NUMERICAL METHODS FOR ODEs

Consider the initial value problem

$$y' = f(x, y), \quad x_0 \le x_0 \le b, \quad y(x_0) = Y_0$$

and denote its solution by $Y(x)$. Most numerical methods solve this by finding values at a set of node points:

$$x_0 < x_1 < \cdots < x_N \le b$$

The approximating values are denoted in this book in various ways. Most simply, we have

$$y_1 \approx Y(x_1), \ \cdots, \ y_N \approx Y(x_N)$$

We also use

$$y(x_i) \equiv y_i, \quad i = 0, 1, ..., N$$

To begin with, and for much of our work, we use a fixed stepsize h, and we generate the node points by

$$x_i = x_0 + i\,h, \quad i = 0, 1, ..., N$$

Then we also write

$$y_h(x_i) \equiv y_i \approx Y(x_i), \quad i = 0, 1, ..., N$$

EULER'S METHOD

Euler's method is defined by

$$y_{n+1} = y_n + h\, f(x_n, y_n), \quad n = 0, 1, ..., N-1$$

with $y_0 = Y_0$. Where does this method come from?

There are various perspectives from which we can derive numerical methods for solving

$$y' = f(x, y), \quad x_0 \le x_0 \le b, \quad y(x_0) = Y_0$$

and Euler's method is simplest example of most such perspectives. Moreover, the error analysis for Euler's method is introduction to the error analysis of most more rapidly convergent (and more practical) numerical methods.

A GEOMETRIC PERSPECTIVE

Look at the graph of $y = Y(x)$, beginning at $x = x_0$. Approximate this graph by the line tangent at $(x_0, Y(x_0))$:

$$\begin{aligned} y &= Y(x_0) + (x - x_0)Y'(x_0) \\ &= Y(x_0) + (x - x_0)f(x_0, Y_0) \end{aligned}$$

Evaluate this tangent line at x_1 and use this value to approximate $Y(x_1)$. This yields Euler's approximation.

We could generalize this by looking for more accurate means of approximating a function, e.g. by using a higher degree Taylor approximation.

