Initial Value Problems for Ordinary Differential Equations.

Many natural, scientific & engineering problems can be described in terms of differential equations. Differential equations give us a way to mathematically express rates of change.

We will be considering methods for treating ordinary differential equations (ODES). ODES only consider derivatives with respect to one variable.

Ex. Let $y(t)$ denote the number of individuals in a certain population. If this population has a constant growth rate α (the difference between a constant birth rate and death rate) then the differential equation

$$y'(t) = \alpha y(t)$$

with initial condition $y(0) = y_0$ describes the population growth.

Any textbook on ordinary differential equations details a number of methods for explicitly finding solutions to first order initial value problems.

Ex (cont)

$$\frac{y'(t)}{y(t)} = \alpha$$

$$D_t\left[\ln(y(t))\right] = \alpha$$

$$\ln(y(t)) = \alpha t + \text{const.}$$

$$y(t) = c\,e^{\alpha t} \qquad c = e^{\text{const}}$$

$$y(0) = y_0 \Rightarrow y(t) = y_0\,e^{\alpha t}.$$

To include other effects such as overcrowding, competition for food etc, one might introduce a second term into the equation

$$y'(t) = \alpha y(t) - \beta\left[y(t)\right]^2$$

where $\beta > 0$ and β is small. Introducing a nonlinear term makes the problem much more difficult to study analytically.

Indeed, few problems originating from the study of physical phenomena can be solved exactly.

We will begin by studying numerical methods for approximating the solution $y(t)$ to a problem

$$\frac{dy}{dt} = f(t, y) \qquad \text{for } a \leq t \leq b$$

subject to the initial condition,

$$y(a) = \alpha.$$

The Elementary Theory of Initial Value Problems.

In order to develop numerical methods for initial value problems we will need some definitions and theoretical results.

__Def.__ A function $f(t,y)$ satisfies a __Lipschitz condition__ in the variable y on a set $D \subset \mathbb{R}^2$ if a constant $L > 0$ exists such that

$$|f(t,y_1) - f(t,y_2)| \leq L|y_1 - y_2|$$

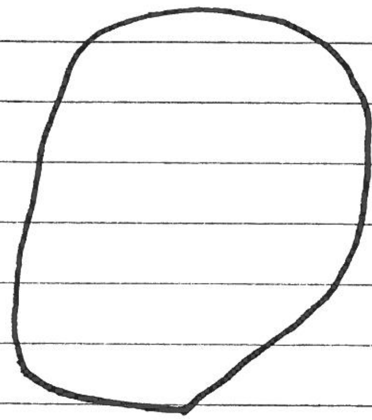
whenever $(t,y_1), (t,y_2) \in D$. The constant L is called a Lipschitz constant for f.

__Def.__ A set $D \subset \mathbb{R}^2$ is said to be __convex__ if whenever (t_1, y_1) and (t_2, y_2) belong to D, the point

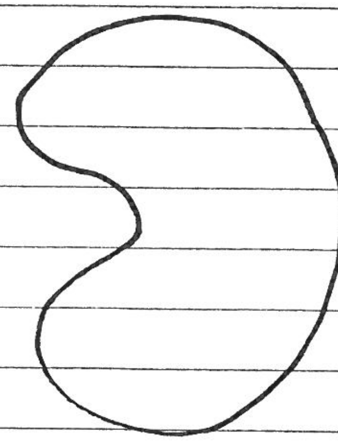
$$((1-\lambda)t_1 + \lambda t_2,\ (1-\lambda)y_1 + \lambda y_2)$$

also belongs to D for each $\lambda \in [0,1]$.

Ex



convex



not convex.

Geometrically, a set is convex provided that whenever two points belong to the set, the entire straight line segment between the points also belongs to the set.

Exercise (not to be handed in)

Show that the set

$$D = \{(t,y) \mid a \le t \le b,\ -\infty < y < \infty\}$$

where a and b are constants is convex.

<u>Thm.</u> Suppose $f(t,y)$ is defined on a convex set $D \subset \mathbb{R}^2$. If a constant $L > 0$ exists with

$$\left| \frac{\partial f}{\partial y}(t,y) \right| \le L$$

for all $(t,y) \in D$ then f satisfies a Lipschitz condition on D in the variable y with Lipschitz constant L.

