U/\partial x)_Q](x_R-x_Q) + \frac{1}{2}[(\partial U/\partial y)_R+(\partial U/\partial y)_Q](y_R-y_Q)$. These two *equations* will give two values for U_R; and these should be in **good** agreement if the *approximations* were justified.

22nd November 2000

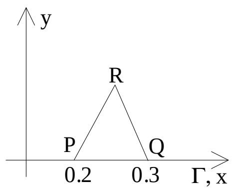
Features of Hyperbolic Equations



(1) To **extend** the solution, we need initial values on a curve Γ which is not a *characteristic*. (2) In the figure, the initial curve Γ is AB. The region in which the solution may be found in is **bounded** by the characteristics AC and BC. Region ABC is called the *domain of dependency of AB*. Values of the solution within ABC are not affected at all by the values on the **continuation** of AB; and this is in contrast with *elliptic equations* — where **all** boundary curves affect the solution.



(3) The **solution** at C is determined from the solution within ABC. AB is called the *interval of dependence of C*. (4) Values at C influence the solution ahead of C in a region determined by the **characteristics** through C. This region is called the *range of influence* of C.



Example: Solve $\partial^2U/\partial x^2 - U^2\partial^2U/\partial y^2 = 0$ at the *first grid point* between x = 0.2 and x = 0.3 (with y > 0), **given** that on y = 0, $U = 0.2+5x^2$, and $\partial U/\partial y = 3x$. A The *characteristics* are given by $a\,\partial^2U/\partial x^2+b\,\partial^2U/\partial x\partial y+c\,\partial^2U/\partial y^2+e = 0$; $dy/dx = {}^{b\pm\sqrt{(b^2-4ac)}}/_{2a}$. *Here*, a = 1, b = 0, and c = $-U^2$, so that $dy/dx = {}^{\pm\sqrt{(4U^2)}}/_2 = \pm U$. Alternatively, we go *directly* to $(dy/dx)^2-U^2 = 0$; $dy/dx = \pm U$.

Note that the *characteristics* depend on the solution U, and cannot be drawn in **advance**. **Solution:** Take P = (0.2, 0); take Q = (0.3, 0); take $f_P = (dy/dx)_P = U_P$; and take $g_Q = (dy/dx)_Q = -U_Q$. Now we know that $U = 0.2+5x^2$ on y = 0, so that $f_P = 0.2+5(0.2)^2 = 0.4$, and that $g_Q = -(0.2+5(0.3)^2) = -0.65$.

(1) Find *the co-ordinates of R*. At P, $y_R-y_P = 0.4(x_R-x_P)$; and *therefore* $y_R = 0.4(x_R-0.2)$ (---(1)). **Similarly**, at Q, $y_R = -0.65(x_R-0.3)$ (---(2)). Solving (1) and (2) *simultaneously* gives $x_R = 0.26190$. and $y_R = 0.02476$. This is our first approximation for R. (2) Find $\partial U/\partial x$ and $\partial U/\partial y$ at R. Method: we use the *finite difference formula* as given in the theory, i.e. $a\,d(\partial U/\partial x)\,dy/dx + c\,d(\partial U/\partial y) + e\,dy = 0$. Now a = 1; c = $-U^2$; and e = 0, which is used at P and Q. On y = 0, we have $U = 0.2+5x^2$, so that $\partial U/\partial x = 10x$. At this *stage* in proceedings, we *know* U, $\partial U/\partial x$, and $\partial U/\partial y$ on y = 0. Therefore, we have $(\partial U/\partial x)_P = 2$; $(\partial U/\partial y)_P = 0.6$; $(\partial U/\partial x)_Q = 3$; and $(\partial U/\partial y)_Q = 0.9$.

Along PR, $(0.4)[(\partial U/\partial x)_R-2] - 0.16[(\partial U/\partial y)_R-0.6] = 0$; and along QR, $(-0.65)[(\partial U/\partial x)_R-3] - 0.4225[(\partial U/\partial y)_R-0.9] = 0$. The *solution* of these two simultaneous equations for $(\partial U/\partial x)_R$ and $(\partial U/\partial y)_R$ gives $(\partial U/\partial x)_R = 2.45524$, and $(\partial U/\partial y)_R = 1.73810$. (3) Find *U at R* (at both P and Q) by using $dU = \partial U/\partial x\,dx + \partial U/\partial y\,dy$. At P, $U_R-0.4 = \frac{1}{2}(2+2.45524)(0.0619) + \frac{1}{2}(1.73810+0.6)(0.024)$; $U_R = 0.56684$. At Q, $U_R = 0.55237$. These two values agree **only to 1 significant figure**, so we should try to *improve* the solution.

Second Approximation. We have $f_R = U_R$, and $g_R = -U_R$. If we *choose* $U_R = 0.56684$ (we could well have just **chosen** U_R to be 0.55237), *then* $f_R = 0.56684$, and $g_R = -0.56684$. **Stage 1**: find a new approximation for the *position* of R. Along PR, $y_R = \frac{1}{2}(0.4+0.56684)(x_R-0.2)$. (f_P+f_R). Along QR, $y_R = -\frac{1}{2}(0.65+0.56684)(x_R-0.3)$. (g_Q+g_R). Solving this pair of *simultaneous equations* yields $x_R = 0.25572$, and $y_R = 0.02694$.

Stage 2: Find $\partial U/\partial x$ and $\partial U/\partial y$ at R. To do this, use *averages* for f, g and c. Along PR, $\frac{1}{2}(0.4+0.56684)(\partial U/\partial x|_R - 2.0) - \frac{1}{2}(0.16+0.32131)(\partial U/\partial y|_R - 0.6) = 0$. $(f_P+f_R$, *then* $c_P+c_R)$. Along QR, $\frac{1}{2}(0.65+0.56684)(\partial U/\partial x|_R - 3.0) - \frac{1}{2}(0.4225+0.32131)(\partial U/\partial y|_R - 0.9) = 0$. $(g_Q+g_R$, *then* $c_Q+c_R)$. Solving *simultaneously* gives $\partial U/\partial x|_R = 2.53117$, and $\partial U/\partial y|_R = 1.6670$. **Stage 3**: Find U at R. Now $U_R = 0.55677$ (*along* PR), and $U_R = 0.55673$ (*along* QR). This time, the answers **agree** to 4 d.p.

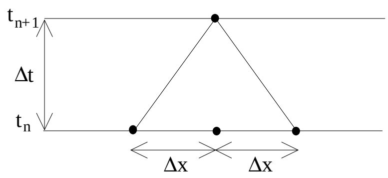
The Finite Difference Solution of Hyperbolic Equations

Consider the *wave equation* $\partial^2U/\partial t^2 = c^2\partial^2U/\partial x^2$. Discretising using *central differences* gives $u_{j,n+1}-2u_{j,n}+u_{j,n-1}/\Delta t^2 = c^2(u_{j-1,n}-2u_{j,n}+u_{j+1,n}/\Delta x^2)$, **or** $u_{j,n+1} = r^2(u_{j+1,n}+u_{j-1,n}) + 2(1-r^2)u_{j,n} - u_{j,n-1}$, where $r = c\Delta t/\Delta x$. **Notes**: (i) There is a *different definition* of r as compared with the parabolic case, where r was $\Delta t/\Delta x^2$; and (ii) $r = c\Delta t/\Delta x$ is called the *Courant number*.

Stability Analysis. Using the *Fourier* (von Neumann) method, substitute $u_{j,n}$ by a Fourier mode (of error $\hat{\epsilon}_n\exp(ikx_j)$) in the *difference scheme*. For **stability**, we require $\hat{\epsilon}_{n+1} = g\hat{\epsilon}_n$ where $|g| \leq 1$. *Writing* $\hat{\epsilon}_{n+1} = g\hat{\epsilon}_n = g^2\hat{\epsilon}_{n-1}$, etc., we obtain $g^2-2g(1-2r^2\sin^2(k\Delta x/2))+1 = 0$, or $g^2-2Ag+1 = 0$, *where* $A = 1-2r^2\sin^2(k\Delta x/2)$. *Note* that $A \leq 1$.

The **roots** of the quadratic for g are as follows: $g_1 = A+\sqrt{(A^2-1)}$; and $g_2 = A-\sqrt{(A^2-1)}$. (a) If $|A| > 1$ (i.e. $A \leq -1$), *then* $A-\sqrt{(A^2-1)} < -1$, so that $|g_2| > 1$, which leads to **instability**. (b) If $|A| \leq 1$, the roots are *complex conjugates:* $g_1 = A+i\sqrt{(1-A^2)}$, and $g_2 = A-i\sqrt{(1-A^2)}$; so that $|g_1| = |g_2| = 1$ gives *stability*. Therefore, we **must** have $|A| \leq 1$, i.e. $|1-2r^2\sin^2(k\Delta x/2)| \leq 1$, i.e. $-1 \leq 1-2r^2\sin^2(k\Delta x/2) \leq 1$.

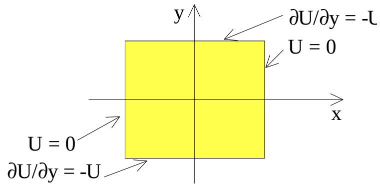
The RHS inequality is **always** satisfied. The LHS inequality gives $2r^2\sin^2(k\Delta x/2) \leq 2$; $r^2\sin^2(k\Delta x/2) \leq 1$. This must be *true* for **all** Fourier modes k, **so** that $r^2 \leq 1$, or that $r \leq 1$. *Therefore*, $c\Delta t/\Delta x \leq 1$, or $\Delta t \leq \Delta x/c$, or $c \leq \Delta x/\Delta t$. This condition is called the *Courant-Friedricks-Leury* condition (the **CFL** condition). It restricts the *timestep* which may be used with a particular spatial mesh; and may be thought of as shown *in the diagram*.