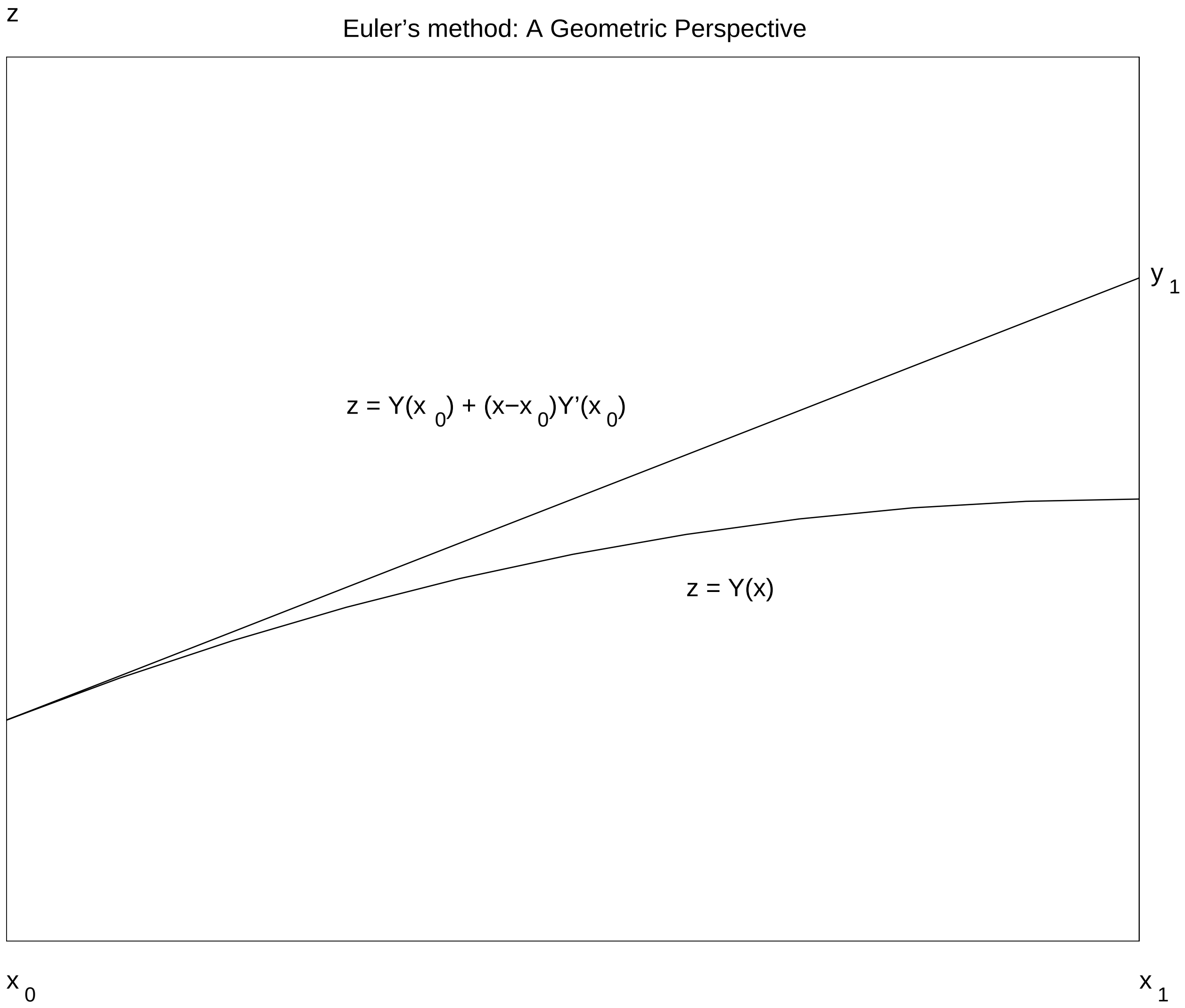

z
Euler’s method: A Geometric Perspective
y 1
z = Y(x 0) + (x−x 0)Y’(x 0)
z = Y(x)
x 0
x 1

Figure 1: A geometric perspective on Euler’s method

TAYLOR'S SERIES

Approximate $Y(x)$ about x_0 by a Taylor polynomial approximation of some degree:

$$\begin{aligned} Y(x_0+h) \approx Y(x_0) + h\,Y'(x_0) + \frac{h^2}{2!}Y''(x_0) \\ + \cdots + \frac{h^p}{p!}Y^{(p)}(x_0) \end{aligned}$$

Euler's method is the case $p = 1$:

$$\begin{aligned} Y(x_0+h) &\approx Y(x_0) + h\,Y'(x_0) \\ &= y_0 + h\,f(x_0, y_0) \equiv y_1 \end{aligned}$$

We have an error formula for Taylor polynomial approximations; and in this case,

$$Y(x_1) - y_1 = \frac{h^2}{2}Y''(\xi_0)$$

with some $x_0 \le \xi_0 \le x_1$.

GENERAL ERROR FORMULA

In general,

$$y_{n+1} = y_n + h\, f(x_n, y_n), \quad n = 0, 1, ..., N-1$$

$$\begin{aligned} Y(x_{n+1}) &= Y(x_n) + h\, Y'(x_n) + \frac{h^2}{2} Y''(\xi_n) \\ &= Y(x_n) + h\, f(x_n, Y(x_n)) + \frac{h^2}{2} Y''(\xi_n) \end{aligned}$$

with some $x_n \le \xi_n \le x_{n+1}$.

We will use this as the starting point of our error analyses of Euler's method. In particular,

$$\begin{aligned} Y(x_{n+1}) - y_{n+1} = \; & Y(x_n) - y_n \\ & + h\,[\, f(x_n, Y(x_n)) - f(x_n, y_n)] \\ & + \frac{h^2}{2} Y''(\xi_n) \end{aligned}$$

