Exam 2, Mathematics 4701, Section TY2 21 copies
1:25 pm–3:25 pm, March 28, 2019, IH-137
Instructor: Attila Máté[1]

1.*a*) Consider the formula

$$\bar f(x,h)=\frac{f(x+h)-f(x-h)}{2h}$$

to approximate the derivative of a function f. Assume we are able to evaluate f with about 5 decimal precision. Assume, further, that $f'''(1)\approx 1$. What is the best value of h to approximate the derivative?

Solution. We have

$$f'(x)=\frac{f(x+h)-f(x-h)}{2h}-\frac{f'''(x)}{3!}h^2+O(h^3).$$

We are able to evaluate $f(x)$ with 5 decimal precision, i.e., with an error of $5\cdot 10^{-6}$. Thus, the (absolute value of the maximum) error in evaluating $\frac{f(x+h)-f(x-h)}{2h}$ is $5\cdot 10^{-6}/h$. So the total error (roundoff error plus truncation error) in evaluating $f'(x)$ is

$$\frac{5\cdot 10^{-6}}{h}+\frac{f'''(x)}{6}h^2\approx\frac{5\cdot 10^{-6}}{h}+\frac{h^2}{6},$$

as $f'''(x)\approx 1$. The derivative of the right-hand side with respect to h is

$$-\frac{5\cdot 10^{-6}}{h^2}+\frac{h}{3}.$$

Equating this with 0 gives the place of minimum error when $h^3=15\cdot 10^{-6}$, i.e., $h\approx 0.0246$.

b) Given a certain function f, we are using the formula

$$\bar f(x,h)=\frac{f(x+h)-f(x-h)}{2h}$$

to approximate its derivative. We have

$$\bar f(1,0.1)=5.135,466,136\quad\text{and}\quad\bar f(1,0.2)=5.657,177,752$$

Using Richardson extrapolation, find a better approximation for $f'(1)$.

Solution. We have

$$f'(x)=\bar f(x,h)+c_1h^2+c_2h^4\dots$$
$$f'(x)=\bar f(x,2h)+c_1(2h)^2+c_2(2h)^4\dots$$

with some c_1, c_2, ... Multiplying the first equation by 4 and subtracting the second one, we obtain

$$3f'(x)=4\bar f(x,h)-\bar f(x,2h)-12c_2h^4+\dots.$$

That is, with $h=0.1$ we have

$$f'(x)\approx\frac{4\bar f(x,h)-\bar f(x,2h)}{3}\approx\frac{4\cdot 5.135,466,136-5.657,177,752}{3}=4.961,56$$

The function in the example is $f(x)=x\tan x$ and $f'(1)=4.982,93$.

2.*a*) The equation $e^x-x^2-4=0$ has one solution; a good approximation to this solution is 2.16. Find a way to improve this approximation by fixed-point iteration.

[1]All computer processing for this manuscript was done under Fedora Linux. $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX was used for typesetting.

Solution. Write the above equation as $x = f(x)$ with $f(x) = \ln(x^2 + 4)$. Then

$$f'(x) = \frac{2x}{x^2 + 4}.$$

It is easy to see that $|f'(x)| < 1$ for every real x. In fact, the equation $2x = x^2 + 4$ can be written as $(x-1)^2+3 = 0$. Since this equation has no real solution, one can easily conclude that $2x < x^2+4$ (since this inequality holds for negative values of x, and so it must hold for every real value of x by the Intermediate-Value Theorem for continuous functions). Thus it follows that $f'(x) < 1$ for all positive x. Hence it is clear that $|f'(x)| < 1$ for all real x, as claimed.

Starting with $x_0 = 2.16$ and taking $x_{n+1} = f(x_n)$, we have $x_1 = 2.15936$, $x_2 = 2.15904$, $x_3 = 2.15888$, $x_4 = 2.15880$, $x_5 = 2.15877$, $x_6 = 2.15875$, $x_7 = 2.15874$, $x_8 = 2.15873$, $x_9 = 2.15873$, etc. With $x = x_9$, we have $e^x - x^2 - 4 = 0.0000105251$.