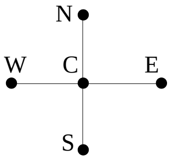
Elliptic Equations

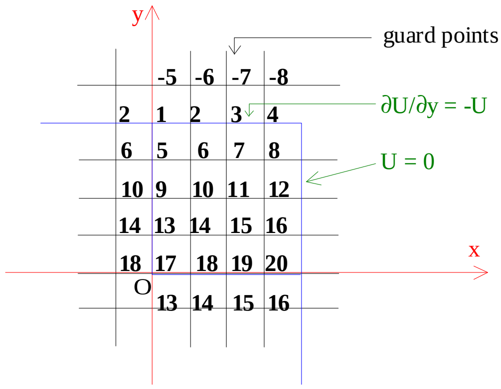
Elliptic equations arise from steady state problems such as *electrostatics*, *magnetostatics*, *perfect fluid flow*, *steady heat flow*, etc. 1-D heat equation: $\partial U/\partial t = \partial^2U/\partial x^2$. 2-D heat equation: $\partial U/\partial t = \partial^2U/\partial x^2+\partial^2U/\partial y^2$. The prototype equation is *Poisson's equation*, $\partial^2U/\partial x^2+\partial^2U/\partial y^2 = f(x,y)$. If $f(x,y) = 0$, the equation is called *Laplace's* equation. **Boundary** conditions: either U or its *derivative* (or a *combination*) is specified on the boundary.

Example: The following equation models the *steady state temperature distribution* in electrical windings, with internal generation of heat, and heat loss by radiation to the surrounding air: $\partial^2U/\partial x^2 + 3\,\partial^2U/\partial y^2 = -16$, defined on the *square* $-1 \le x \le 1$, $-1 \le y \le 1$; with boundary conditions $U = 0$ on $x = \pm 1$; and $\partial U/\partial y = -U$ on $y = \pm 1$.



Note that the problem is symmetric w.r.t. both the **x-axis** and the **y-axis**, so that we need only solve in *one* quadrant, which we shall choose to be the +ve quadrant. We use the **shown** mesh, and we discretise the PDE by using *central difference formulae*. Using the points either side in the usual way (see the diagram on the <u>left</u>), we obtain $(U_W-2U_C+U_E/\Delta x^2) + 3(U_N-2U_C+U_S/\Delta y^2) = f_C$.





For our *problem*, choose $\Delta x = \Delta y = ¼$. so that the **difference** scheme is given by $(U_W-2U_C+U_E/(¼)^2) + 3(U_N-2U_C+U_S/(¼)^2) = -16$, which *gives* $3U_N + U_W - 8U_C + U_E + 3U_S = -1$. For *example*, at **mesh point 10**, $3U_6+U_9-8U_{10}+U_{11}+3U_{14} = -1$; at **mesh point 12**, $3U_8+U_{11}-8U_{12}+0+3U_{16} = -1$; and at **mesh point 9**, $3U_5+U_{10}-8U_9+U_{10}+3U_{13} = -1$.

For the *derivative boundary condition* at $y = 1$, introduce **guard** points outside the domain, and approximate $\partial U/\partial y|_c$ by $U_N-U_S/2\Delta y$. So, for example, at *mesh point 1*, $\partial U/\partial y|_c \approx U_{-5}-U_5/2(¼) = U_{-5}-U_5/½$. The *difference equation* for the PDE at mesh point 1 is given by $3U_{-5}+U_2-8U_1+U_2+3U_5 = -1$. The *derivative boundary condition* at mesh point 1 is given by $U_{-5}-U_5/½ = -U_1$. **Eliminating** U_{-5} *between* these two equations, we obtain $3(U_5-(U_1/2)) + U_2 - 8U_1 + U_2 + 3U_5 = -1$, which *gives* $-8U_1 - {}^3/_2U_1 + 2U_2 + 6U_5 = -1$.

30th November 2000

The set of *20 equations in the 20 unknowns* is as follows: <u>**Point 1:**</u> $-8U_1-{}^3/_2U_1+2U_2+6U_5 = -1$; <u>**Point 2:**</u> $U_1-8U_2-{}^3/_2U_2+U_3+6U_6 = -1$; <u>**Point 3:**</u> $U_2-8U_3-{}^3/_2U_3+U_4+6U_7 = -1$; <u>**Point 4:**</u> $U_3-8U_4-{}^3/_2U_4+6U_8 = -1$; <u>**Point 5:**</u> $3U_1-8U_5+2U_6+3U_9 = -1$; <u>**Point 6:**</u> $3U_2+U_5-8U_6+U_7+3U_{10} = -1$ *(N+W+C+E+S = -1)*; <u>**Point 7:**</u> $3U_3+U_6-8U_7+U_8+3U_{11} = -1$; <u>**Point 8:**</u> $3U_4+U_7-8U_8+3U_{12} = -1$.

<u>**Points 9 to 12**</u>: *Add 4 onto node numbers 5 to 8*. <u>**Points 13 to 16**</u>: Add another 4 onto the *node/mesh point numbers*. <u>**Point 17:**</u> $6U_{13}-8U_{17}+2U_{18} = -1$ *(N and S + C + E and W)*; <u>**Point 18:**</u> $6U_{14}+U_{17}-8U_{18}+U_{19} = -1$ *(N and S + W + C + E)*; <u>**Point 19:**</u> $6U_{15}+U_{18}-8U_{19}+U_{20} = -1$; and <u>**Point 20:**</u> $6U_{16}+U_{19}-8U_{20} = -1$.

$$A = \begin{pmatrix} B-\frac{3}{2}I & 6I & 0 & 0 & 0 \\ 3I & B & 3I & 0 & 0 \\ 0 & 3I & B & 3I & 0 \\ 0 & 0 & 3I & B & 3I \\ 0 & 0 & 0 & 6I & B \end{pmatrix}, \text{ where } B = \begin{pmatrix} -8 & 2 & 0 & 0 \\ 1 & -8 & 1 & 0 \\ 0 & 1 & -8 & 1 \\ 0 & 0 & 1 & -8 \end{pmatrix}, \text{ and } I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This *matrix equation* may be written as $A\underline{u} = \underline{b}$, where $\underline{u} = (U_1U_2U_3....)^T$ is the *vector* of the **unknown** values at the 20 mesh points, and $\underline{b} = -(1\ 1\ 1\ ....\ 1)^T$ is a *vector of ones*. The matrix A may be written in **block** form as shown above.

Exercises 5

Let the PDE ${}^{\partial U}/_{\partial t}$ - ${}^{\partial^2 U}/_{\partial x^2}$ = 0 be **approximated** at the point ($j\Delta x$, $n\Delta t$) by the *difference equation* $\theta[u_{j,n+1}-u_{j,n-1}/2\Delta t] + (1-\theta)[u_{j,n}-u_{j,n-1}/\Delta t] - (u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2) = 0$. Show that the *local truncation error* at this point is given by $-½\Delta t(1-\theta)({}^{\partial^2 U}/_{\partial t^2})_{j,n} - {}^1/_{12}\Delta x^2(\partial^4 U/\partial x^4)_{j,n} + O(\Delta t^2) + O(\Delta x^4)$. For *what* value of θ will this error **reduce** to one of order Δt^2 and Δx^4? **With** this value of θ, and the choice $\Delta x^2 = 6\Delta t$, show by the *Fourier method* that the scheme is **stable**.

A: $u_{j,n\pm1} = u_{j,n} \pm ({}^{\partial U}/_{\partial t})_{j,n}\Delta t + ½({}^{\partial^2 U}/_{\partial t^2})_{j,n}\Delta t^2 \pm ...$ **And** $u_{j\pm1,\,n} = u_{j,n} \pm ({}^{\partial U}/_{\partial x})_{j,n}\Delta x + ½({}^{\partial^2 U}/_{\partial x^2})_{j,n}\Delta x^2 \pm ...$ Substituting *into the difference equation* gives ${}^{\theta}/_{2\Delta t}[u_{j,n} + ({}^{\partial U}/_{\partial t})_{j,n}\Delta t + ½({}^{\partial^2 U}/_{\partial t^2})\Delta t^2 + ... - (u_{j,n} - ({}^{\partial U}/_{\partial t})_{j,n}\Delta t + - ...)] + {}^{(1-\theta)}/_{\Delta t}[u_{j,n} - (u_{j,n} - ({}^{\partial U}/_{\partial t})_{j,n}\Delta t + ½({}^{\partial^2 U}/_{\partial t^2})\Delta t^2 - ...] - {}^1/_{\Delta x^2}[2(½({}^{\partial^2 U}/_{\partial x^2})_{j,n}\Delta x^2 + {}^1/_{4!}(\partial^4 U/\partial x^4)\Delta x^4 + ...)] = 0$. *Simplifying*, we get $(1-\theta)[-½({}^{\partial^2 U}/_{\partial t^2})\Delta t - {}^1/_{4!}(\partial^4 U/\partial t^4)\Delta t^3 + ...] + [({}^{\partial U}/_{\partial t}) + {}^1/_{3!}({}^{\partial^3 U}/_{\partial t^3})\Delta t^2 + ...] - 2[½({}^{\partial^2 U}/_{\partial x^2}) + {}^1/_{4!}(\partial^4 U/\partial x^4)\Delta x^2 + ...] = 0$ (---(1)).