NUMERICAL DIFFERENTIATION

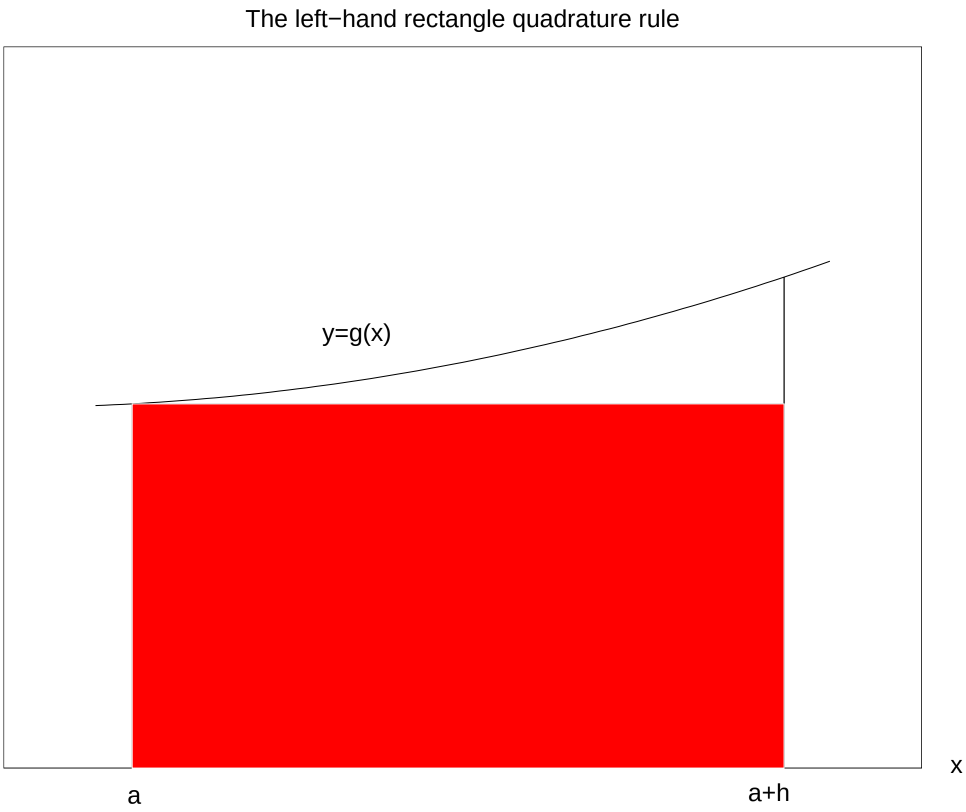
From beginning calculus,

$$Y'(x_n) \approx \frac{Y(x_n + h) - Y(x_n)}{h}$$

This leads to

$$\begin{aligned} Y(x_n + h) &\approx Y(x_n) + h\,Y'(x_n) \\ &= Y(x_n) + h\ f(x_n, Y(x_n)) \\ &\approx y_n + h\,f(x_n, y_n) \end{aligned}$$

Most numerical differentiation approximations can be used to obtain numerical methods for solving the initial value problem. However, a number of such formulas turn out to be poor methods for solving differential equations, and we will see an example of this in the one of the following sections of the book.



NUMERICAL INTEGRATION

Consider the numerical approximation

$$\int_a^{a+h} g(x)\,dx \approx hg(a)$$

which is called the *left-hand rectangle rule*. It is the area of the rectangle with base $[a, a+h]$ and height $g(a)$.

Return to the differential equation $y' = f(x, y)$ and substitute the solution $Y(x)$ for y:

$$Y'(x) = f(x, Y(x))$$

Integrate this over the interval $[x_n, x_{n+1}]$,

$$\int_{x_n}^{x_{n+1}} Y'(x)\, dx = \int_{x_n}^{x_{n+1}} f(x, Y(x))\, dx$$

$$Y(x_{n+1}) = Y(x_n) + \int_{x_n}^{x_{n+1}} f(x, Y(x))\, dx$$

Integrate this with the left-hand rectangle rule,

$$Y(x_{n+1}) \approx Y(x_n) + h\, f(x_n, Y(x_n))$$

Again this leads to Euler's method.

EXAMPLE

Consider the problem

$$y' = -y + 2\cos x, \qquad y(0) = 1$$

We solve this on the interval $0 \le x \le 5$. Look at the behaviour of the error

$$e_h(x) = Y(x) - y_h(x)$$

as a function of both h and x.

1. For a particular x, the error appears to be halved when h is halved.
2. For a fixed h, the error varies with x, and it appears to be oscillating in sign.

From this,

$$e_h(x) \approx c(x)h$$

seems accurate empirically, with $c(x)$ an oscillating function of x.

