Lecture 29 (hrs. 58,59) - December 2, 2025, 16:30 - 18:30 F3

(4.05) Definition (total error).

Let t(k) be an integration instant and let x(k) be the corresponding approximation generated by a numerical method for the approximation of the solution of the problem

$$(\S) \qquad x'(t) = F(t, x(t)) \quad , \quad x(t_0) = x_0 \quad , \quad t \in [t_0, t_f]$$

The column:

$$et(k) = x(k) - y(t(k); x_0, t_0) \in R^n$$

is called the *total error at time* t(k). The norm of et(k), which is indicated by ET(k), is *a measure of how much the method fails, at time* t(k), *in following the solution of the problem* (§).

(4.06) Definition (convergent method for $E \to 0$).

A numerical method for approximating the solution of the problem (§) is *convergent for* $E \to 0$ if: *for every* $\Delta > 0$ there exists E_* such that *if* $E < E_*$ *then* for the instants $t(0) = t_0, \ldots, t(N)$ and the columns $x(0) = x_0, \ldots, x(N)$ determined by the method we have:

$$t(N) = t_f \qquad \text{and} \qquad \max\{ ET(0), \ldots, ET(N) \} < \Delta$$

(4.07) Definition (local error).

Let t(k-1) and t(k) be two consecutive integration instants and x(k-1), x(k) be the corresponding approximations generated by a numerical method for the approximation of the solution of the problem (§). The column:

$$el(k) = x(k) - y(t(k); x(k-1), t(k-1)) \in R^n$$

is called the *local error at time* t(k). The norm of el(k), which is indicated by EL(k), is *a measure of how much the method fails, at time* t(k), *in following the solution of the differential equation* x'(t) = F(t,x(t)) *which at time* t(k-1) *passes through* x(k-1).

(4.08) Remark (relation between local and total error).

It is:

$$et(k) = x(k) - y(t(k); x_0, t_0) = (x(k) - y(t(k); x(k-1), t(k-1))) +$$

$$+ (y(t(k); x(k-1), t(k-1)) - y(t(k); x_0, t_0))$$

hence:

$$et(k) = el(k) + (y(t(k); x(k-1), t(k-1)) - y(t(k); x_0, t_0))$$

Using the notation:

$$\Delta y(t''; s, t') = y(t''; y(t'; x_0, t_0) + s, t') - y(t''; y(t'; x_0, t_0), t')$$

we may rewrite:

$$et(k) = el(k) + \Delta y(t(k);\ et(k-1),\ t(k-1))$$

The quantity Δy(t"; s, t') describes how *the differential equation* transmits at time t" the deviation, s, at time t', from the solution $y(t;\ x_0,t_0)$ of the problem (§).

(4.A) TS(1) METHOD - EXPLICIT EULER

(4.09) Hypothesis (solution regularity).

Suppose that *all* solutions of the differential equation x'(t) = F(t,x(t)) have *continuous second derivatives*.

The condition is certainly satisfied if *all* first partial derivatives of the function F(t,x) *exist* and are *continuous* functions of t and x.

(In fact: if y(t) is a solution to the differential equation we have:

$$y''(t) = (y'(t))' = (\ F(\ t,y(t)\ ))' = \partial_t F(\ t,y(t)\ ) + \partial_x F(\ t,y(t)\ ) \cdot y'(t)$$

which is continuous because $\partial_t F(\ t,x\ )$, $\partial_x F(\ t,x\ )$, y(t) and y'(t) are.)

(4.10) Definition (TS(1) method - explicit Euler).

The *TS(1) method* (or *explicit Euler method*) is defined by the following procedures.

- CHOICE of h(k). Given E > 0 and λ > 0, for each k we set:

  $$d(k) = \max\{\ \lambda\ ,\ \|\ y''(t(k);\ x(k),t(k))\ \|\ \}$$

  then:

  $$h(k) = \min\left\{\ \sqrt{\frac{2E}{d(k)}}\ ,\ t_f - t(k)\ \right\}$$

- CALCULATION of x(k+1). After choosing h(k) we set:

  $$x(k+1) = x(k) + F(\ t(k),x(k)\ )\ h(k)$$

The name of the method is a consequence of the fact that the function x(k) + F(t(k),x(k)) h is obtained by *truncating the Taylor series* of y(t(k) + h; x(k),t(k)) *at the first term* in h = 0.

(4.11) Remark (on the choice of h(k)).

Let y(t) indicate the solution y(t; x(k),t(k)) of the differential equation, and let s be the function from R to R^n defined by:

$$s(h) = x(k) + F(\ t(k),x(k)\ )\ h - y(t(k) + h)$$

Let G be the graph of y(t). The value s(h) represents the *deviation* between G and the tangent line to G at (t(k),x(k)), measured at time t(k) + h. *For* h > 0 *the quantity* s(h) *is the local error at time* t(k) + h.

Since y(t) has a continuous second derivative, s(h) *also has a continuous second derivative*. By Taylor's formula at h = 0 with Lagrange remainder, there exists a function z from R to R^n such that:

$$s(h) = s(0) + s'(0)\ h + \frac{1}{2}\ s''(0)\ h^2 + z(h)\ h^2 \qquad \text{and} \qquad z(h) \to 0 \text{ as } h \to 0$$

hence, since s(0) = x(k) - y(t(k)) = 0, s'(0) = F(t(k),x(k)) - y'(t(k)) = 0 and s"(0) = - y"(t(k)):

$$s(h) = -\ \frac{1}{2}\ y''(t(k))\ h^2 + z(h)\ h^2 \qquad \text{where} \qquad z(h) \to 0 \text{ as } h \to 0$$

If $y''(t(k)) \neq 0$ *then*:

- *When* h *is small*: $-\ \frac{1}{2}\ y''(t(k))\ h^2$ *is a good estimate of* s(h)

  (in the sense that the relative error tends to zero as h → 0)

- It is:

$$\left\|\ -\ \frac{1}{2}\ y''(t(k))\ h^2\ \right\| = E \quad \Leftrightarrow \quad h = \sqrt{\frac{2E}{\|\ y''(t(k))\ \|}}$$

The choice of h(k) guarantees that, in any case and for every $\lambda > 0$, we have:

$$\left\|\ -\ \frac{1}{2}\ y''(t(k))\ h(k)^2\ \right\| \leqslant E$$

The parameter λ is intended to prevent d(k) = 0 and also ensures that:

$$\textit{for every } k: \quad d(k) \geqslant \lambda \quad \textit{hence} \quad h(k) \leqslant \sqrt{\frac{2E}{\lambda}}$$