To *get* the l.t.e., do (1) - PDE, **giving** $(1-\theta)[-½({}^{\partial^2 U}/_{\partial t^2})\Delta t - O(\Delta t^3)] + ({}^1/_{3!}({}^{\partial^3 U}/_{\partial t^3})\Delta t^2 + ...) - 2({}^1/_{4!}(\partial^4 U/\partial x^4)\Delta x^2 + O(\Delta x^4)) = 0$; $(1-\theta)[-½({}^{\partial^2 U}/_{\partial t^2})\Delta t] + O(\Delta t^2) - {}^1/_{12}\Delta x^2(\partial^4 U/\partial x^4) + O(\Delta x^4) = 0$. **<u>QED</u>**. The next step is to *choose a value of* θ that will <u>eliminate</u> two of the above terms. The choice θ = 1 eliminates the **first** term, but we **also** need a term of the form ${}^1/_{12}\Delta x^2(\partial^4 U/\partial x^4)$. *Try* $\theta = 1 + {}^1/_6{}^{\Delta x^2}/_{\Delta t}$ to start with. **Therefore,** $(1-1-{}^{\Delta x^2}/_{6\Delta t})[-½({}^{\partial^2 U}/_{\partial t^2})\Delta t] + O(\Delta t^2) - {}^1/_{12}\Delta x^2(\partial^4 U/\partial x^4) + O(\Delta x^4) = 0$; ${}^1/_{12}\Delta x^2({}^{\partial^2 U}/_{\partial t^2}) - {}^1/_{12}\Delta x^2(\partial^4 U/\partial x^4) + O(\Delta t^2) + O(\Delta x^4) = 0$.

But ${}^{\partial U}/_{\partial t} = {}^{\partial^2 U}/_{\partial x^2}$, so that ${}^{\partial^2 U}/_{\partial t^2} = \partial^4 U/\partial x^4$. Therefore, *substituting in the above* gives just $O(\Delta t^2) + O(\Delta x^4)$ as required. So $\theta = 1 + {}^{\Delta x^2}/_{6\Delta t}$ is our choice. Further, if $\Delta x^2 = 6\Delta t$, we get $\theta = 1 + {}^{6\Delta t}/_{6\Delta t} = 2$. *Substitute* in for θ, and the scheme becomes $u_{j,n+1}-u_{j,n}/\Delta t = u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2$ (---(2)).

Now show that the scheme is **stable** by using the *Fourier* method. Let $\epsilon_{j,n} = \hat{\epsilon}_n\exp(ikx_j)$. **Substituting** this into (2) gives the following expression: $({}^1/_{\Delta t})[\hat{\epsilon}_{n+1}\exp(ikx_j) - \hat{\epsilon}_n\exp(ikx_j) = ({}^1/_{\Delta x^2})[\hat{\epsilon}_n\exp(ik(x_j+\Delta x)) - 2\hat{\epsilon}_n\exp(ikx_j) + \hat{\epsilon}_n\exp(ik(x_j-\Delta x))]$; $\hat{\epsilon}_{n+1}\exp(ikx_j) = \hat{\epsilon}_n\exp(ikx_j) + {}^{\Delta t}/_{\Delta x^2}(...)$. **Cancelling** the $\exp(ikx_j)$'s, we obtain $\hat{\epsilon}_{n+1} = \hat{\epsilon}_n + {}^{\Delta t}/_{\Delta x^2}\hat{\epsilon}_n(\exp(-ik\Delta x) + \exp(ik\Delta x) - 2)$; $\hat{\epsilon}_{n+1} = \hat{\epsilon}_n + {}^{\Delta t}/_{\Delta x^2}\hat{\epsilon}_n(2\cos(k\Delta x) - 2)$.

Now $\cos 2\theta = 1-2\sin^2\theta$, so that $\hat{\epsilon}_{n+1} = \hat{\epsilon}_n + {}^{\Delta t}/_{\Delta x^2}\hat{\epsilon}_n(2-4\sin^2({}^{k\Delta x}/_2)-2)$. Manipulating, we get the following: $\hat{\epsilon}_{n+1} = \hat{\epsilon}_n - 4{}^{\Delta t}/_{\Delta x^2}\hat{\epsilon}_n\sin^2({}^{k\Delta x}/_2)$. **Now** $\hat{\epsilon}_{n+1} = g\hat{\epsilon}_n$, so that $g\hat{\epsilon}_n = \hat{\epsilon}_n - 4{}^{\Delta t}/_{\Delta x^2}\hat{\epsilon}_n\sin^2({}^{k\Delta x}/_2)$; $g = 1-4{}^{\Delta t}/_{\Delta x^2}\sin^2({}^{k\Delta x}/_2)$. **But** $\Delta x^2 = 6\Delta t$, so that $g = 1-{}^{4\Delta t}/_{6\Delta t}\sin^2({}^{k\Delta x}/_2)$; $g = 1-{}^2/_3\sin^2({}^{k\Delta x}/_2)$. Now $0 \le \sin^2({}^{k\Delta x}/_2) \le 1$; $0 \le {}^2/_3\sin^2({}^{k\Delta x}/_2) \le {}^2/_3$; $-{}^2/_3 \le -{}^2/_3\sin^2({}^{k\Delta x}/_2) \le 0$; ${}^1/_3 \le 1-{}^2/_3\sin^2({}^{k\Delta x}/_2) \le 1$; ${}^1/_3 \le g \le 1$. Therefore, the scheme is *stable* for this value of θ — and the **choice** $\Delta x^2 = 6\Delta t$.

Q: Using *forward difference* for time, and *central differences* for space, construct a finite difference scheme to solve the PDE ${}^{\partial U}/_{\partial t} = a{}^{\partial^2 U}/_{\partial x^2} + b{}^{\partial U}/_{\partial x} + cU$, where *a, b and c* are constants, and the **other** symbols have their usual meanings. Show, with the *usual* notation, that the <u>truncation error</u> of this scheme is $½\Delta t{}^{\partial^2 U}/_{\partial t^2} - a{}^{\Delta x^2}/_{12}(\partial^4 U/\partial x^4) - b{}^{\Delta x^2}/_6{}^{\partial^3 U}/_{\partial x^3} + O(\Delta t^2) + O(\Delta x^4)$. Now assume that b = c = 0. Show that the scheme is **fourth** order accurate with the *choice* $\Delta t = {}^{\Delta x^2}/_{6a}$. Show also that the scheme is *stable* with this choice of Δt.

A: This is similar to the *previous question*. The difference scheme we use is $u_{j,n+1}-u_{j,n}/\Delta t - a(u_{j-1,n}-2u_{j,n}+u_{j+1,n}/\Delta x^2) - b(u_{j+1,n}-u_{j-1,n}/2\Delta x) - cu_{j,n} = 0$. Use **Taylor expansions** to get the local truncation error $T_{j,n}$ at point (x_j,t_n). We get, *after substitution*, $T_{j,n} = ({}^{\partial U}/_{\partial t})_{j,n} + ({}^{\partial^2 U}/_{\partial t^2})_{j,n}{}^{\Delta t}/_2\ \ O(\Delta t^3) - a[({}^{\partial^2 U}/_{\partial x^2})_{j,n} + {}^1/_{12}(\partial^4 U/\partial x^4)_{j,n}\Delta x^2 + O(\Delta x^4)] - b[({}^{\partial U}/_{\partial x})_{j,n} + {}^1/_6({}^{\partial^3 U}/_{\partial x^3})_{j,n}\Delta x^2 + O(\Delta x^4)] - cU_{j,n}$.

Taking away the PDE gives $T_{j,n} = \frac{1}{2}\Delta t\,{}^{\partial^2 U}/_{\partial t^2} - a^{\Delta x^2}/_{12}(\partial^4 U/\partial x^4) - b^{\Delta x^2}/_6\,{}^{\partial^3 U}/_{\partial x^3} + O(\Delta t^2) + O(\Delta x^4)$. **QED**. If $b = c = 0$, then the PDE **becomes** ${}^{\partial U}/_{\partial t} = a^{\partial^2 U}/_{\partial x^2}$; the *difference equation* becomes $u_{j,n+1}-u_{j,n}/\Delta t - a(u_{j+1,n}-2u_{j,n}+u_{j+1,n}/\Delta x^2) = 0$; and the truncation error becomes $T_{j,n} = {}^{\Delta t}/_2({}^{\partial^2 U}/_{\partial t^2})_{j,n} - a^{\Delta x^2}/_{12}(\partial^4 U/\partial x^4)_{j,n} + O(\Delta t^2) + O(\Delta x^4)$.