<u>Proof.</u> Let (t,y_1) and (t,y_2) be in D. Holding t fixed, define $g(y) = f(t,y)$.

Suppose $y_1 < y_2$. Since the line joining (t,y_1) to (t,y_2) lies in D and f is continuous on D, we have $g \in C[y_1, y_2]$. Furthermore $g'(y) = \frac{\partial f(t,y)}{\partial y}$.

Using the Mean Value Theorem on g, a number ξ_1 with $y_1 < \xi < y_2$ exists so that

$$g(y_2) - g(y_1) = g'(\xi)(y_2 - y_1)$$

$$\Longrightarrow \quad f(t,y_2) - f(t,y_1) = \frac{\partial f(t,\xi)}{\partial y}(y_2 - y_1)$$

$$\Longrightarrow \quad |f(t,y_2) - f(t,y_1)| \le L\,|y_2 - y_1|$$

So f satisfies a Lipschitz condition on D in the variable y with Lipschitz constant L. ■

The previous theorem in combination with the next is particularly fundamental for showing the existence and uniqueness of solutions to ODEs.

<u>Thm</u>. Suppose that $D = \{(t,y) \mid a \le t \le b, -\infty < y < \infty\}$ and that $f(t,y)$ is continuous on D.

If f satisfies a Lipschitz condition on D in the variable y, then the initial value problem

$$y'(t) = f(t,y), \quad a \le t \le b, \quad y(a) = \alpha$$

has a unique solution $y(t)$ for $a \le t \le b$.

Ex. Show that the initial-value problem

$$y' = y\cos t,\quad 0 \le t \le 1,\quad y(0) = 1$$

has a unique solution.

Soln. Since $f(t,y) = y\cos t$, we have $\frac{\partial f}{\partial y}(t,y) = \cos t$

$\Rightarrow$ f satisfies a Lipschitz condition in y with $L = 1$ on

$$D = \{(t,y) \mid 0 \le t \le 1,\ -\infty < y < \infty\}.$$

Also, f is continuous on D so there exists a unique solution.

We also need to know if small changes in the statement of the problem introduce correspondingly small changes in the solution.

Def. The initial value problem

$$\frac{dy}{dt} = f(t, y), \quad a \le t \le b, \; y(a) = \alpha$$

is said to be a well-posed problem if:

1. A unique solution, $y(t)$, to the problem exists

2. There exist constants $\varepsilon_0 > 0$ and $K > 0$ such that for any ε with $\varepsilon_0 > \varepsilon > 0$ whenever $\delta(t)$ is continuous with $|\delta(t)| < \varepsilon$ for all t in $[a, b]$ and when $|\delta_0| < \varepsilon$, the initial value problem

$$\frac{dz}{dt} = f(t, z) + \delta(t), \quad a \le t \le b, \quad z(a) = \alpha + \delta_0$$

has a unique solution $z(t)$ that satisfies

$$|z(t) - y(t)| < K\varepsilon$$

for all t in $[a, b]$

The perturbed problem assumes the possibility of an error $\delta(t)$ being introduced in the statement of the differential equation as well as an error δ_0 being present in the initial condition.

Numerical methods always solve perturbed problems since roundoff errors perturb the original problem.
$\Rightarrow$ It only makes sense to approximate well-posed problems.

<u>Thm</u>. Suppose $D = \{(t,y) \mid a \le t \le b \;\&\; -\infty < y < \infty\}$

If f is continuous and satisfies a Lipschitz condition in the variable y on the set D, then the initial value problem

$$\frac{dy}{dt} = f(t,y), \quad a \le t \le b, \; y(a) = \alpha$$

is well-posed.

Ex Show that the initial-value problem

$$y' = t^2 y + 1, \quad 0 \le t \le 1, \quad y(0) = 1$$

is well-posed.

