U.P.B. Sci. Bull., Series , Vol. 68, No. 2, 2006

ERREURS DE TRONCATURE ET D'ARRONDI

E. BRĂILEANU*

În acest articol, este prezentată metoda cu un pas a lui Euler de determinare a soluției aproximative a unui sistem diferențial. Sunt demonstrate trei rezultate adaptate de la pasul fix la pasul mobil. Se obține o mai bună evaluare a erorilor de discretizare şi rotunjire.

We present the problem of the approached resolution of the differential systems by successive approximations and we give the addaption of the demonstration for three results from the stationary to the variable step method. Thus, we obtain an amelioration of the truncation and round – off errors.

On présente le problème de la résolution approchée des systèmes différentiels par des approximations succesives et on donne trois démonstrations adaptées pour l'extension de la méthode du pas constant à la méthode du pas variable. Ainsi, on obtient une amélioration de l'erreur de troncature et d'arrondi.

Mots-clef: pas, erreurs, arrondi.

Introduction

Soit le problème Cauchy

$$\begin{cases} \boldsymbol{y}'(x) = \boldsymbol{f}(x, \boldsymbol{y}) \\ \boldsymbol{y}(a) = \boldsymbol{\alpha} \end{cases} \tag{1}$$

où $x \in [a,b]$, $\boldsymbol{y} \in \mathbf{R}^s$, $s \in \mathbf{N}^*$, $\boldsymbol{\alpha}$ est fixé dans $\mathbf{R}^s$.

Pour la fonction continue

$$\boldsymbol{f} : [a,b] \times \mathbf{R}^s \to \mathbf{R}$$

avec la propriété qu'il existe $L \geq 0$ tel que

$$\left\| \boldsymbol{f}(x, \boldsymbol{y}) - \boldsymbol{f}(x, \boldsymbol{y}^*) \right\| \leq L \left\| \boldsymbol{y} - \boldsymbol{y}^* \right\|$$

($\boldsymbol{f}$ est une fonction L – Lipschitz dans la variable vectorielle), le problème Cauchy admet une solution unique.

Par la méthode d'Euler du pas constant, la solution $\boldsymbol{y}$ est approchée par l'ensemble $(y_n)_{n=0,1,2,...}$ obtenue par les formules:

* Lecturer, Dept. of Mathematics, University "Politehnica" of Bucharest, ROMANIA

$$\begin{cases} \boldsymbol{y}_0 = \boldsymbol{\alpha} \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\,\phi(x_n, \boldsymbol{y}_n; h) \end{cases} \tag{2}$$

où $h \geq 0$ est une constante qui s'apelle le pas (constant) de la méthode; ϕ est la fonction de croissance de la méthode.

Les points $(x_n)_{n=0,1,2,\ldots}$ des formules (2) sont obtenus par le processus itératif:

$$\begin{cases} x_0 = a \\ x_{n+1} = x_n + h \end{cases} \tag{3}$$

où $x_n + h \in [a,b]$ pour chaque n considéré.

Soit $\theta : [a,b] \to (0,1]$ une fonction continue par morceaux. Pour la méthode à un pas variable, nous avons les suivants processus itératifs:

$$\begin{cases} \boldsymbol{y}_0 = \boldsymbol{\alpha} \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\,\theta(x_n)\phi(x_n; \boldsymbol{y}_n; h\theta(x_n)) \end{cases} \tag{4}$$

respectivement

$$\begin{cases} x_0 = a \\ x_{n+1} = x_n + h\theta(x_n) \end{cases} \tag{5}$$

Le vecteur $\boldsymbol{e}_n = \boldsymbol{y}_n - \boldsymbol{y}(x_n)$ est appellé l'erreur de troncature.

Pour $h \geq 0$ tel que $x + h \in [a,b]$, la fonction de croissance rélative du système est définie par

$$\Delta(x, \boldsymbol{y}(x); h\theta(x)) = \begin{cases} \dfrac{\boldsymbol{y}(x + h\,\theta(x)) - \boldsymbol{y}(x)}{h\,\theta(x)}, & h \neq 0 \\ f(x,y) & , h = 0 \end{cases} \tag{6}$$

La fonction

$$E_L(x) = \begin{cases} \dfrac{e^{Lx} - 1}{L}, & L \neq 0 \\ x & , L = 0 \end{cases} \tag{7}$$

s'appelle la fonction Lipschitz.

Pour $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_s \end{pmatrix} \in \mathbf{R}^s$, nous avons $\|\mu\| = |\mu_1| + \ldots + |\mu_s|$.