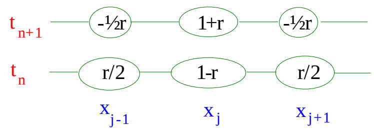
From the (*new*) PDE, ${}^{\partial^2 U}/_{\partial t^2} = a^{d}/_{dt}({}^{\partial^2 U}/_{\partial x^2}) = a^{\partial^2}/_{\partial x^2}({}^{\partial U}/_{\partial t}) = a^{\partial^2}/_{\partial x^2}a({}^{\partial^2 U}/_{\partial x^2}) = a^2(\partial^4 U/\partial x^4)$. *Substituting* for ${}^{\partial^2 U}/_{\partial t^2}$ gives $T_{j,n} = {}^{\Delta t}/_2 a^2(\partial^4 U/\partial x^4)_{j,n} - a^{\Delta x^2}/_{12}(\partial^4 U/\partial x^4)_{j,n} + O(\Delta t^2) + O(\Delta x^4)$; $T_{j,n} = ({}^{a^2}/_2\Delta t - a^{\Delta x^2}/_{12})(\partial^4 U/\partial x^4)_{j,n} + O(\Delta t^2) + O(\Delta x^4)$. The *first term* vanishes if ${}^{a^2}/_2\Delta t = a^{\Delta x^2}/_{12}$, i.e. if $\Delta t = {}^{\Delta x^2}/_{6a}$. In this case, $T_{j,n} = O(\Delta t^2) + O(\Delta x^4)$ as *required*.

To **investigate** stability, use the von Neumann method, expressing the error as a *Fourier series*. Because of linearity, we can study the propagation of a single Fourier mode, $\hat{\epsilon}_n\exp(ikx_j)$, where k is the *node number* — and this is propagated by the **same** difference equation as $u_{j,n}$. As before, let $\epsilon_{j,n} = \hat{\epsilon}_n\exp(ikx_j)$, and **substitute and manipulate** to give $g = 1-4a^{\Delta t}/_{\Delta x^2}\sin^2({}^{k\Delta x}/_2)$. Now if $\Delta t = {}^{\Delta x^2}/_{6a}$, then $g = 1-{}^2/_3\sin^2({}^{k\Delta x}/_2)$. Therefore, *as before*, ${}^1/_3 \le g \le 1$, and the scheme is **stable** for $\Delta t = {}^{\Delta x^2}/_{6a}$.

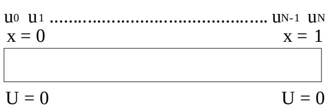
Q: The **Crank-Nicolson** implicit scheme is used to obtain an *approximate solution* to the PDE ${}^{\partial^2 U}/_{\partial x^2} = {}^{\partial U}/_{\partial t}$, subject to the boundary condition $U = 0$ at $x = 0$ and $x = 1$. Show that the scheme can be written in *matrix form* as $u_{n+1} = Gu_n$, where $G = (2I-rT)^{-1}(2I+rT)$, in which T is a *tridiagonal* matrix; I is the **identity** matrix; and $r = \Delta t/\Delta x^2$. Write down the *entries* in T.

Use the **matrix** method to show that the scheme is *unconditionally stable*. You may **assume** that the eigenvalues λ_s of the N×N tridiagonal matrix $A = [a_{i,j}]$, *where* $a_{i,i} = -2$; $a_{i-1,i} = a_{i+1,i} = 1$; and *otherwise* $a_{i,j} = 0$, are given by $\lambda_s = -4\sin^2({}^{s\pi}/_{2(N+1)})$, for s = 1,...,N. You may also *assume* that G is **real** and **symmetric**, so that $\|G\| = \rho(G)$, and that the *eigenvalues* of $[f(A)]^{-1}g(A)$, where f(A) and g(A) are *polynomials* in the matrix A, are **given** by $[f(\lambda)]^{-1}g(\lambda)$, where λ is an *eigenvalue* of A.

A: The Crank-Nicolson implicit scheme comes from **applying** $\theta = \frac{1}{2}$ to $u_{j,n+1}-u_{j,n}/\Delta t = (1-\theta)[u_{j+1,n}-2u_{j,n}+u_{j-1,n}/\Delta x^2] + \theta[u_{j+1,n+1}-2u_{j,n+1}+u_{j-1,n+1}/\Delta x^2]$, where $0 \le \theta \le 1$. So we have $-r\frac{1}{2}u_{j-1,n+1}+(1+r)u_{j,n+1}-r\frac{1}{2}u_{j+1,n+1} = r\frac{1}{2}u_{j-1,n}+(1-r)u_{j,n}+r\frac{1}{2}u_{j+1,n}$. The *computational molecule / stencil* is as shown on the right, so that in the implicit scheme, **three** unknowns at time t_{n+1} are related to three values at **time** t_n.



We obtain a matrix equation to *solve* at each timestep. The form of the matrix equation is $A\underline{u}_{n+1} = B\underline{u}_n+\underline{b}_n$, where $\underline{u}_{n+1}$ and $\underline{u}_n$ are the vectors of the values of the *dependent variable* at the mesh points; A and B are given by the **difference** scheme; and $\underline{b}_n$ gives the *boundary conditions*. In this question, the **boundary conditions** are *zero*, so that $\underline{b}_n = 0$.



The *matrix equation* looks like as shown on the right. As you can see, $\underline{A}$ and $\underline{B}$ are *tridiagonal* matrices. Multiplying **both** sides by 2, and taking it inside $\underline{A}$ and $\underline{B}$ respectively, we

$$\begin{pmatrix} (1+r) & -r/2 & & & 0 \\ -r/2 & \ddots & \ddots & & \\ & -r/2 & (1+r) & -r/2 & \\ & & \ddots & \ddots & -r/2 \\ 0 & & & -r/2 & (1+r) \end{pmatrix} \begin{pmatrix} u_1 \\ .. \\ u_{j-1} \\ u_j \\ u_{j+1} \\ .. \\ u_{N-1} \end{pmatrix}_{n+1} = \begin{pmatrix} (1-r) & r/2 & & & 0 \\ r/2 & \ddots & \ddots & & \\ & r/2 & (1-r) & r/2 & \\ & & \ddots & \ddots & r/2 \\ 0 & & & r/2 & (1-r) \end{pmatrix} \begin{pmatrix} u_1 \\ .. \\ u_{j-1} \\ u_j \\ u_{j+1} \\ .. \\ u_{N-1} \end{pmatrix}_{n} + \overset{(\underline{b}_n)}{\underline{0}}$$

$$\underline{A} \qquad \underline{u}_{n+1} = \underline{B} \qquad \underline{u}_n + \underline{b}_n$$

$$\underline{T} = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & \cdots & \cdots & \cdots \\ & & & 1 & -2 & 1 \end{bmatrix}$$

are then able to write $\underline{A} = (2\underline{I}-r\underline{T})$, and $\underline{B} = (2\underline{I}+r\underline{T})$, where $\underline{T}$ is *as shown on the left*. Therefore, $\underline{u}_{n+1} = \underline{A}^{-1}\underline{B}\underline{u}_n$. **QED**.

Let us now use the *matrix method* to show that the scheme is unconditionally stable. We assume that the *eigenvalues* of $\underline{T}$ are $\lambda_s = -4\sin^2({}^{s\pi}/_{2(N+1)})$, with s = 1,..,N. Let $f(\underline{T}) = 2\underline{I}-\underline{T}$, and let $g(\underline{T}) = 2\underline{I}+r\underline{T}$. For the error to remain **bounded**, i.e. that the scheme is *unconditionally stable*, we must have $\rho(G) \le 1$, where $||G|| = \rho(G)$, the *spectral radius* of G. In other words, the **largest magnitude** of the largest eigenvalue of G must have *modulus* ≤ 1.

So we *want* $|[f(\lambda)]^{-1}g(\lambda)| \le 1$; $|[2\underline{I}-r\lambda_s]^{-1}(2\underline{I}+r\lambda_s)| \le 1$; $|2+r\lambda_s/2-r\lambda_s| \le 1$; $|{}^{2-4\sin^2(s\pi/2(N+1))}/_{2+4\sin^2(s\pi/2(N+1))}| \le 1$; or $-1 \le ... \le 1$. The *RHS* inequality **gives** $2-4\sin^2({}^{s\pi}/_{2(N+1)}) \le 2+4\sin^2({}^{s\pi}/_{2(N+1)})$; $-\sin^2({}^{s\pi}/_{2(N+1)}) \le \sin^2({}^{s\pi}/_{2(N+1)})$, which is **always true** as $\sin^2 x \ge 0$ for all x. The *LHS* inequality gives $-2-4\sin^2({}^{s\pi}/_{2(N+1)}) \le 2-4\sin^2({}^{s\pi}/_{2(N+1)})$; $0 \le 4$, all right. Therefore, the *inequality holds*, and we can say that the scheme **is** unconditionally stable. **QED**.

Q: Show that if **central differences** are used throughout to construct a difference scheme to approximate the *one-dimensional heat equation*, ${}^{\partial U}/_{\partial t} = {}^{\partial^2 U}/_{\partial x^2}$, then the scheme is unconditionally unstable. Suppose now that the **space** derivative ${}^{\partial^2 U}/_{\partial x^2}$ is replaced by the expression $u_{j+1,n}-u_{j,n+1}-u_{j,n-1}+u_{j-1,n}/\Delta x^2$, where the notation used is *standard*. Show that the **modified** scheme is unconditionally stable.

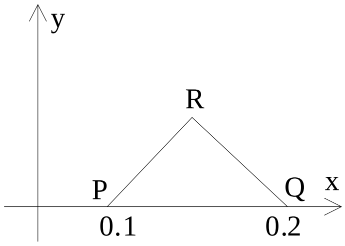
A: The *difference scheme* is $u_{j,n+1}-u_{j,n-1}/2\Delta t = u_{j-1,n}-2u_{j,n}+u_{j+1,n}/\Delta x^2$. We have **already shown** that this is unconditionally unstable (see *page 17, the Leap Frog Scheme*). In summary, use the **Fourier** method ($\epsilon_{j,n} = \hat{\epsilon}_n \exp(ikx_j)$); manipulate; let $\hat{\epsilon}_n = g\hat{\epsilon}_{n-1}$, etc.; substitute and manipulate to get $g^2 + 8({}^{\Delta t}/_{\Delta x^2})\sin^2({}^{k\Delta x}/_2)g - 1 = 0$; and then **factorise** to prove that $|g| \nleq 1$.

