Lecture 32 (hrs. 64,65) - December 9, 2025, 16:30 - 18:30 F3

(4.C) RUNGE-KUTTA METHODS

(4.21) Example.

In the TS(2) method the user is required to determine and realize the functions:

$$G_2(t,x) \quad \text{to compute } x(k+1)$$

and:

$$G_3(t,x) \quad \text{in the choice of } h(k)$$

In general, the task is more difficult the higher the order of the method: in the TS(p) method the user must determine and realize the functions:

$$G_2(t,x),\dots,G_p(t,x) \quad \text{to compute } x(k+1)$$

and:

$$G_{p+1}(t,x) \quad \text{in the choice of } h(k)$$

The *Runge-Kutta methods* are designed to eliminate this burden.

To introduce the *structure* of the methods, let's see, in the TS(2) method, how the computation of x(k+1) is transformed when we use a numerical estimate of the value $G_2(t,x)$.

(4.22) Remark (numerical estimate of G_2).

The value $G_2(t(k),x(k)) = y''(t(k))$ can be *estimated* with the following considerations:

(a) By definition:

$$\frac{y'(t(k) + \tau) - y'(t(k))}{\tau} \to y''(t(k)) \quad \text{as} \quad \tau \to 0$$

hence:

$$\textit{when } \tau \textit{ is small} \quad \frac{y'(t(k) + \tau) - y'(t(k))}{\tau} \quad \textit{is a good approximation of } y''(t(k))$$

(b) Since y(t) is the solution of the differential equation whose value in t(k) is x(k) we have:

$$y'(t(k)) = F(\ t(k),y(t(k))\ ) = F(\ t(k),x(k)\ )$$

and:

$$y'(t(k) + \tau) = F(\ t(k) + \tau\ ,\ y(t(k) + \tau)\ )$$

This last value is *not computable* because, given τ, the procedure does not know $y(t(k) + \tau)$. Then:

we approximate $y(t(k) + \tau)$ *with* $y(t(k)) + y'(t(k))\,\tau = x(k) + F(\,t(k),x(k)\,)\,\tau$

Overall:

using a small value of τ, *we estimate* $G_2(t(k),x(k)) = y''(t(k))$ *with*

$$\frac{F(\,t(k) + \tau\,,\ x(k) + F(\,t(k),x(k)\,)\,\tau\,) - F(\,t(k),x(k)\,)}{\tau}$$

This quantity, given τ, *is computable without using* G_2.

The estimate is *reasonable*. In fact, set $F(k) = F(\,t(k),x(k)\,)$ and consider the function of τ:

$$H(\tau) = F(\,t(k) + \tau\,,\ x(k) + F(k)\,\tau\,)$$

Since we are assuming $F(t,x)$ to have *continuous first partial derivatives*, H also has continuous first derivatives. Then:

$$H(\tau) = H(0) + H'(0)\,\tau + z(\tau)\,\tau \quad \text{where } z(\tau) \to 0 \text{ as } \tau \to 0$$

But: $H(0) = F(k)$ and

$$H'(0) = \partial_t F(\,t(k),x(k)\,) + \partial_x F(\,t(k),x(k)\,) \cdot F(\,t(k),x(k)\,) = G_2(\,t(k),x(k)\,) = y''(t(k))$$

hence:

$$H(\tau) = F(k) + y''(t(k))\,\tau + z(\tau)\,\tau$$

and:

$$\frac{H(\tau) - F(k)}{\tau} - y''(t(k)) = z(\tau) \to 0 \text{ as } \tau \to 0$$

(4.23) <u>Remark</u> (use of the numerical estimate).

In TS(2):

$$x(k+1) = x(k) + F(\,t(k),x(k)\,)\,h(k) + \frac{1}{2}\,G_2(\,t(k),x(k)\,)\,h(k)^2$$

Using $\tau = h(k)$ in the estimate of Remark (4.22) we get:

$$G_2(\,t(k),x(k)\,) = \frac{F(\,t(k) + h(k)\,,\ x(k) + F(\,t(k),x(k)\,)\,h(k)\,) - F(\,t(k),x(k)\,)}{h(k)}$$

hence (remember that $F(k) = F(\,t(k),x(k)\,)$):

$$x(k+1) = x(k) + F(k)\,h(k) + \frac{1}{2}\,[\,F(\,t(k) + h(k)\,,\ x(k) + F(k)\,h(k)\,) - F(k)\,]\,h(k)$$

$$= x(k) + \frac{1}{2}\,[\,F(k) + F(\,t(k) + h(k)\,,\ x(k) + F(k)\,h(k)\,)\,]\,h(k)$$

This procedure for calculating x(k+1) can be rewritten, in a more compact way, as follows: *the value* x(k+1) *is obtained*, after having chosen h(k), *by setting*:

- ST_1 = F(t(k),x(k))
- ST_2 = F(t(k) + h(k) , x(k) + ST_1 h(k))

and then

$$x(k+1) = x(k) + \frac{1}{2} (ST_1 + ST_2)\, h(k)$$

(4.24) Definition (two and three stages Runge-Kutta methods).