Lemme. Soit l'ensemble de nombre réels $\{t_0, t_1, \ldots, t_n, \ldots\}$ tel que $|t_{n+1}| \leq a|t_n| + b$, $(\forall)\, n = 0,1,2,\ldots$, où a et b sont des constantes positives. Alors,

$$|t_n| \le a^n |t_0| + \begin{cases} \dfrac{a^n - 1}{a-1} b, & a \ne 1 \\ nb & , a = 1 \end{cases}, \ (\forall)\, n = 0,1,2,\dots$$

La notation $\tilde{\boldsymbol{y}}$ représente un arrondissement pour $\boldsymbol{y}$ et le vecteur $\boldsymbol{r}_n = \tilde{\boldsymbol{y}}_n - \boldsymbol{y}_n$ s'appelle l'erreur d'arrondi.

Le plus petit nombre utilisé par l'ordinateur s'appelle l'unité de base. La notation x^* reprèsente le plus approché multiple de l'unité pour le nombre x.

Le vecteur $\eta_{n+1} = \left(h\tilde{\phi}(x_n, \boldsymbol{y}_n; h)\right)^* - h\tilde{\phi}(x_n, \tilde{\boldsymbol{y}}_n; h)$ s'appelle l'erreur induite, le vecteur $\rho_{n+1} = h\tilde{\phi}(x_n, \tilde{\boldsymbol{y}}_n; h) - h\phi(x_n, \tilde{\boldsymbol{y}}_n; h)$ s'appelle l'erreur inhérente et le vecteur $\varepsilon_{n+1} = \eta_{n+1} + \rho_{n+1}$ s'appelle l'erreur d'arrondi locale.

1. Méthode à un pas variable

Théorème 1. Soit $\phi : [a,b] \times \mathbf{R}^s \times [s, h_0] \to \mathbf{R}$, $\phi = \phi(x; \boldsymbol{y}; h\theta(x))$ une fonction de croissance continue et L – Lipschitz dans la variable vecteur tel que $\|\phi(x, \boldsymbol{y}; h\theta(x)) - \Delta(x, \boldsymbol{y}; h\theta(x))\| \le ch^p$, $(\forall) x \in [a,b]$, $h \in [0, h_0]$, où c et p sont des constantes positives. Alors,

$$\|\boldsymbol{e}_n\| \le \begin{cases} \dfrac{(1 + h\theta(x_n)L)^n - 1}{h\theta(x_n)L}, & L \ne 0 \\ nh^{p+1}\theta(x_n)c & , L = 0. \end{cases}$$

Démonstration. En utilisant les rélations (5), on obtient

$$\boldsymbol{y}(x_{n+1}) = \boldsymbol{y}(x_n) + \frac{\boldsymbol{y}(x_n + h\theta(x_n)) - \boldsymbol{y}(x_n)}{h\theta(x_n)} \cdot h\theta(x_n).$$

Donc, $\boldsymbol{y}(x_{n+1}) = \boldsymbol{y}(x_n) + h\theta(x_n)\Delta(x_n, \boldsymbol{y}(x_n); h\theta(x_n))$

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\theta(x_n)\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)).$$

Par la soustraction, on trouve

$$\begin{aligned} \boldsymbol{e}_{n+1} &= e_n + h\theta(x_n)[\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) - \Delta(x_n, \boldsymbol{y}(x_n); h\theta(x_n))] = \\ &= \boldsymbol{e}_n + h\theta(x_n)[\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) - \phi(x_n, \boldsymbol{y}(x_n); h\theta(x_n))] + \\ &+ h\theta(x_n) \cdot [\phi(x_n, \boldsymbol{y}(x_n); h\theta(x_n)) - \Delta(x_n, \boldsymbol{y}(x_n); h\theta(x_n))]. \end{aligned}$$

On obtient

$$\|\boldsymbol{e}_{n+1}\| \le \|\boldsymbol{e}_n\| + h\theta(x_n)L\|\boldsymbol{e}_n\| + h\theta(x_n)ch^p,$$

$$\|\boldsymbol{e}_{n+1}\| \le (1 + h\theta(x_n)L)\|\boldsymbol{e}_n\| + h^{p+1}\theta(x_n)c.$$

En appliquant la lemme, on déduit

$$\|\boldsymbol{e}_n\| \le (1+h\theta(x_n)L)^n \|\boldsymbol{e}_0\| + \begin{cases} \dfrac{(1+h\theta(x_n)L)^n - 1}{h\theta(x_n)L}, & L \ne 0 \\ nh^{p+1}\theta(x_n)c, & L = 0. \end{cases}$$

Mais $\boldsymbol{y}_0 = \boldsymbol{y}(a) = \alpha$, c'est-à-dire $\|\boldsymbol{e}_0\| = 0$.