Similarly, we have already proved (see *page 18, Dufort Frankell*) that the scheme $u_{j,n+1}-u_{j,n-1}/2\Delta t = u_{j+1,n} - u_{j,n+1} - u_{j,n-1} + u_{j-1,n}/\Delta x^2$ is unconditionally **stable**. Here, we proceed as above, and get the *quadratic equation* $g^2(1-2r) + 4rg\cos(k\Delta x) + (1-2r)$, where $r = {}^{\Delta t}/_{\Delta x^2}$. Apply the *quadratic formula,* and we find **bounds** for the solutions — finding that $|g| \le 1$ *always*.

Exercises 6

Q: Prove that the *slopes* of the characteristics of the quasi-linear PDE $a{}^{\partial^2 U}/_{\partial x^2}+b{}^{\partial^2 U}/_{\partial x\partial y}+c{}^{\partial^2 U}/_{\partial y^2}+e = 0$ satisfy the *quadratic equation* $a({}^{dy}/_{dx})^2-b({}^{dy}/_{dx})+c = 0$, and that on the **characteristics,** the equation reduces to $a{}^{d}/_{dx}({}^{\partial U}/_{\partial x}){}^{dy}/_{dx} + c{}^{d}/_{dx}({}^{\partial U}/_{\partial y}) + e{}^{dy}/_{dx} = 0$.

Use the method of *characteristics* to find a first approximation to the solution of the PDE ${}^{\partial^2 U}/_{\partial x^2}-(U^2\cos^2 U){}^{\partial^2 U}/_{\partial y^2} = 0$ from P = (0.1, 0) and Q = (0.2, 0) to a **new** point R. The equation is to be *solved* in the domain y > 0, and the initial conditions are $U = 1+5x^2$ and ${}^{\partial U}/_{\partial y} = 10x$ on y = 0. Work to 3 significant figures.



A: The *first part* is theory given in the **lecture notes.** Now we have initial conditions $U = 1+5x^2$ on $y = 0$, (i); and $\partial U/\partial y = 10x$ on $y = 0$, (ii). *Differentiating* (i) w.r.t. x gives $\partial U/\partial x = 10x$ on $y = 0$, (iii). For this **particular** equation (as compared with the *general PDE considered by the theory),* $a = 1$; $b = 0$; $c = -U^2\cos^2U$; and $e = 0$, so that the **characteristics** satisfy $(dy/dx)^2 - U^2\cos^2U = 0$. *Therefore,* $dy/dx = \pm U\cos U$.

Along a characteristic, the PDE **reduces** to $\frac{d}{dx}(\partial U/\partial x)\,dy/dx - U^2\cos^2U\,\frac{d}{dx}(\partial U/\partial y) = 0$. Substituting *for* dy/dx, we obtain $\pm U\cos U\,d(\partial U/\partial x) + U^2\cos^2U\,d(\partial U/\partial y) = 0$, (iv), *along* a characteristic. To proceed, we require the values of **several** variables at P and Q. At P = (0.1, 0), (i) $\Rightarrow U = 1+5(0.1)^2 = 1.05$; (ii) $\Rightarrow \partial U/\partial y = 10(0.1) = 1$; and (iii) $\Rightarrow \partial U/\partial x = 10(0.1) = 1$.

Now $U = 1.05 \Rightarrow U\cos U = 1.05\cos(1.05) = (1.05)(0.4976) = 0.5224$; **and thus** $U^2\cos^2U = 0.27295$. At P, we *take the characteristic with +ve slope*, i.e. $dy/dx = +U\cos U = 0.5224$. Now at Q = (0.2, 0), (i) $\Rightarrow U = 1+5(0.2)^2 = 1.2$; (ii) $\Rightarrow \partial U/\partial y = 10(0.2) = 2$; and (iii) $\Rightarrow \partial U/\partial x = 10(0.2) = 2$. Now $U = 1.2 \Rightarrow U\cos U = 1.2\cos(1.2) = (1.2)(0.3624) = 0.4348$; and thus $U^2\cos^2U = 0.1891$. At Q, *we take the characteristic with -ve slope*, i.e. $dy/dx = -U\cos U = -0.4348$.

<u>Next Step</u>: get the <u>position of R</u>. Assume that the characteristics are *straight lines through P and Q* that meet at R, i.e. that $y_R-y_P/x_R-x_P = (dy/dx)_P$; and that $y_R-y_Q/x_R-x_Q = (dy/dx)_Q$. *Therefore,* $y_R-0/x_R-0.1 = 0.5224$, *and* $y_R-0/x_R-0.2 = -0.4348$. **So** $y_R - 0.5224x_R = -0.05224$, and $y_R+0.4348x_R = 0.08696$. *Subtracting, we obtain* $0.9572x_R = 0.1392$; $x_R = 0.145$.

Substituting back yields $y_R = 0.08696 - (0.4348)(0.145) = 0.0238$. So **R is at position** (0.145, 0.0238). *We now want the values* of $\partial U/\partial x$ and $\partial U/\partial y$ at R: use (iv) *at P and Q*. At P, $(0.5224)[(\partial U/\partial x)_R-(\partial U/\partial x)_P] + (0.27295)[(\partial U/\partial y)_R-(\partial U/\partial y)_P] = 0$. **Therefore**, $(0.5224)[(\partial U/\partial x)_R-1] + (0.27295)[(\partial U/\partial y)_R-1] = 0$; $0.5224(\partial U/\partial x)_R + 0.27295(\partial U/\partial y)_R = 0.79535$, (v).

At Q, $-(0.4348)[(\partial U/\partial x)_R-(\partial U/\partial x)_Q] + (0.1891)[(\partial U/\partial y)_R-(\partial U/\partial y)_Q] = 0$; $-0.4348[(\partial U/\partial x)_R-2.0] + 0.1891[(\partial U/\partial y)_R-2.0] = 0$; $0.4348(\partial U/\partial x)_R - 0.1891(\partial U/\partial y)_R = 0.4914$, (vi). We now *solve (v) and (vi) simultaneously* for $(\partial U/\partial x)_R$ and $(\partial U/\partial y)_R$. (v)/0.5224 $\Rightarrow (\partial U/\partial x)_R + 0.5225(\partial U/\partial y)_R = 1.5225$. (vi)/0.4348 $\Rightarrow (\partial U/\partial x)_R-0.4349(\partial U/\partial y)_R = 1.130$.

Subtracting gives $0.9574(\partial U/\partial y)_R = 0.3925$, so that $(\partial U/\partial y)_R = 0.410$, to 3 s.f.; and that $(\partial U/\partial x)_R = 1.130 - (0.4349)(0.410) = 0.952$, to 3 s.f. <u>Next Step</u>: get the <u>value of U at R</u> (use $dU = \partial U/\partial x\,dx + \partial U/\partial y\,dy$ at P and Q, and **compare the solutions** to ascertain the *accuracy* obtained). At P, $U_R-U_P = \frac{1}{2}[(\partial U/\partial x)_R+(\partial U/\partial x)_P](x_R-x_P) + \frac{1}{2}[(\partial U/\partial y)_R + (\partial U/\partial y)_P](y_R-y_P)$.

Therefore, $U_R-1.05 = (0.5)(0.952+1)(0.145-0.1) + (0.5)(0.410+1)(0.0238-0)$; $U_R = 1.05 + (0.5)(1.952)(0.045) + (0.5)(1.410)(0.0238)$; $U_R = 1.111$, (vii). *At Q*, $U_R-U_Q = \frac{1}{2}[(\partial U/\partial x)_R + (\partial U/\partial x)_Q](x_R-x_Q) + \frac{1}{2}[(\partial U/\partial y)_R + (\partial U/\partial y)_Q](y_R-y_Q)$; $U_R-1.2 = (0.5)(0.952+2.0)(0.145-0.2) + (0.5)(0.410+2.0)(0.0238-0)$; $U_R = 1.2 + (0.5)(2.952)(-0.055) + (0.5)(2.410)(0.0238)$; $U_R = 1.147$, (viii). *Comparing (vii) and (viii)* shows that there is 2 s.f. of **accuracy**.

Q: The PDE $\partial^2U/\partial x^2 - U^2 \partial^2U/\partial y^2 = 2y$ is to be **solved** in the domain $x > 0$ and $y > 1$, subject to the *initial condition* $U(x,1) = \partial U/\partial y(x,1) = x^2$, the **initial** curve being the line $y = 1$. Use the numerical method of *characteristics* to find a first approximation for U, at the first characteristic grid point R between P = (1, 1) and Q = (1.1, 1).

Verify that $U = x^2y$ is the *exact solution* of the problem, and compare your numerical answer with the exact solution. You may assume that, on the *characteristics*, the PDE, with the usual notation, reduces to $a\, d(\partial U/\partial x)\, dy/dx + c\, d(\partial U/\partial y) + e\, dy = 0$.

A: Compared to the *general PDE*, we have $a = 1$; $b = 0$; $c = -U^2$; and $e = -2y$. Thus the **characteristics** satisfy $(dy/dx)^2 - U^2 = 0$, and *therefore* $dy/dx = \pm U$. *Along a characteristic*, the PDE reduces to $d(\partial U/\partial x)\, dy/dx - U^2 d(\partial U/\partial y) - 2y\,dy = 0$. *Substituting* for dy/dx gives $\pm U d(\partial U/\partial x) - U^2 d(\partial U/\partial y) - 2y\,dy = 0$ (---(1)) along a *characteristic*.