Two-stage Runge-Kutta (RK) *methods* are those in which, after choosing appropriate real numbers c_2, a_{21}, b_1 and b_2, the value x(k+1) is obtained, after having chosen h(k), by setting:

- ST_1 = F(t(k),x(k))
- ST_2 = F(t(k) + c_2 h(k) , x(k) + a_{21} ST_1 h(k))

and then

- x(k+1) = x(k) + (b_1 ST_1 + b_2 ST_2) h(k)

Three-stage Runge-Kutta (RK) *methods* are those in which, after having chosen appropriate real numbers c_2, c_3, a_{21}, a_{31}, a_{32}, b_1, b_2 and b_3, the value x(k+1) is obtained, after having chosen h(k), by setting:

- ST_1 = F(t(k),x(k))
- ST_2 = F(t(k) + c_2 h(k) , x(k) + a_{21} ST_1 h(k))
- ST_3 = F(t(k) + c_3 h(k) , x(k) + [a_{31} ST_1 + a_{32} ST_2] h(k))

and then

- x(k+1) = x(k) + (b_1 ST_1 + b_2 ST_2 + b_3 ST_3) h(k)

(4.25) Definition (order of a method as h → 0)

Let s(h) be the deviation function for the method under consideration. The integer p is called the *order of the method as* h → 0 if:

$$s^{(m)}(0) = 0 \quad \text{for } m = 0,\dots,p \quad \text{and} \quad s^{(p+1)}(0) \neq 0$$

that is, if the first term of the Taylor expansion of s(h) for h = 0 is that of order p+1:

$$s(h) = \frac{1}{(p+1)!}\, s^{(p+1)}(0)\, h^{p+1} + \dots$$

(4.26) Remark (how to find the parameters in a Runge-Kutta method).

In a multi-stage Runge-Kutta method the values of the parameters c_i, a_{ij}, b_i are determined (*not uniquely*) by the condition that: for *each* function F defining the Cauchy Problem, the order of the method for $h \to 0$, is *as high as possible*.

(4.27) Example.

Consider the two-stage Runge-Kutta method. For each k, let y(t) = y(t;x(k),t(k)). Then we have:

$$s(h) = x(k) + [b_1\, ST_1 + b_2\, ST_2(h)]\, h - y(t(k) + h)$$

Hence:

$$s^{(1)}(h) = b_1\, ST_1 + b_2\, ST_2'(h)\, h + b_2\, ST_2(h) - F[t(k) + h, y(t(k) + h)]$$

$$s^{(2)}(h) = b_2\, ST_2''(h)\, h + 2\, b_2\, ST_2'(h) - \partial_t F[t(k) + h, y(t(k) + h)] -$$

$$- F[t(k) + h, y(t(k) + h)] \cdot \partial_t F[t(k) + h, y(t(k) + h)]$$

and, since $ST_2(0) = ST_1 = F[t(k),x(k)]$:

$$s^{(0)}(0) = s(0) = x(k) - y(t(k)) = 0$$

$$s^{(1)}(0) = (b_1 + b_2 - 1)\, F[t(k),x(k)]$$

$$s^{(2)}(0) = 2\, b_2\, ST_2'(0) - \partial_t F[t(k),x(k)] - F[t(k),x(k)] \cdot \partial_t F[t(k),x(k)]$$

Then, set $F_k(h) = F[t(k) + c_2\, h\ ,\ x(k) + a_{21}\, ST_1\, h]$ and hence $ST_1 = F[t(k),x(k)] = F_k(0)$, it is:

$$ST_2'(h) = c_2\, \partial_t F_k(h) + a_{21}\, F_k(0)\, \partial_x F_k(h)$$

and:

$$s^{(2)}(0) = (2\, b_2\, c_2 - 1)\, \partial_t F_k(0) + (2\, b_2\, a_{21} - 1)\, F_k(0)\, \partial_x F_k(0)]$$

Finally:

$$s^{(1)}(0) = 0 \text{ for every } F \quad \Leftrightarrow \quad b_1 + b_2 - 1 = 0$$
$$s^{(2)}(0) = 0 \text{ for every } F \quad \Leftrightarrow \quad 2\, b_2\, c_2 - 1 = 0 \quad \text{and} \quad 2\, b_2\, a_{21} - 1 = 0$$

and the method is of order *at least* two for $h \to 0$ if and only if:

$$b_1 + b_2 = 1 \quad , \quad 2\, b_2\, c_2 = 1 \quad , \quad 2\, b_2\, a_{21} = 1$$

For example:

$$b_1 = b_2 = 1/2 \quad , \quad c_2 = a_{21} = 1 \quad \text{(Heun method)}$$ [1]

$$b_1 = 0 \quad , \quad b_2 = 1 \quad , \quad c_2 = a_{21} = 1/2 \quad \text{(modified Euler method or midpoint method)}$$

(4.28) Remark.

For a method of order p we have:[2]

- $N \to \infty$ *as* $1/\sqrt[p+1]{E}$;
- *For every* k: $ET(k) \to 0$ *as* $\sqrt[p+1]{E^p} = E^{\frac{p}{p+1}}$

1 It is the method of Example (4.23), also called ‘improved Euler method’.

2 Proof omitted.

(4.29) Remark.

Definition (4.24) extends to methods with any number of stages. Furthermore: the maximum order of a one-stage method is *one* (there is only one one-stage method of order one: the TS(1) method), of a two-stage method it is *two*, and of a three-stage method it is *three*. In general, *the maximum order of a method is less than or equal to the number of stages.*