It also follows that there is a number $q < 1$ such that $|f'(x)| \le q$ for every real x. The smallest such q can be found with a little effort, but the existence of such a q can easily be established as follows. We have $\lim_{x\to\pm\infty} f'(x) = 0$. Therefore, for each positive $\epsilon < 1$ there is an $R > 0$ such that $|f'(x)| < \epsilon$ whenever $|x| > R$. Let M be the maximum of $|f'(x)|$ on the interval $[-R, R]$; there is such a maximum M by the Maximum-Value Theorem. Clearly, $M < 1$, since $|f'(x)| < 1$ for every x. Take

$$q = \max(M, \epsilon).$$

As $|f'(x)| < q$ for every x, fixed point iteration will converge with any starting point.

Note. The equation $x = g(x)$ with $g(x) = \sqrt{e^x - 4}$ does not work with fixed point iteration to solve the above equation. In fact,

$$g'(x) = \frac{e^x}{2\sqrt{e^x - 4}},$$

and $g'(2.16) \approx 2.00602$.

b) State the usual sufficient condition for the fixed-point iteration to converge when solving the equation $x = f(x)$.

Solution. The condition is described be the following

Theorem. *Assume $x = c$ is a solution of the equation $x = f(x)$. Assume further that there are numbers $r > 0$ and q with $0 \le q < 1$ such that*

$$|f'(x)| \le q \quad \textit{for all} \quad x \textit{ with } c - r < x < c + r.$$

Then starting with any value $x_1 \in (c - r, c + r)$, the sequence $\{x_n\}_{n=1}^{\infty}$ defined by $x_n = f(x_{n-1})$ for $n > 2$ converges to c.

3. We want to evaluate

$$\int_0^1 e^{x^2}\, dx$$

using the composite trapezoidal rule with three decimal precision, i.e., with an error not exceeding $5 \cdot 10^{-4}$. What value of n should one use when dividing the interval $[0, 1]$ into n parts?

Solution. The error term in the composite trapezoidal rule when integrating f on the interval $[a, b]$ and dividing the interval into n parts is

$$-\frac{(b-a)^3}{12n^2} f''(\xi)$$

with some $\xi \in (a, b)$. We want to use this with $a = 0$, $b = 1$, and $f(x) = e^{x^2}$. We have

$$f''(x) = (4x^2 + 2)e^{x^2}.$$

This function is increasing on the interval $[0,1]$ (because both factors are increasing. Hence it assumes its maximum at the right end point, that is, at $x=1$. We have $f''(1)=6e$ ($\approx 16.309,691$). Since $f''(x)>0$ on $[-1,1]$, we therefore have $|f''(x)|<6e$ for $x\in(0,1)$. So, noting that $a=0$ and $b=1$, the absolute value of the error is

$$\frac{(b-a)^3}{12n^2}|f''(\xi)| = \frac{1}{12n^2}|f''(\xi)| < \frac{6e}{12n^2} = \frac{e}{2n^2}.$$

In order to ensure that this error is less than $5\cdot 10^{-4}$, we need to have $1/n^2 < 10^{-3}/e$, i.e.,

$$n > \sqrt{1000e} \approx \sqrt{2718.282} \approx 52.137.$$

So one needs to make sure that $n \geq 53$. Thus one needs to divide the interval $[0,1]$ into (at least) 53 parts in order to get the result with 4 decimal precision while using the trapezoidal rule.

4.*a*) Let h and k be numbers, and let $f(x,y)$ be a function that is differentiable sufficiently many times (so all required derivatives exist, and the order of mixed derivatives is interchangeable). Evaluate

$$\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^3 f.$$

In your answer, you may write f_x, f_{xy}, f_{yy}, ..., for the various derivatives of f.