To proceed, we require the **values** of several variables at P and Q. At P = (1,1), $U = \partial U/\partial y = 1^2 = 1$; and $\partial U/\partial x = 2x = 2$. At P, we take the *characteristic* with +ve slope, i.e. $dy/dx = U = 1$. Now at Q = (1.1, 1), $U = \partial U/\partial y = 1.1^2 = 1.21$; and $\partial U/\partial x = 2x = 2.2$. At Q, we take the *characteristic with -ve slope*, i.e. $dy/dx = -U = -1.1$.

Next Step: find the position of R. Assume that the characteristics are *straight lines through P and Q meeting at R*, i.e. that $y_R - y_P / x_R - x_P = (dy/dx)_P$; and that $y_R - y_Q / x_R - x_Q = (dy/dx)_Q$. *Therefore*, $y_R - 1 / x_R - 1 = 1$, and $y_R - 1 / x_R - 1.1 = -1.1$. **It follows that** $y_R - 1 = x_R - 1$, and that $y_R - 1 = -1.1x_R + 1.1^2$. Therefore, $y_R = x_R$, and this *implies that* $2.1x_R = 1.1^2 + 1$; $2.1x_R = 2.21$; $x_R = 1.05238 = y_R$. So R is *at position* (1.05238, 1.05238).

Next Step: obtain the **values** of $\partial U/\partial x$ and $\partial U/\partial y$ at R, by using (1) *at P and Q*. **At P,** $(1)[(\partial U/\partial x)_R - (\partial U/\partial x)_P] - (1^2)[(\partial U/\partial y)_R - (\partial U/\partial y)_P] - 2y(y_R - y_P) = 0$; $(\partial U/\partial x)_R - (\partial U/\partial y)_R - 1 - 2(1.05238-1) = 0$ (---(2)). **At Q,** $(1.21)[(\partial U/\partial x)_R - 2.2] - (1.21^2)[(\partial U/\partial y)_R - 1.21] - 2(1)(y_R - y_Q) = 0$; $1.21(\partial U/\partial x)_R - 1.4641(\partial U/\partial y)_R - 0.890439 - 2(1.05238-1) = 0$ (---(3)). Solve (2) and (3) *simultaneously* to get $(\partial U/\partial y)_R = 0.2641/0.21 + (-2/0.21 + 2/0.2541)(0.05238) = 1.171$; $(\partial U/\partial x)_R = 1 + 0.2641/0.21 + (2 - 2/0.21 + 2/0.2541)(0.05238) = 2.2758$.

Next Step: obtain the **value** of U at R, using $dU = \partial U/\partial x\, dx + \partial U/\partial y\, dy$ at P and Q; and *compute the solutions to ascertain the accuracy obtained*. At P, $U_R - U_P = ½[(\partial U/\partial x)_R + (\partial U/\partial x)_P](x_R - x_P) + ½[(\partial U/\partial y)_R + (\partial U/\partial y)_P](y_R - y_P)$; $U_R - 1 = ½[(\partial U/\partial x)_R + 2](1.05238-1) + ½[(\partial U/\partial y)_R + 1](1.05238-1)$; ....; $U_R =$ 1.1688. At Q, $U_R - U_Q = ½[(\partial U/\partial x)_R + (\partial U/\partial x)_Q](x_R - x_Q) + ½[(\partial U/\partial y)_R + (\partial U/\partial y)_Q](y_R - y_Q)$; $U_R - 1.21 = ½[(\partial U/\partial x)_R + 2.2](1.05238-1.1) + ½[(\partial U/\partial y)_R + 1.21](1.05238-1)$; ....; $U_R =$ 1.1658.

We now *verify* that $U = x^2y$ is the **exact** solution: $\partial U/\partial x = 2xy$; $\partial^2U/\partial x^2 = 2y$; $\partial U/\partial y = x^2$; and $\partial^2U/\partial y^2 = 0$. On *substitution*, we obtain $2y = 2y$, and therefore, $U = x^2y$ **is** a valid solution. At R, (1.05238, 1.05238), U_R (*Analytical*) $= x^2y = 1.05238^3 =$ 1.165515. As you can see, we have agreement in the **red** values good *to 2 d.p.*

Q: State, *without* proof, the **Lax Equivalence theorem**. Explain briefly why the theorem is of *central* importance in the application of the finite difference method to solutions of PDE's. Let the equation $\partial U/\partial t + \nu \partial U/\partial x = 0$ (where ν is a constant) be *approximated* by the difference scheme $u_{j,n+1} = (1-\theta-r)u_{j+1,n} + (\theta+r)u_{j-1,n}$, where $0 < \theta < 1$; $r = \nu\Delta t/\Delta x$; and the *rest* of the symbols have their **usual** meaning.

Prove that if the *time step* Δt is chosen to satisfy $\Delta t = \Delta x/c|\nu|$, where c is a *positive* constant, then the scheme is **consistent** if and only if $\theta = ½$. Show (by the Fourier method) that with the choice $\theta = ½$, the constant c must be chosen to be **greater than or equal to unity** for the scheme to be stable.

A: **Lax's Equivalence Theorem**: Given a *properly posed* linear initial value problem, and a linear finite-difference approximation to it that satisfies the **consistency** condition, stability is the necessary and sufficient condition for convergence. The theorem is important because in order to prove convergence, which is of *vital importance*, "all" that we need to prove is that the scheme is **consistent** and **stable**.

Note: a problem is *properly posed* if (i) the solution is **unique** when it exists; (ii) the solution depends **continuously** on the initial data; and (iii) a solution always exists for initial data that is **arbitrarily close** to initial data for which no solution exists. Justification: use *Taylor series expansions*, and apply to the difference scheme.

$u_{j,n+1} = (1-\theta-r)u_{j+1,n} + (\theta+r)u_{j-1,n}$; $u_{j,n+1} = u_{j+1,n} + (\theta+r)(u_{j-1,n} - u_{j+1,n})$; $u_{j,n} + (\partial U/\partial t)\Delta t + \frac{1}{2!}(\partial^2 U/\partial t^2)\Delta t^2 + \ldots = u_{j,n} + (\partial U/\partial x)\Delta x + \ldots + (\theta+r)[(u_{j,n} - (\partial U/\partial x)\Delta x + (\partial^2 U/\partial x^2)\Delta x^2/2 - (\partial^3 U/\partial x^3)\Delta x^3/3! + \ldots) - (u_{j,n} + (\partial U/\partial x)\Delta x + (\partial^2 U/\partial x^2)\Delta x^2/2 + (\partial^3 U/\partial x^3)\Delta x^3/3! + \ldots)]$; $(\partial U/\partial t)\Delta t + (\partial^2 U/\partial t^2)\Delta t^2/2 + \ldots = (\partial U/\partial x)\Delta x + (\partial^2 U/\partial x^2)\Delta x^2/2 + \ldots + (\theta+r)[-2[(\partial U/\partial x)\Delta x + (\partial^3 U/\partial x^3)\Delta x^3/3! + \ldots]]$; $(\ldots) - (\ldots) + 2(\theta+r)(\ldots) = 0$.

Now *take away the PDE*, so that the l.t.e. is $(\partial U/\partial t)\Delta t + (\partial^2 U/\partial t^2)\Delta t^2/2 + \ldots - [(\partial U/\partial x)\Delta x + (\partial^2 U/\partial x^2)\Delta x^2/2 + \ldots] + 2(\theta+r)[(\partial U/\partial x)\Delta x + (\partial^3 U/\partial x^3)\Delta x^3/3! + \ldots] - \partial U/\partial t - \nu \partial U/\partial x = 0$ (the red bit is 0). So the l.t.e. is $(\partial U/\partial t)\Delta t + (\partial^2 U/\partial t^2)\Delta t^2/2 + \ldots - [(\partial U/\partial x)\Delta x + (\partial^2 U/\partial x^2)\Delta x^2/2 + \ldots] + 2(\theta+r)[(\partial U/\partial x)\Delta x + (\partial^3 U/\partial x^3)\Delta x^3/3! + \ldots]$, or the l.t.e. is $(\partial U/\partial t)\Delta t + (2\theta+2r-1)(\partial U/\partial x)\Delta x + O(\Delta t^2) + O(\Delta x^2)$. We require the l.t.e. to tend to zero as $\Delta x \to 0$ and $\Delta t \to 0$ independently.

First, *let* $\Delta t \to 0$. Then we **have** $(2\theta+2r-1)(\partial U/\partial x)\Delta x = 0$ left, which *reduces* to $(2\theta-1)(\partial U/\partial x)\Delta x = 0$ (---(1)), as $r = \nu\Delta t/\Delta x \to 0$ as $\Delta t \to 0$. For (1) to *disappear*, we must have $2\theta-1 = 0$; $\theta = ½$. **Now** let $\Delta x \to 0$. Then, we have $(\partial U/\partial t)\Delta t = 0$. But *this disappears as well*, because $\Delta t = \Delta x/c|\nu| \to 0$ as $\Delta x \to 0$. So we have **QED**, and it is only *consistent* iff $\theta = ½$.