Ans. Since

$$\left| \frac{\partial (t^2 y + 1)}{\partial y} \right| = |t^2| \le 1$$

$t^2 y + 1$ satisfies a Lipschitz condition on D with Lipschitz constant 1. Since $(t^2 y + 1)$ is continuous on

$$\{ (t, y) \mid 0 \le t \le 1, -\infty < y < \infty \}$$

we know that the problem is well-posed.

Euler's Method.

Our first numerical scheme for initial value problems will be Euler's Method — a very simple but low order method!

Consider the initial value problem

$$y' = f(t, y), \quad y(a) = y_0$$

$$a \le t \le b.$$

We will compute an approximation to the solution at the mesh points

$$t_k = a + kh \qquad k = 0, 1, \ldots, N$$

where $h = \frac{b-a}{N}$ is called the <u>step size</u>. Here we have assumed h is a constant, although variable step sizes are also useful.

Euler's Method can be derived using a Taylor Expansion:

$$y(t_{k+1}) = y(t_k + h) = y(t_k) + h y'(t_k) + \frac{h^2}{2} y''(\xi_k)$$

$$= y(t_k) + h f(t_k, y(t_k)) + \frac{h^2}{2} y''(\xi_k)$$

Euler's Method constructs an approximation

$$w_k \approx y(t_k)$$

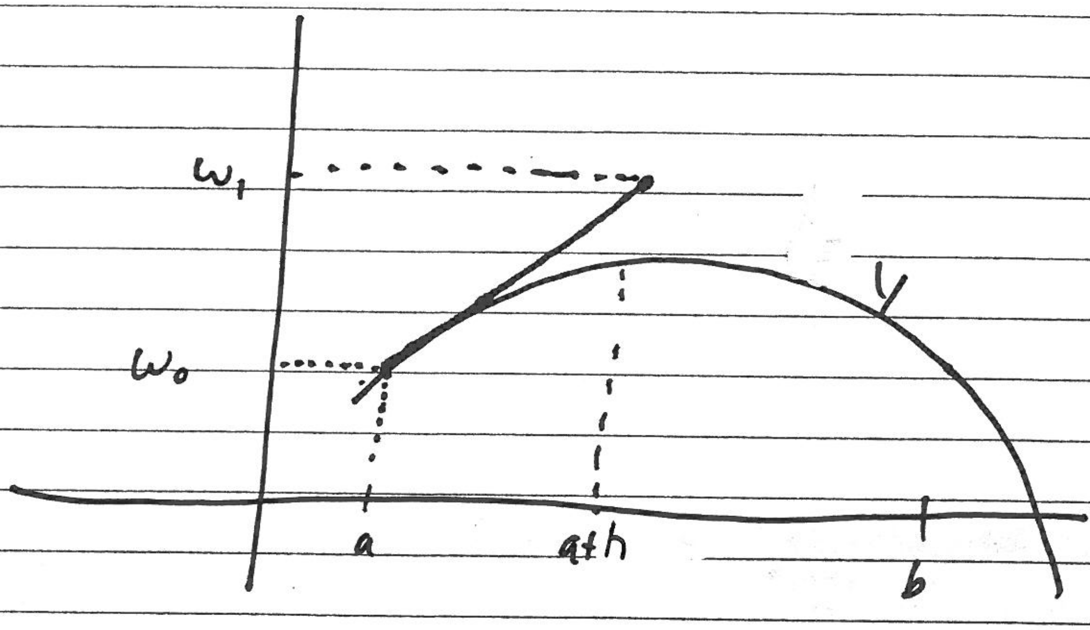
by dropping the remainder term

$$w_0 = y_0$$

$$w_k = w_{k-1} + h f(t_{k-1}, w_{k-1}) \qquad 1 \le k \le N$$

Euler's Method also has a simple geometric interpretation

At a given point (t_n, w_n) we march for a given finite step size h in a direction given by the slope $f(t_n, w_n)$



Euler's Method has a particularly straightforward error analysis.

The analysis is also interesting because it can be extended to the higher order methods we will consider later.