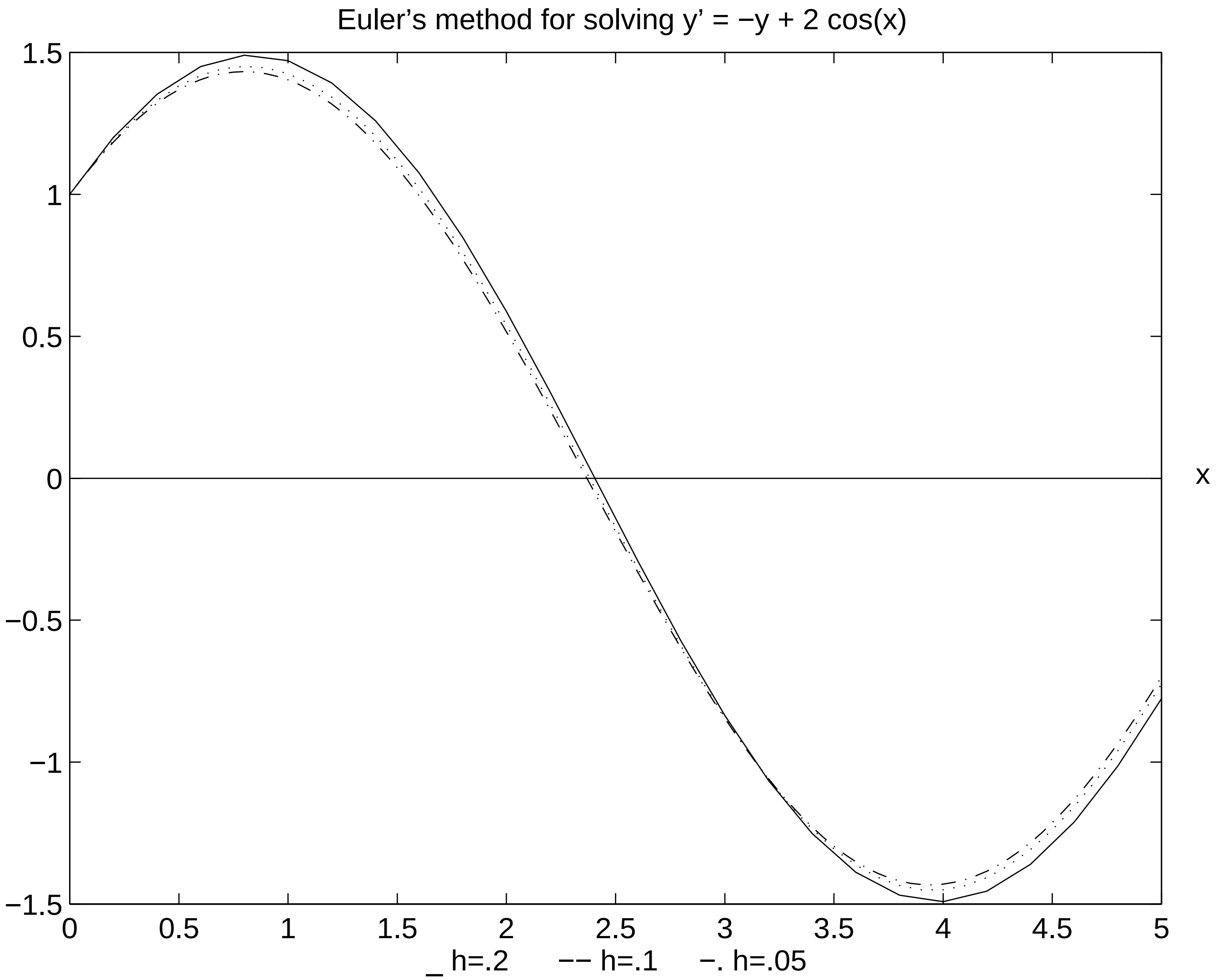


Figure 2: Solution by Euler's method

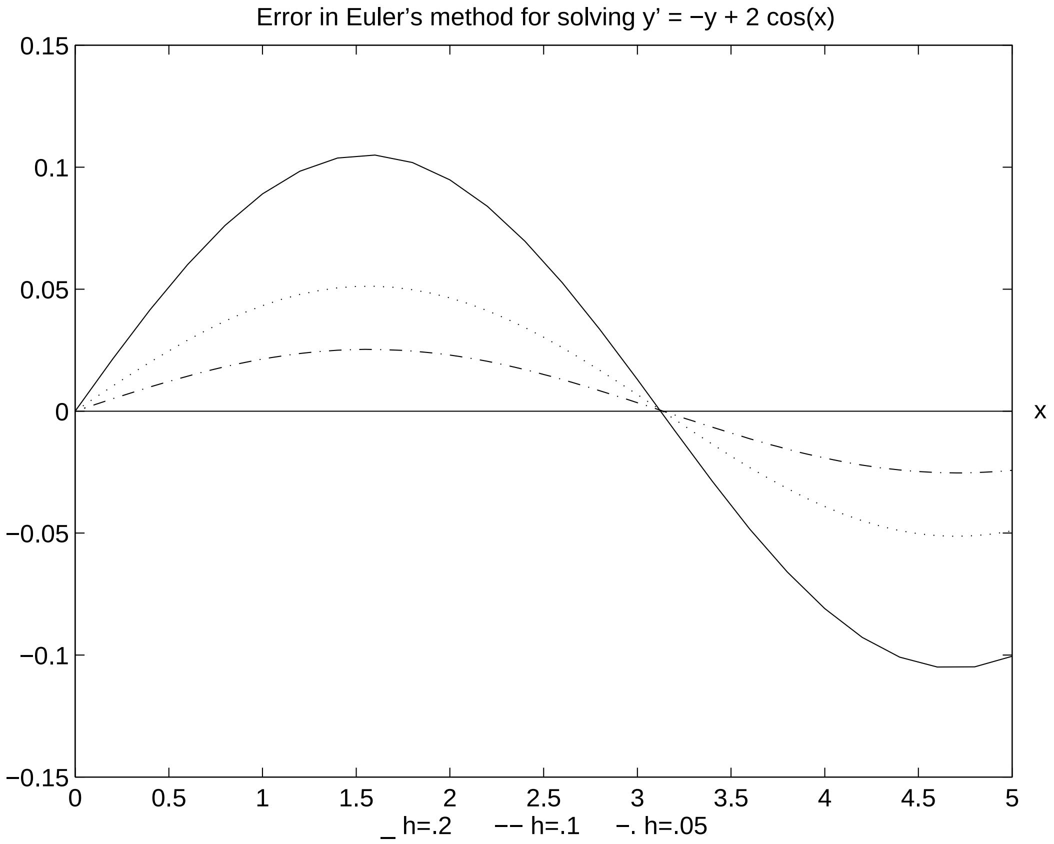


Figure 3: Error in Euler solution

ERROR ANALYSIS - SPECIAL CASES

We begin with a couple of special cases, to obtain some additional intuition on the behaviour of the error $e_h(x) = Y(x) - y_h(x)$. Consider