Now choose $\theta = ½$, and show that for the scheme to be *stable*, we must have $c \geq 1$. Therefore, $u_{j,n+1} = (½-r)u_{j+1,n} + (½+r)u_{j-1,n}$. Let $\in_{j,n} = \hat{\in}_n \exp(ikx_j)$, so that we have $\hat{\in}_{n+1}\exp(ixk_j) = (½-r)\hat{\in}_n \exp(ik(x_j+\Delta x)) + (½+r)\hat{\in}_n \exp(ik(x_j-\Delta x))$; $\hat{\in}_{n+1} = (½-r)\hat{\in}_n \exp(ik\Delta x) + (½+r)\hat{\in}_n \exp(-ik\Delta x)$. **Now** as $\hat{\in}_{n+1} = g\hat{\in}_n$, we have $g = (½-r)\exp(ik\Delta x) + (½+r)\exp(-ik\Delta x)$. We *want* $|g| \leq 1$ for stability, where $g = ½[\exp(ik\Delta x) + \exp(-ik\Delta x)] + r[\exp(-ik\Delta x) - \exp(ik\Delta x)]$. **Now** $e^{\pm i\theta} = \cos\theta \pm i\sin\theta$, so that $e^{i\theta}+e^{-i\theta} = 2\cos\theta$, and that $e^{-i\theta}-e^{i\theta} = -2i\sin\theta$. *Therefore*, $g = \cos(k\Delta x) - 2ir\sin(k\Delta x)$: g is a **complex** quantity. Now continue to find out that $|g| \leq 1$ only *when* $c \geq 1$.

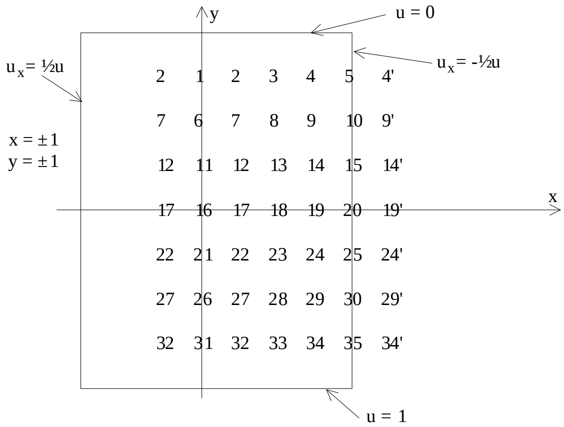
Exercises 7

Q: Using the *simplest central difference formula*, and choosing $\Delta x = \Delta y = h$, write down a **finite difference scheme** for the PDE $\partial^2U/\partial x^2 + \partial^2U/\partial y^2 - \alpha U = 0$. The equation is to be solved inside the *square domain* determined by the lines $x = \pm1$, $y = \pm1$, subject to the **boundary conditions** (i) $U(x,1) = 0$, $-1 \le x \le 1$; (ii) $U(x,-1) = 1$, $-1 \le x \le 1$; (iii) $\partial U/\partial x(1,y) = -\frac{1}{2}U$, $-1<y<1$; and (iv) $\partial U/\partial x(-1,y) = \frac{1}{2}U$, $-1<y<1$.

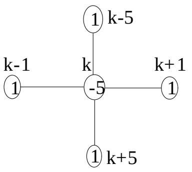
Consider the **case** $\alpha = 16$, with $h = \frac{1}{4}$. Label the *points* with co-ordinates (0, 3/4), (1/4, 3/4), (1/2, 3/4), (3/4, 3/4), (1, 3/4), (0, 1/2), (1/4, 1/2), etc. as 1, 2, 3, 4, 5, 6, 7, etc. Show that we need to solve **thirty five** finite difference equations, which can be written in *matrix* form as $A\underline{u} = \underline{b}$, where $\underline{u}$ is a **column** vector whose *transpose* is $(u_1, u_2, ..., u_{35})$; A is a matrix which can be written in partitioned form *as shown;* and I is the 5×5 **unit** matrix. Find B and $\underline{b}$.

$$\begin{bmatrix} B & I & & & & & \\ I & B & I & & & & \\ & I & B & I & & & \\ & & I & B & I & & \\ & & & I & B & I & \\ & & & & I & B & I \\ & & & & & I & B \end{bmatrix}$$

A: *The central difference approximation* to $u_{xx}+u_{yy}-\alpha u = 0$ (with $\Delta x = \Delta y = h$) is $\frac{1}{h^2}(u_{i+1,j} - 2u_{ij} + u_{i-1,j}) + \frac{1}{h^2}(u_{i,j+1} - 2u_{ij} + u_{i,j-1}) - \alpha u_{ij} = 0$ (---(1)). By **symmetry** about the y-axis, we can *solve* for $x \ge 0$. (The boundary condition on $x = -1$ is **symmetrical** to the boundary condition on $x = 1$ — so the *derivative* of u in the outward direction normal to $x = -1$ is $-\frac{1}{2}u$, i.e. $\partial u/\partial n = -\partial u/\partial x = -\frac{1}{2}u$ on $x = -1$, *and* $\partial u/\partial n = \partial u/\partial x = -\frac{1}{2}u$ on $x = 1$).



We have **35 unknowns** $(u_1, ..., u_{35})$ at the points *shown* in the diagram, where $h = 0.25$. Introduce **guard points** 4', 9', ..., 34' at $x = 1.25$ to enable *second order approximation* for the Neumann boundary condition on $x = 1$. Also, **label** the points at $x = -0.25$ symmetrically about $x = 0$, so as to incorporate *symmetry* into the approximation on $x = 0$.



For $\alpha = 16$ and $h = \frac{1}{4}$, (1) **becomes** $16[u_{i+1,j} - 5u_{ij} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}] = 0$. Therefore, $-5u_{ij} + u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} = 0$. (---(2)). The *computational stencil* is as shown on the left. Using the **numbering scheme** on the diagram, (2) becomes $u_{k-5} + u_{k-1} - 5u_k + u_{k+1} + u_{k+5} = 0$ (---(3)).

On $x = 1$, *eliminate guard points* using central difference approximations to the boundary condition: $u_x = -\frac{1}{2}u$: $u_{(k-1)'} - u_{k-1}/2h = -\frac{1}{2}u_k$; $u_{(k-1)'} - u_{k-1}/2.\frac{1}{4} = -\frac{1}{2}u_k$; $u_{(k-1)'} = u_{k-1} - \frac{1}{4}u_k$ (---(4)). On $x = 1$, (3) takes the **form** $u_{k-5} + u_{k-1} - 5u_k + u_{(k-1)'} + u_{k+5} = 0$. **Substituting** for $u_{(k-1)'}$ from (4) $\Rightarrow$ $u_{k-5} + 2u_{k-1} - \frac{21}{4}u_k + u_{k+5} = 0$ (---(5)). On $x = 0$, (3) takes the **form** $u_{k-5} - 5u_k + 2u_{k+1} + u_{k+5} = 0$. (---(6)).

Matrix 1

$$\begin{bmatrix} -5 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & -9/4 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ \ldots \\ \ldots \\ x_{10} \end{bmatrix}$$

B

$$\begin{bmatrix} -5 & 2 & 0 & 0 & 0 \\ 1 & -5 & 1 & 0 & 0 \\ 0 & 1 & -5 & 1 & 0 \\ 0 & 0 & 1 & -5 & 1 \\ 0 & 0 & 2 & -9/4 & 0 \end{bmatrix}$$

Matrix 3

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -5 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & -9/4 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ \ldots \\ \ldots \\ x_{15} \end{bmatrix}$$

Matrix 4

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -5 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & -5 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & -9/4 & 0 \end{bmatrix} \begin{bmatrix} x_{26} \\ x_{27} \\ \ldots \\ \ldots \\ \ldots \\ x_{35} \end{bmatrix}$$

A

$$\begin{bmatrix} B & I & & & & & \\ I & B & I & & & & 0 \\ & I & B & I & & & \\ & & I & B & I & & \\ & & & I & B & I & \\ 0 & & & & I & B & I \\ & & & & & I & B \end{bmatrix}$$

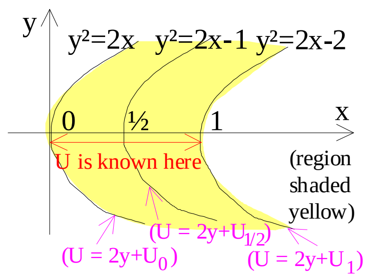
On $y = 3/4$, the u_{k-5} are *not present* (because $u = 0$ on $y = 1$), so (3) becomes $u_{k-1} - 5u_k + u_{k+1} + u_{k+5} = 0$. **On** $y = -3/4$, the u_{k+5} are all 1 (because *$u = 1$ on $y = -1$*), so (3) becomes $u_{k-5} + u_{k-1} - 5u_k + u_{k+1} = -1$. So for <u>points 1 to 5</u>, we have matrix 1, i.e. $(B\ I)(x_1....x_{10})^T = 0$, where B is *as shown*, and I is the 5×5 unit matrix. For <u>points 6 to 10</u>, we have matrix 3, i.e. $(I\ B\ I)(x_1...x_{15})^T = 0$. Similarly for the points *down to 30*, giving a total of 5 **similar** equations. For <u>points 31 to 35</u>, we have matrix 4, i.e. $(I\ B)(x_{26}....x_{35})^T = (-1....-1)^T$. The **complete** system is therefore summarised by the expression $A\underline{u} = \underline{b}$, where A is *as shown;* B is as before; I is the 5×5 **unit** matrix; and $\underline{b} = (0.....0\text{-}1....\text{-}1)^T$. (rows 1 to 25; rows 26 to 35).