To derive the proof of convergence we need the following

<u>Lemma</u>: If s and t are positive real numbers $\{a_i\}_{i=0}^{K}$ is a sequence satisfying

$$a_0 \geq -t/s,$$

$$a_{i+1} \leq (1+s)a_i + t \qquad i = 0, 1, \ldots, K$$

then $a_{i+1} \leq e^{(i+1)s}(a_0 + t/s) - t/s$

<u>Proof</u>. $a_{i+1} \leq (1+s)\,a_i + t$

$\leq (1+s)\left[(1+s)a_{i-1}+t\right]+t$

$\vdots$

$\leq (1+s)^{i+1}a_0 + \underbrace{\left[1+(1+s)+(1+s)^2+\cdots+(1+s)^i\right]}_{\substack{\text{geometric series} \\ \text{sums to } \frac{1}{s}\left[(1+s)^{i+1}-1\right]}} t$

$= (1+s)^{i+1}\left(a_0 + t/s\right) - t/s$

$\leq \left(1+s+\frac{1}{2}s^2+\frac{1}{3!}s^3+\ldots\right)^{i+1}\left(a_0+t/s\right) - t/s$

$= e^{(i+1)s}\left(a_0 + t/s\right) - t/s.$

Thm. Suppose f is continuous a satisfies a Lipschitz condition with constant L on

$$D = \{(t,y) \mid a \le t \le b, -\infty < y < \infty\}$$

and that a constant M exists with the property that

$$|y''(t)| \le M$$

Let $y(t)$ denote the unique solution to the initial value problem

$$y' = f(t,y), \quad y(a) = y_0 \quad a \le t \le b$$

and $w_0, w_1, \dots, w_N$ be the approximations generated by Euler's Method

Then for each $i = 0, 1, \dots, N$

$$|y(t_i) - w_i| \le \frac{hM}{2L}\left[e^{L(t_i - a)} - 1\right].$$

Proof. When $i=0$ the result is clearly true.

Consider $i=1,2,\ldots$

$$|y(t_{i+1})-w_{i+1}| = |y(t_i)+hf(t_i,y_i)+\tfrac{h^2}{2}y''(\xi_i)-(w_i+hf(t_i,w_i))|$$
$$= |y(t_i)-w_i+h[f(t_i,y_i)-f(t_i,w_i)]+\tfrac{h^2}{2}y''(\xi_i)|$$
$$\le |y(t_i)-w_i|+h|f(t_i,y_i)-f(t_i,w_i)|+\tfrac{h^2}{2}|y''(\xi_i)|$$

$\le L|y(t_i)-w_i|$ since f satisfies a Lipschitz condition with constant L

$|y''(t)|\le M$

$$\le |y(t_i)-w_i|+hL|y(t_i)-w_i|+\tfrac{h^2}{2}M$$
$$=(1+hL)|y(t_i)-w_i|+\tfrac{h^2}{2}M$$
$$\le e^{(i+1)hL}\left(|y(t_0)-w_0|+\frac{h^2M}{2hL}\right)-\frac{h^2M}{2hL}$$

By Lemma with
$a_j = |y_j - w_j|$
$s = hL$
$t = h^2M/2$

$$= e^{(t_{i+1}-t_0)L}\left(\frac{hM}{2L}\right)-\frac{hM}{2L}$$
$$=\left(e^{(t_{i+1}-a)L}-1\right)\frac{hM}{2L}$$

$\blacksquare$

<u>Note</u> that the theorem requires that

$$|y''(t)| \le M.$$

The second derivative $y''(t)$ may not be known, but if $\partial f/\partial t$ and $\partial f/\partial y$ exist then

$$y''(t) = \frac{d}{dt} y'(t) = \frac{df}{dt}(t, y(t))$$

$$= \frac{\partial f}{\partial t}(t, y(t)) + \frac{\partial f}{\partial y}(t, y(t)) \cdot f(t, y(t))$$

<u>Example</u>. What value of h is needed to ensure that

$$|y(t_i) - w_i| \le 0.1$$

for the initial value problem

$$y' = \frac{2}{t} y + t^2 e^t \qquad 1 \le t \le 2$$

$$y(1) = 0$$

You are given $y''(t) = (2 + 4t + t^2)e^t - 2e$.