$$y' = 2x, \quad y(0) = 0$$

This has the solution $Y(x) = x^2$. Euler's method becomes

$$y_{n+1} = y_n + 2x_n h, \quad y_0 = 0$$

$$\begin{array}{rcll} y_1 & = & y_0 + 2x_0 h & = x_1 x_0 \\ y_2 & = & y_1 + 2x_1 h & = x_1 x_0 + 2x_1 h = x_2 x_1 \\ y_3 & = & y_2 + 2x_2 h & = x_2 x_1 + 2x_2 h = x_3 x_2 \end{array}$$

By induction,

$$y_n = x_n x_{n-1}, \quad n \geq 1$$

For the error,

$$Y(x_n) - y_n = x_n^2 - x_n x_{n-1} = x_n h$$

Return to our error equation

$$\begin{aligned} Y(x_{n+1}) - y_{n+1} &= Y(x_n) - y_n \\ &\quad + h\,[\,f(x_n, Y(x_n)) - f(x_n, y_n)] \\ &\quad + \frac{h^2}{2} Y''(\xi_n) \end{aligned} \tag{1}$$

With the mean value theorem,

$$f(x_n, Y(x_n)) - f(x_n, y_n) = \frac{\partial f(x_n, \zeta_n)}{\partial y}\,[Y(x_n) - y_n]$$

with ζ_n between $Y(x_n)$ and y_n. Then we can write

$$e_h(x_{n+1}) = \left[1 + h\frac{\partial f(x_n, \zeta_n)}{\partial y}\right] e_h(x_n) + \frac{h^2}{2} Y''(\xi_n) \tag{2}$$

with $e_h(x_0) = 0$. We also will assume henceforth that

$$K \equiv \max_{\substack{x_0 \le x \le b \\ -\infty < y < \infty}} \left| \frac{\partial f(x, y)}{\partial y} \right| < \infty$$

Consider those differential equations with

$$\frac{\partial f(x,y)}{\partial y} \leq 0, \quad x_0 \leq x \leq b, \quad -\infty < y < \infty$$

Then

$$-1 \leq 1 + h\frac{\partial f(x_n,\zeta_n)}{\partial y} \leq 1$$

provided h is chosen sufficiently small, i.e.

$$h \leq \frac{2}{K}$$

Using this in our error formula (2),

$$|e_h(x_{n+1})| \leq |e_h(x_n)| + \frac{h^2}{2}\left\|Y''\right\|_\infty, \quad n \geq 0 \quad (3)$$

in which

$$\left\|Y''\right\|_\infty = \max_{x_0 \leq t \leq b}\left|Y''(t)\right|$$

Using induction with (3), we can prove

$$|e_h(x_n)| \leq \frac{h}{2}(x_n - x_0)\left\|Y''\right\|_\infty$$

Again the error is bounded by something of the form $c(x_n)h$.

GENERAL ERROR ANALYSIS

Return to

$$e_h(x_{n+1}) = \left[1 + h\frac{\partial f(x_n, \zeta_n)}{\partial y}\right] e_h(x_n) + \frac{h^2}{2} Y''(\xi_n)$$

For comparison with other numerical methods, we introduce

$$\tau_n = \frac{h}{2} Y''(\xi_n)$$

$$\tau(h) = \frac{h}{2} \left\| Y'' \right\|_\infty$$

Our error equation becomes

$$e_h(x_{n+1}) = \left[1 + h\frac{\partial f(x_n, \zeta_n)}{\partial y}\right] e_h(x_n) + h\tau_n \quad (4)$$

Take bounds to obtain

$$\begin{aligned} |e_h(x_{n+1})| &\leq (1 + hK) |e_h(x_n)| + h |\tau_n| \\ &\leq (1 + hK) |e_h(x_n)| + h\tau(h) \end{aligned}$$

By induction, we can show this implies

$$\begin{aligned} &|e_h(x_n)| \leq (1 + hK)^n |e_h(x_0)| \\ &\quad + \left\{1 + (1 + hK) + ... + (1 + hK)^{n-1}\right\} h\tau(h) \end{aligned}$$