<u>Past Paper Question</u>: Consider the *first order quasi-linear PDE* $a\,\partial U/\partial x + b\,\partial U/\partial y = c$, where a, b and c are in general *functions of x, y and U*. Show that the equation may be **transformed** into 2 ODE's if integration is performed *along a particular family of curves* known as the characteristics.

A: Let $a\,\partial U/\partial x + b\,\partial U/\partial y = c$ (---(1)). Consider the *increment* in U between the two points $P = (x, y)$ and $Q = (x+dx, y+dy)$. The *total differential* of U is $dU = \partial U/\partial x\,dx + \partial U/\partial y\,dy$ (---(2)). Substituting for $\partial U/\partial x$ from (1) into (2) gives $dU = {}^{[c-b(\partial U/\partial y)]}/_{a}\,dx + \partial U/\partial y\,dy$, or $\partial U/\partial y(ady-bdx) + (cdx-adU) = 0$ (---(3)). (1) is *quasi-linear,* so that (3) does **not** involve $\partial U/\partial x$. Now *choose* dy/dx at P so as to make $ady-bdx = 0$, i.e. choose $dy/dx = b/a$. This defines the **characteristic direction** at P, and the PDE *reduces* to $cdx-adU = 0$, i.e. $dU/dx = c/a$. The two ODE's *may be written* as $dx/a = dy/b = dU/c$.

Q: Show that the **characteristics** of the equation $y\,\partial U/\partial x + \partial U/\partial y = 2$ are the curves $y^2 = 2(x-c)$, where *c is a constant of integration*. Derive the equations for the 3 characteristics passing through the **points** (0, 0), (1/2, 0), and (1, 0); and sketch them. Indicate on the diagram the region in which a *knowledge* of U on the part of the real axis ($0 \le x \le 1$) determines the solution uniquely. Derive the *solution* for U along a <u>general</u> characteristic, and hence write down the solutions *along the 3 characteristics* passing through the points (0, 0), (1/2, 0), and (1, 0).

A: *Here,* $dx/y = dy/1 = dU/2$. Now $1dx = ydy$; $x = \tfrac{1}{2}y^2+c$; and $y^2 = 2(x-c)$, where *c is a constant of integration*. **QED**. If it passes **through** (0, 0), this implies that $c = 0$, so that $y^2 = 2x$. If it passes **through** (½, 0), this implies that $0 = 1-2c$, so that $c = ½$, and so $y^2 = 2x-1$. And if it passes **through** (1, 0), this implies that $0 = 2-2x$, so that $c = 1$, and so $y^2 = 2x-2$. The sketch is *as shown on the right*.



<u>Now</u> solve for U. $dU/dy = 2$, so that $U = 2y+d$, where d is a *constant of integration*. If U is **known** for $0 \le x \le 1$, and for $y = 0$, then the constant d is given by the *value of U*, where the characteristic <u>crosses</u> the x-axis. If the values of U at $x = 0$; $x = ½$; and $x = 1$ are U_0, $U_{1/2}$ and U_1 respectively, then the *equations for U along the characteristics* passing through these points are $U = 2y+U_0$; $U = 2y+U_{1/2}$; and $U = 2y+U_1$.

Exam Paper: January 2001

Answer 3 questions out of 5 (Questions Done: 2, 4, 5)

(1) (a) Consider the first order ordinary differential equation ${}^{dy}/_{dt} = f(t,y)$ for $t > t_0$ with initial value $y(t_0) = y_0$. Write down the following numerical schemes for such an initial value problem using a uniform timestep Δt: (i) Euler, (ii) trapezoidal Euler, (iii) backward Euler. **[4 marks]**

(b) Write down the difference equations given by the three schemes when applied to the initial value problem ${}^{dy}/_{dt} = -\lambda y$ for $t>0$ with initial value $y(0) = 1$, where $\lambda > 0$ is a constant. **[4 marks]**

(c) Show that the exact solution $y = e^{-\lambda t}$ implies that $y(t_{n+1}) = e^{-\lambda\Delta t}y(t_n)$. Compare this exact expression with the approximations given by the three difference equations derived above. Hence comment on the stability of the three difference schemes. **[4 marks]**

(d) Illustrate your conclusions by comparing the exact solution with approximate solutions for $\lambda = 2$ using $\Delta t = ¼$ and $\Delta t = {}^{3}/_{2}$ for 4 timesteps. **[4 marks]**

(e) Consider the logistic equation ${}^{dy}/_{dt} = y(1-y)$. Set up a predictor-corrector scheme using the Euler scheme for the predictor and the trapezoidal scheme for the corrector. Derive the difference equations, but do not attempt to solve them. **[4 marks]**

(2) The partial differential equation ${}^{\partial U}/_{\partial t} - {}^{\partial^2 U}/_{\partial x^2} = 0$ is approximated at the point $x_j = j\Delta x$, $t_n = n\Delta t$ by the difference equation

$$\theta\left[\frac{u_{j,n+1}-u_{j,n-1}}{2\Delta t}\right] + (1-\theta)\left[\frac{u_{j,n}-u_{j,n-1}}{\Delta t}\right] - \frac{u_{j+1,n}-2u_{j,n}+u_{j-1,n}}{\Delta x^2} = 0.$$

(a) Expand U about this point in terms of Taylor series in time up to Δt^2 and Taylor series in space up to Δx^5. Hence show that the local truncation error at this point is $-½\Delta t(1-\theta)({}^{\partial^2 U}/_{\partial t^2})_{j,n} - {}^{1}/_{12}\Delta x^2(\frac{\partial^4 U}{\partial x^4})_{j,n} + O(\Delta t^2) + O(\Delta x^4)$. **[8 marks]**

(b) Show that setting $\theta = 1 + {}^{\Delta x^2}/_{6\Delta t}$ reduces the local truncation error to one of order Δt^2 and Δx^4. **[3 marks]**

(c) With this value of θ and the choice $\Delta x^2 = 6\Delta t$, show by the Fourier method that the scheme is stable. **[9 marks]**

(3) Consider the initial value problem $\partial U/\partial t = \partial^2 U/\partial x^2$ with initial condition $U(x,0)$ given for $0 \le x \le 1$ and boundary conditions $U(0,t) = g(t)$, $U(1,t) = h(t)$ given for $t>0$.

Construct an explicit finite difference scheme for the problem using forward difference for the time derivative and central differences for the space derivative.

Write down the difference equations in matrix form and show that the stability of the scheme depends on the magnitude of the spectral radius of a certain matrix. Deduce that the scheme is stable when $\Delta t \le (\Delta x)^2/2$.

You may assume that the eigenvalues of the real symmetric $M \times M$ matrix

$$\begin{pmatrix} a & b & & & & & \\ b & a & b & & & & \\ & b & a & b & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & b & a & b \\ & & & & & b & a \end{pmatrix}$$

are given by $\lambda_s = a + 2b(1-2\sin^2(s\pi/M+1))$ for $s = 1, 2, ..., M$. **[20 marks]**

(4) The partial differential equation $\partial^2 U/\partial x^2 - U^2\, \partial^2 U/\partial y^2 = 2y$ is to be solved in the domain $x > 0$, $y > 1$. The initial curve is the line $y=1$ and the initial conditions are $U(x,1) = \partial U/\partial y(x,1) = x^2$.

Use the numerical method of characteristics to find a first approximation for U at the first characteristic grid point R between P = (1,1) and Q = (1.1, 1). **[16 marks]**

Assume that $U = x^2y$ is the exact solution of the problem and compare your numerical answer with the exact solution. **[4 marks]**

You may assume that the partial differential equation $a\,\partial^2 U/\partial x^2 + b\,\partial^2 U/\partial x\partial y + c\,\partial^2 U/\partial y^2 + e = 0$, with the usual notation, reduces to $a\, d(\partial U/\partial x)\, dy/dx + c\, d(\partial U/\partial y) + e\, dy = 0$ on the characteristic curves given by the roots of $a(dy/dx)^2 - b(dy/dx) + c = 0$.

(5) Using the simplest central difference formulae and choosing a uniform mesh $\Delta x = \Delta y = \Delta$, write down a finite difference scheme for the partial differential equation $\partial^2 U/\partial x^2 + 3\partial^2 U/\partial y^2 = \alpha$, where α is a constant. **[5 marks]**

The equation is to be solved inside the square domain determined by the lines $x = \pm 1$, $y = \pm 1$, subject to the boundary conditions (i) $U(1,y) = 0$ for $-1 \leq y \leq 1$ and (ii) $U(x,1) = 1$ for $-1 < x < 1$, and the problem is symmetric with respect to Ox, Oy.

Consider the case $\alpha = -16$ and $\Delta = \frac{1}{4}$. Label the points with coordinates (0, 3/4), (1/4, 3/4), (1/2, 3/4), (3/4, 3/4), (0, 1/2), (1/4, 1/2), ... as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ... etc. Show that it is necessary to solve sixteen finite difference equations which can be written in matrix form as $A\mathbf{u} = \mathbf{b}$, where $\mathbf{u}$ is a column vector whose transpose is $(u_1, u_2, \ldots, u_{16})$ and A is a matrix which can be written in partitioned form as

$$\begin{bmatrix} B & 3I & & \\ 3I & B & 3I & \\ & 3I & B & 3I \\ & & 6I & B \end{bmatrix}$$

where I is the 4×4 unit matrix. Find B and $\mathbf{b}$. **[15 marks]**