23.9

Ans. $y''(t)$ is increasing and positive on $[1,2]$

so $|y''(t)| \le |y''(2)|$

$= |4e^2 - 2e$

$= 98.0102$

Since $\left|\frac{\partial}{\partial y}\left(\frac{2}{t}y + t^2 e^t\right)\right|$

$\le \left|\frac{2}{t}\right|$

≤ 2

a Lipschitz constant for

$f(t,y) = \frac{2}{t}y + t^2 e^t$ is $L = 2$

$\therefore \quad |y(t_i) - w_i| \le \frac{hM}{2L}\left[e^{L(t_i - a)} - 1\right] < .0.1$

We need to choose h so that

$$\frac{98.0102\,h}{4}\left[e^{2(2-1)} - 1\right] < .0.1$$

$\Rightarrow \quad h < \frac{0.4}{98.0102(e^2 - 1)} = .00064.$

We need a way to compare the efficiency of different approximation methods.

One approach is to compare how much the exact solution to the differential equation fails to satisfy the difference equation being used for the approximation.

<u>Def</u>. The difference method

$$w_0 = \alpha$$
$$w_{i+1} = w_i + h\,\phi(t_i, w_i)$$

has local truncation error

$$\tau_{i+1}(h) = \frac{y(t_{i+1}) - \left(y(t_i) + h\,\phi(t_i, y(t_i))\right)}{h}$$

$$= \frac{y(t_{i+1}) - y(t_i)}{h} - \phi(t_i, y(t_i))$$

$i = 0, 1, \ldots, N-1.$

Ex Euler's Method.

The difference method

$$w_0 = \alpha$$

$$w_{i+1} = w_i + h f(t_i, w_i)$$

has local truncation error

$$\tau_{i+1}(h) = \frac{y(t_{i+1}) - y(t_i)}{h} - f(t_i, y(t_i))$$

$$= \frac{h}{2} y''(\xi_i) \qquad \text{for some } \xi_i \in (t_i, t_{i+1})$$

When $|y''(t)| \le M$ on $[a,b]$

we have

$$|\tau_{i+1}(h)| \le \frac{h}{2} M$$

So the local truncation error in Euler's Method is $O(h)$.

Local truncation errors are called local because they measure the accuracy of the method at a specific step, assuming the method was exact at the previous steps.

We obviously want the local truncation error to be small.

Often, methods for solving ordinary differential equations are derived so that the local truncation errors are of the form

$$O(h^p)$$

for the largest p possible, while keeping the number of operations reasonable.

How to obtain improved accuracy?

__ie__ a larger p in the $O(h^p)$ local truncation error

Suppose we want to approximate the solution to the initial value problem

$$y' = f(t, y) \qquad a \le t \le b$$
$$y(0) = \alpha$$

where $y(t) \in C^{(n+1)}[a, b]$.

One approach is to expand the solution in terms of its n^{th} Taylor polynomial about t_i.

$$y(t_{i+1}) = y(t_i) + h\underset{\substack{\uparrow \\ f(t,y(t))}}{y'(t_i)} + \frac{h^2}{2}\underset{\substack{\uparrow \\ f'(t,y(t))}}{y''(t_i)} + \dots + \frac{h^n}{n!}\underset{\substack{\uparrow \\ f^{(n-1)}(t,y(t))}}{y^{(n)}(t_i)} + R$$

where
$$R = \frac{h^{n+1}}{(n+1)!}y^{(n+1)}(\xi_i)$$

If we drop the remainder term we obtain the

Taylor Method of Order n

$$w_0 = \alpha$$

$$w_{i+1} = w_i + h\,T^{(n)}(t_i, w_i) \qquad i = 0, 1, \ldots N-1$$

where

$$T^{(n)}(t_i, w_i) = f(t_i, w_i) + \frac{h}{2} f'(t_i, w_i) + \cdots + \frac{h^{n-1}}{n!} f^{(n-1)}(t_i, w_i)$$

Note: Euler's Method is just Taylor's Method of order one.