This leads to

$$|e_h(x_n)| \leq (1+hK)^n |e_h(x_0)| + \frac{(1+hK)^n - 1}{K}\tau(h) \tag{5}$$

We need the following:

$$(1+t)^n \leq e^{nt}, \quad n \geq 0, \quad -1 \leq t < \infty$$

Then (5) leads to

$$|e_h(x_n)| \leq e^{nhK} |e_h(x_0)| + \frac{e^{nhK} - 1}{K}\tau(h)$$

Since $nh = x_n - x_0$, we have

$$|e_h(x_n)| \leq e^{(x_n - x_0)K} |e_h(x_0)| + \frac{e^{(x_n - x_0)K} - 1}{K}\tau(h)$$

For $e_h(x_0) = 0$, we have that the error has the general form

$$|e_h(x_n)| \leq c(x_n)h$$

STABILITY ANALYSIS

We have

$$y_{n+1} = y_n + h\, f(x_n, y_n), \quad y_0 = Y_0$$

Now consider perturbing this to

$$z_{n+1} = z_n + h\left[\, f(x_n, z_n) + \delta(x_n)\right], \quad z_0 = Y_0 + \epsilon$$

In which $\delta(x)$ is a bounded function on $[x_0, b]$.

Let $e_n = z_n - y_n$. Subtracting above,

$$e_{n+1} = e_n + h\left[f(x_n, z_n) - f(x_n, y_n)\right] + h\delta(x_n)$$

with $e_0 = \epsilon$. Using the type of analysis used above, we have

$$|e_{n+1}| \leq (1 + hK)\,|e_n| + h\,\|\delta\|_\infty\,, \quad n \geq 0$$

This yields

$$|e_n| \le e^{(x_n - x_0)K} |\epsilon| + \frac{e^{(x_n - x_0)K} - 1}{K} \|\delta\|_\infty$$

Thus we have a type of stability, in which the change in the numerical solution is bounded by a constant times the change in the data of the initial value problem, *independent of* h.

In the case our differential equation satisfies

$$\frac{\partial f(x,y)}{\partial y} \le 0, \quad x_0 \le x \le b, \quad -\infty < y < \infty$$

and the stepsize satisfies $hK \le 1$, we can show the much better result

$$|e_n| \le |\epsilon| + (x_n - x_0) \|\delta\|_\infty, \quad n \ge 0$$

EFFECT OF ROUNDING ERROR

Again consider the Euler method

$$y_{n+1} = y_n + h\, f(x_n, y_n), \quad y_0 = Y_0$$

Now consider that rounding errors occur in the calculation of y_{n+1} from y_n. Let $\widetilde{y}_n$ denote the numerical values actually computed. Then we have

$$\widetilde{y}_{n+1} \approx \widetilde{y}_n + h\, f(x_n, \widetilde{y}_n), \quad \widetilde{y}_0 \approx Y_0$$

To have an equation, write

$$\widetilde{y}_{n+1} = \widetilde{y}_n + h\, f(x_n, \widetilde{y}_n) + \rho_n, \quad \widetilde{y}_0 \approx Y_0 \tag{6}$$

with ρ_n the rounding error. Usually, ρ_n is proportional to the <u>unit round</u> u of the computer, and

$$|\rho_n| \le k\, u\, |Y(x_n)| \text{ or } k\, u\, |y_n| \tag{7}$$

For single precision in IEEE arithmetic, $u = 5.96 \times 10^{-8}$.

Now recall the equation satisfied by the true solution $Y(x)$:

$$Y(x_{n+1}) = Y(x_n) + h\,f(x_n, Y(x_n)) + \frac{h^2}{2} Y''(\xi_n)$$

Let $\widetilde{e}_n = Y(x_n) - \widetilde{y}_n$. Subtract (6) from this equation and proceed as before in the derivation of error formulas. This yields

$$\begin{aligned}\widetilde{e}_h(x_{n+1}) &= \left[1 + h\frac{\partial f(x_n, \zeta_n)}{\partial y}\right] \widetilde{e}_h(x_n) \\ &\qquad + \frac{h^2}{2} Y''(\xi_n) - \rho_n\end{aligned}$$

Write the last two terms as

$$h\left[\frac{h}{2} Y''(\xi_n) - \frac{\rho_n}{h}\right]$$

and identify this with $h\tau_n$ in the earlier error analysis.

Using the earlier error analysis, together with (7) for ρ_n, we get

$$\begin{aligned}|\widetilde{e}_h(x_n)| &\leq e^{(x_n-x_0)K}\,|Y_0-\widetilde{y}_0| \\ &\quad+\frac{e^{(x_n-x_0)K}-1}{K}\left[\frac{h}{2}\,\|Y''\|_\infty+\frac{k\,u\,\|Y\|_\infty}{h}\right]\end{aligned}$$

This says that as h decreases, the error will initially be proportional to h, what we denote by $O(h)$. But eventually, the error will begin to increase again as h decreases. An example of this is shown in the textbook (page 351). That example also shows that the error is much worse with chopped arithmetic than with rounded arithmetic.