Solution. The operators $h\partial/\partial x$ and $k\partial/\partial y$ commute, and so we can use the binomial theorem:

$$\begin{aligned}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^3 f &= \left(\left(h\frac{\partial}{\partial x}\right)^3 + 3\left(h\frac{\partial}{\partial x}\right)^2 k\frac{\partial}{\partial y} + 3h\frac{\partial}{\partial x}\left(k\frac{\partial}{\partial y}\right)^2 + \left(k\frac{\partial}{\partial y}\right)^3\right) f \\ &= h^3 f_{xxx} + 3h^2 k f_{xxy} + 3hk^2 f_{xyy} + k^3 f_{yyy}.\end{aligned}$$

b) Let $f(x,y)$ be a function that is differentiable sufficiently many times. Evaluate

$$\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right)^2 f.$$

Solution. The operators $\partial/\partial x$ and $f\partial/\partial y$ do not commute, so the binomial theorem is not valid in this case, and the square has to be evaluated by multiplying it out directly, while being careful not to interchange the order of operators. We have

$$\begin{aligned}\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right)^2 f &= \left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial y}\right) f = \frac{\partial}{\partial x}\frac{\partial}{\partial x}f + \frac{\partial}{\partial x}f\frac{\partial}{\partial y}f + f\frac{\partial}{\partial y}\frac{\partial}{\partial x}f + f\frac{\partial}{\partial y}f\frac{\partial}{\partial y}f \\ &= f_{xx} + \frac{\partial}{\partial x}ff_y + f\frac{\partial}{\partial y}f_x + f\frac{\partial}{\partial y}ff_y = f_{xx} + \frac{\partial}{\partial x}(ff_y) + f\frac{\partial}{\partial y}f_x + f\frac{\partial}{\partial y}(ff_y) \\ &= f_{xx} + (f_x f_y + ff_{yx}) + ff_{xy} + f(f_y f_y + ff_{yy}) = f_{xx} + f_x f_y + 2ff_{xy} + f(f_y)^2 + f^2 f_{yy};\end{aligned}$$

in the second line on the right, we used parentheses to make it clear that the product rule of differentiation needs to be used to obtain the next expression. Note that the expression on the right-hand side is the same as the expression for y''' in the differential equation $y'=f(x,y)$. This is not accidental, since $y''' = (d/dx)^3 y = (d/dx)^2 y' = (d/dx)^2 f(x,y)$, and we have

$$\frac{d}{dx} = \frac{\partial}{\partial x} + \frac{dy}{dx}\frac{\partial}{\partial y} = \frac{\partial}{\partial x} + f\frac{\partial}{\partial y}$$

according to the chain rule and the equation $dy/dx = f(x,y)$.

5. Consider the differential equation $y' = f(x, y)$ with initial condition $y(x_0) = y_0$. Show that, with $x_1 = x_0 + h$, the solution at x_1 can be obtained with an error $O(h^3)$ by the formula

$$y_1 = y_0 + \frac{h}{4} f(x_0, y_0) + \frac{3h}{4} f\left(x_0 + \frac{2h}{3}, y_0 + \frac{2h}{3} f(x_0, y_0)\right).$$

In other words, this formula describes a Runge-Kutta method of order 2.

Solution. Writing f, f_x, f_y for f and its derivatives at (x_0, y_0), we have

$$f\left(x_0 + \frac{2h}{3}, y_0 + \frac{2h}{3} f(x_0, y_0)\right) = f + \frac{2h}{3} f_x + \frac{2h}{3} f \cdot f_y + O(h^2).$$

according to Taylor's formula in two variables. Substituting this into the above formula for y_1, we obtain

$$y_1 = y_0 + \frac{h}{4} f + \frac{3h}{4}\left(f + \frac{2h}{3} f_x + \frac{2h}{3} f \cdot f_y + O(h^2)\right) = y_0 + hf + \frac{h^2}{2}(f_x + f f_y) + O(h^3).$$

This agrees with the Taylor expansion of y_1 (given in the preceding program) with error $O(h^3)$, showing that this is indeed a correct Runge-Kutta method of order 2.

This method is called the *optimal Runge-Kutta method of order 2*, because it can be shown that among Runge-Kutta methods of order 2 it minimizes the local truncation error.