Ex Use Taylor's method of order two to approximate the solution for the initial value problem

$$y' = te^{3t} - 2y, \quad 0 \le t \le 1$$

$$y(0) = 0 \qquad \text{with} \quad h = 0.5$$

Ans. The first approximation is

$$w_1 = w_0 + h(t_0 e^{3t_0} - 2w_0) + \frac{h^2}{2}(t_0 e^{3t_0} + e^{3t_0} + 4w_0)$$

$$= 0 + 0.5(0 - 0) + \frac{(0.5)^2}{2}(0 + 1 + 0) = 0.125$$

and the second is

$$w_2 = w_1 + h(t_1 e^{3t_1} - 2w_1) + \frac{h^2}{2}(t_1 e^{3t_1} + e^{3t_1} + 4w_1)$$

$$= 0.125 + 0.5(0.5e^{1.5} - 2(0.125)) + \frac{(0.5)^2}{2}(0.5e^{1.5} + e^{1.5} + 4(0.125))$$

$$= 2.02323897$$

Note that if we want to determine an approximation at some intermediate point (eg for some $t \in (t_{i-1}, t_i)$) then cubic Hermite interpolation based on $(y(t_{i-1}), y'(t_{i-1}), y(t_i), y'(t_i))$ is a particularly natural choice for a Taylor method of degree ≤ 4.

Such an interpolation has the advantages that it can be constructed locally and that $y'(t) = f(t, y(t))$ is given.

To interpolate results from very high order Taylor methods $(n > 4)$ we will need higher order oscillating polynomials to preserve the overall accuracy of the results.

The local truncation error for Taylor's method of order n is also easily derived:

$$y_{i+1} - y_i - hf(t_i, y_i) - \frac{h^2}{2} f'(t_i, y_i) - \cdots - \frac{h^n}{n!} f^{(n-1)}(t_i, y_i)$$

$$= \frac{h^{n+1}}{(n+1)!} f^{(n)}(\xi_i, y(\xi_i))$$

where $y_i \equiv y(t_i)$

Thus the local truncation error is

$$\tau_{i+1}(h) = \frac{y_{i+1} - y_i}{h} - f(t_i, y_i) - \frac{h}{2} f'(t_i, y_i) - \cdots - \frac{h^{n-1}}{n!} f^{(n-1)}(t_i, y_i)$$

$$= \frac{h^n}{(n+1)!} f^{(n)}(\xi_i, y(\xi_i))$$

Thus if $y \in C^{n+1}[a,b]$

$\Rightarrow y^{(n+1)}(t) = f^{(n)}(t, y(t))$ is bounded

and $\tau_i = O(h^n)$ for each $i = 1, 2, \ldots, N$.

Runge - Kutta Methods.

Taylor Methods are seldom used in practice because they require the computation and evaluation of the derivatives of $f(t, y)$. These evaluations can be complicated and time - consuming.

Runge - Kutta methods have the high local truncation error of the Taylor methods but do not need to compute and evaluate derivatives of $f(t, y)$.

To give some idea of how Runge - Kutta methods are developed, we will show the derivation of a simple second order method. Here, the increment to w is a weighted average of two estimates of the increment which we will call K_1 and K_2:

$$
* \left\{ \begin{aligned} w_{n+1} &= w_n + aK_1 + bK_2 \\ K_1 &= h f(t_n, w_n) \\ K_2 &= h f(t_n + \alpha h, w_n + \beta K_1) \end{aligned} \right.
$$

We can think of K_1 and K_2 as estimates of the change in y when t advances by h because they are the product of the change in t and a value for the slope of the curve.

Runge Kutta methods often use the simple Euler estimate as the first estimate of Δy.

Now, our problem is to devise a scheme by choosing the four parameters a, b, α, β. We do so by making the local truncation error of (*) as small as possible ...