AN ASYMPTOTIC ERROR FORMULA

Recall the error formula

$$e_h(x_{n+1}) = \left[1 + h\frac{\partial f(x_n, \zeta_n)}{\partial y}\right] e_h(x_n) + \frac{h^2}{2} Y''(\xi_n)$$

for the error $e_h(x_n) = Y(x_n) - y_n$. In this, ζ_n is between $Y(x_n)$ and y_n; and ξ_n is between x_n and x_{n+1}. We now replace ζ_n by $Y(x_n)$ and ξ_n by x_n, to try to find the dominant part of the error $e_h(x_n)$. This yields a new error formula

$$g_{n+1} = \left[1 + h\frac{\partial f(x_n, Y(x_n))}{\partial y}\right] g_n + \frac{h^2}{2} Y''(x_n) \quad (8)$$

with $g_n \approx e_n$, and $g_0 = 0$.

We expect g_n to be proportional to h, and therefore we write $g_n = h\delta_n$. Substituting this in (8), cancelling h, and re-arranging the equation, we obtain

$$\delta_{n+1} = \delta_n + h\left[\frac{\partial f(x_n, Y(x_n))}{\partial y}\delta_n + \frac{1}{2}Y''(x_n)\right]$$

with $\delta_0 = 0$. This is Euler's method applied to the differential equation

$$D'(x) = \frac{\partial f(x, Y(x))}{\partial y}D(x) + \frac{1}{2}Y''(x), \quad D(x_0) = 0 \tag{9}$$

Thus

$$\begin{aligned} \delta_n &\approx D(x_n) \\ g_n &\approx hD(x_n) \\ e_n &\approx hD(x_n) \end{aligned}$$

In the book, it is shown that

$$e_n = hD(x_n) + O(h^2) \tag{10}$$

This is called an *asymptotic error formula*. It tells us how the error behaves as h becomes small.

EXAMPLE

Consider

$$y' = -y^2, \quad y(0) = 1$$

The true solution is $Y(x) = 1/(1+x)$. The equation (9) becomes

$$D'(x) = \frac{-2}{1+x} D(x) + \frac{1}{(1+x)^3}, \quad D(0) = 0$$

and its solution is

$$D(x) = \frac{\log(x+1)}{(x+1)^2}$$

Thus

$$Y(x_n) - y_n = \frac{\log(x_n+1)}{(x_n+1)^2} h + O(h^2)$$

RICHARDSON EXTRAPOLATION

Since

$$Y(x) - y_h(x) \approx hD(x)$$

for any node point x, we also have

$$Y(x) - y_{2h}(x) \approx 2hD(x)$$

Combining these, we have

$$Y(x) - y_{2h}(x) \approx 2\left[Y(x) - y_h(x)\right]$$

$$Y(x) \approx y_h(x) + \left[y_h(x) - y_{2h}(x)\right] \qquad (11)$$

$$Y(x) - y_h(x) \approx y_h(x) - y_{2h}(x) \qquad (12)$$

The formula (11) is called "Richardson's extrapolation formula"; and (12) is called "Richardson's error estimate".

SYSTEMS OF EQUATIONS

Consider a system of 2 first order equations:

$$\begin{aligned} y_1' &= f_1(x, y_1, y_2), \quad y_1(x_0) = Y_{1,0} \\ y_2' &= f_2(x, y_1, y_2), \quad y_2(x_0) = Y_{2,0} \end{aligned}$$

We can apply to each equation the types of approximations used earlier with a single equation. This leads to the numerical method

$$\begin{aligned} y_{1,n+1} &= y_{1,n} + hf_1(x_n, y_{1,n}, y_{2,n}), \quad y_{1,0} = Y_{1,0} \\ y_{2,n+1} &= y_{2,n} + hf_2(x_n, y_{1,n}, y_{2,n}), \quad y_{2,0} = Y_{2,0} \end{aligned}$$

If we write the system in the vector form

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}), \quad \mathbf{y}(x_0) = \mathbf{Y}_0$$

then the numerical method can be written in the vector form

$$\mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{f}(x, \mathbf{y}_n), \quad \mathbf{y}_0 = \mathbf{Y}_0$$

MEAN-VALUE THEOREM

For the error analysis, we first need the following multivariable form of the mean-value theorem. For a function $g(y_1, ..., y_m)$ of m variables, also write it as $g(\mathbf{y})$ with

$$\mathbf{y} = [y_1, ..., y_m]^T$$

We look at what happens to the value of the function when the variables are changed. In particular,

$$g(\mathbf{y}) - g(\mathbf{z}) = \nabla g(\zeta) \cdot (\mathbf{y} - \mathbf{z})$$

with ζ some point on the line segment joining $\mathbf{y}$ and $\mathbf{z}$. For two variables, this says

$$\begin{aligned} &g(y_1, y_2) - g(z_1, z_2) \\ &\quad = \frac{\partial g(\zeta)}{\partial y_1}(y_1 - z_1) + \frac{\partial g(\zeta)}{\partial y_2}(y_2 - z_2) \end{aligned}$$

with ζ on the line joining $\mathbf{y}$ and $\mathbf{z}$.

Consider now

$$\begin{aligned} & f(x, y_1, y_2) - f(x, z_1, z_2) \\ & \quad = \frac{\partial f(x, \zeta)}{\partial y_1}(y_1 - z_1) + \frac{\partial f(x, \zeta)}{\partial y_2}(y_2 - z_2) \end{aligned}$$

We apply this with f_1 and f_2. This yields for $i = 1, 2$,

$$\begin{aligned} & f_i(x, y_1, y_2) - f_i(x, z_1, z_2) \\ & \quad = \frac{\partial f_i(x, \zeta_i)}{\partial y_1}(y_1 - z_1) + \frac{\partial f_i(x, \zeta_i)}{\partial y_2}(y_2 - z_2) \end{aligned}$$

with ζ_i on the line segment joining (y_1, y_2) and (z_1, z_2). We can write this in matrix-vector form as

$$\mathbf{f}(x, \mathbf{y}) - \mathbf{f}(x, \mathbf{z}) = F\,(\mathbf{y} - \mathbf{z})$$

with F the 2 × 2 matrix

$$F = \begin{bmatrix} \dfrac{\partial f_1(x, \zeta_1)}{\partial y_1} & \dfrac{\partial f_1(x, \zeta_1)}{\partial y_2} \\ \dfrac{\partial f_2(x, \zeta_2)}{\partial y_1} & \dfrac{\partial f_2(x, \zeta_2)}{\partial y_2} \end{bmatrix}$$

where both ζ_i and ζ_i are on the line segment joining (y_1, y_2) and (z_1, z_2).

When $\mathbf{y}$ is close to $\mathbf{z}$, the matrix F is close to the Jacobian matrix

$$\mathbf{f}_{\mathbf{y}}(x, \mathbf{y}) = \begin{bmatrix} \dfrac{\partial f_1(x, \mathbf{y})}{\partial y_1} & \dfrac{\partial f_1(x, \mathbf{y})}{\partial y_2} \\ \dfrac{\partial f_2(x, \mathbf{y})}{\partial y_1} & \dfrac{\partial f_2(x, \mathbf{y})}{\partial y_2} \end{bmatrix}$$

For an $m \times 1$ vector v and an $m \times m$ matrix A, we introduce the norms

$$\|v\|_\infty = \max_{1 \le i \le m} |v_i|$$

$$\|A\| = \max_{1 \le i \le m} \sum_{j=1}^{m} \left| A_{i,j} \right|$$

Then it can be shown that

$$\|Av\|_\infty \le \|A\| \, \|v\|_\infty$$

We apply this to

$$\mathbf{f}(x,\mathbf{y}) - \mathbf{f}(x,\mathbf{z}) = F\,(\mathbf{y}-\mathbf{z})$$

to obtain

$$\|\mathbf{f}(x,\mathbf{y}) - \mathbf{f}(x,\mathbf{z})\|_\infty \le \|F\|\,\|\mathbf{y}-\mathbf{z}\|_\infty$$

Looking at the definition of F and of $\|F\|$, we introduce

$$K = \max_{\substack{-\infty<y_1,y_2<\infty \\ x_0\le x\le b}} \left[\max_i \sum_j \left|\frac{\partial f_i(x,y_1,y_2)}{\partial y_j}\right|\right]$$

and we assume it is a finite number. Then

$$\|\mathbf{f}(x,\mathbf{y}) - \mathbf{f}(x,\mathbf{z})\|_\infty \le K\,\|\mathbf{y}-\mathbf{z}\|_\infty$$

This is the replacement to the inequality

$$|f(x,y) - f(x,z)| \le K\,|y-z|$$

with

$$K = \max_{\substack{-\infty<y<\infty \\ x_0\le x\le b}} \left|\frac{\partial f(x,y)}{\partial y}\right|$$

for working with a single equation $y' = f(x,y)$.

We can imitate the earlier proofs to show convergence, with

$$\begin{aligned} \left\|\mathbf{Y}(x_n) - \mathbf{y}_n\right\|_\infty \le{}& e^{(x_n - x_0)K} \left\|\mathbf{Y}_0 - \mathbf{y}_0\right\|_\infty \\ &+ h\frac{e^{(x_n - x_0)K} - 1}{2K} \max_{x_0 \le x \le b} \left\|\mathbf{Y}''(x)\right\|_\infty \end{aligned}$$

The other results for a single equation have similar analogues when solving systems of first order equations. The formula for an asymptotic error formula is given in the text.