Re-writing (*)

$$w_{n+1} = w_n + ah\, f(t_n, w_n) + bh f(t_n + \alpha h, w_n + \beta h f(t_n, w_n))$$

The local truncation error is then

$$T_{n+1}(h) = \frac{y_{n+1} - y_n}{h} - af(t_n, y_n) - bf(t_n + \alpha h, y_n + \beta h f(t_n, y_n))$$

24.6

Well, $y_{n+1} = y_n + h f(t_n, y_n) + \frac{h^2}{2} \underbrace{f'(t_n, y_n)}_{f_t(t_n, y_n) + f_y(t_n, y_n) \cdot f(t_n, y_n)} + O(h^3)$

$$f(t_n + \alpha h, y_n + \beta h f(t_n, y_n))$$
$$= f(t_n, y_n) + f_t(t_n, y_n)\alpha h + f_y(t_n, y_n) f(t_n, y_n) \beta h + O(h^2)$$

$$\therefore \tau_{n+1}(h) = (1 - a - b)\, f(t_n, y_n)$$
$$+ h\left(\tfrac{1}{2} - \alpha b\right) f_t(t_n, y_n)$$
$$+ h\left(\tfrac{1}{2} - \beta b\right) f_y(t_n, y_n)\, f(t_n, y_n)$$

Thus the local truncation error will be order two provided

$$a + b = 1$$
$$\alpha b = 1/2$$
$$\beta b = 1/2$$

but there is not enough flexibility to obtain a third order method. (Proof is an exercise)

Some examples of ~~two~~ second order Runge Kutta schemes

$$\left.\begin{aligned} a &= 0 \\ b &= 1 \\ \alpha &= \tfrac{1}{2} \\ \beta &= \tfrac{1}{2} \end{aligned}\right\} \Rightarrow \quad \begin{aligned} w_0 &= y(t_0) \\ w_{n+1} &= w_n + hf\left(t_n + \tfrac{h}{2}, w_n + \tfrac{h}{2} f(t_n, w_n)\right) \end{aligned}$$

The Midpoint Method.

$$\begin{aligned} a &= \tfrac{1}{2} \\ b &= \tfrac{1}{2} \\ \alpha &= 1 \\ \beta &= 1 \end{aligned} \Rightarrow \quad \begin{aligned} w_0 &= y(t_0) \\ w_{n+1} &= w_n + \tfrac{h}{2}\left[f(t_n, w_n) + f(t_{n+1}, w_n + hf(t_n, w_n))\right] \end{aligned}$$

The Modified Euler Method

$$\begin{aligned} a &= 1/4 \\ b &= 3/4 \\ \alpha &= 2/3 \\ \beta &= 2/3 \end{aligned} \Rightarrow \quad \begin{aligned} w_0 &= y(t_0) \\ w_{n+1} &= w_n + \tfrac{h}{4}\left[f(t_n, w_n) + 3f\left(t_n + \tfrac{2}{3}h, w_n + \tfrac{2}{3}hf(t_n, w_n)\right)\right] \end{aligned}$$

Heun's Method

Third order Runge-Kutta methods are not commonly used.

Fourth order Runge-Kutta methods are widely used and are derived in a similar fashion. Greater complexity results from having to compare terms through h^4 and this gives a set of 11 equations in 13 unknowns. The set of equations can be solved with 2 unknowns being chosen arbitrarily.

The most commonly used set of values leads to the algorithm

$$w_{n+1} = w_n + \tfrac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4)$$

$$K_1 = h f(t_n, w_n)$$

$$K_2 = h f(t_n + \tfrac{1}{2}h, w_n + \tfrac{1}{2}K_1)$$

$$K_3 = h f(t_n + \tfrac{1}{2}h, w_n + \tfrac{1}{2}K_2)$$

$$K_4 = h f(t_n + h, w_n + K_3)$$

"The classical 4^{th} order Runge-Kutta method"

The main computational effort in applying Runge-Kutta methods is the evaluation of f. In the second order methods, the local truncation error is $O(h^2)$ and the cost is two functional evaluations per step.

The Runge-Kutta method of order four requires four evaluations per step and the local truncation error is $O(h^4)$.

We may wonder about higher order formulas...

Butcher has shown that the following relationship holds:

Evaluations per step	2	3	4	$5 \le n \le 7$	$8 \le n \le 9$	$10 \le n$
Best possible local error	$O(h^2)$	$O(h^3)$	$O(h^4)$	$O(h^{n-1})$	$O(h^{n-2})$	$O(h^{n-3})$

This indicates why methods of order less than 5 are often used rather than higher order methods with a larger step size.