Théorème 2. Soit $\boldsymbol{y}$ la solution du problème Cauchy et

$$\phi : [a,b] \times \mathbf{R}^s \times [0,h_0] \to \mathbf{R}$$

la fonction de croissance qui a les propriétés du théorème 1.

Soit l'ensemble $(y_n)_{n=0,1,2,\ldots}$ où

$$\begin{cases} \boldsymbol{y}_0 = \alpha \\ \boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\theta(x_n)\left[\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) + (h\theta(x_n))^q k \mu_n\right] \end{cases} \tag{8}$$

$x_n \in [a,b]$, $n = 0,1,2,\ldots$, où k, q sont des constantes positives, les vecteurs $(\mu_n)_{n=0,1,2,\ldots}$ vérifient les inégalités $\|\mu_n\| \le 1$, $(\forall) n = 0,1,2,\ldots$

Alors, $\|\boldsymbol{y}_n - \boldsymbol{y}(x_n)\| \le h^r M E_L(nhc)$, pour $x_n \in [a,b]$, $h \in [0,h_0]$, où $r = \min(p,q)$, $M = ch^{p-r} + khd^{q-r}$, $d = \sup\limits_{x \in [a,b]} \theta(x)$.

Démonstration. Nous avons

$$\boldsymbol{y}(x_{n+1}) = \boldsymbol{y}(x_n) + h\theta(x_n)\Delta(x_n, \boldsymbol{y}(x_n); h\theta(x_n)) \tag{9}$$

En utilisant (8) – (9), on trouve

$$\boldsymbol{e}_{n+1} = \boldsymbol{e}_n + h\theta(x_n)[\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) - \Delta(x_n, \boldsymbol{y}(x_n); h\theta(x_n))] + (h\theta(x_n))^q k \mu_n .$$

On obtient

$$\left\|\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) - \Delta(x_n, y(x_n); h\theta(x_n)) + (h\theta(x_n))^q k \mu_n\right\| \le$$

$$\le \|\phi(x_n, \boldsymbol{y}_n; h\theta(x_n)) - \phi(x_n, y(x_n); h\theta(x_n))\| +$$

$$+ \|\phi(x_n, \boldsymbol{y}(x_n); h\theta(x_n)) - \Delta(x_n, y(x_n); h\theta(x_n))\| + h^q (\theta(x_n))^q k \|\mu_n\| \le$$

$$\le L\|\boldsymbol{e}_n\| + ch^p + h^q d^q k ,$$

où $d = \sup\limits_{x\in[a,b]} \theta(x) \in (0,1]$.

Puis,

$$\|\boldsymbol{e}_{n+1}\| \le \|\boldsymbol{e}_n\| + (hd)\left(L\|\boldsymbol{e}_n\| + ch^p + h^q d^q k\right) =$$
$$= (1+hdL)\|\boldsymbol{e}_n\| + cdh^{p+1} + h^{q+1}d^{q+1}k.$$

En utilisant la lemme, on trouve $\left(\|\boldsymbol{e}_0\| = 0\right)$:

$$\|\boldsymbol{e}_n\| \le (1+hdL)^n\|\boldsymbol{e}_0\| + \begin{cases} \dfrac{(1+hdL)^n - 1}{hdL}\left(cdh^{p+1} + h^{q+1}d^{q+1}k\right), & L \ne 0 \\ n\left(cdh^{p+1} + h^{q+1}d^{q+1}k\right) & , L = 0 \end{cases} =$$
$$= \left(ch^p + h^q d^q k\right)\cdot\begin{cases} \dfrac{(1+hdL)^n - 1}{L}, & L \ne 0 \\ nhd & , L = 0 \end{cases} \le$$
$$\le h^r\left(c\,h^{p-r} + k\,h^{q-r}d\right)\cdot\begin{cases} \dfrac{e^{nhLd} - 1}{L}, & L \ne 0 \\ nhd & , L = 0 \end{cases} = h^r ME_L(nhd)$$

parce que $e^x > x+1, (\forall)x > 0$.

Supposons $h^* = h, a^* = a$, θ une fonction continue par morceaux, $\theta(x_n)^* = \theta(x_n)$ (donc, $x_n^* = x_n$), $y_0^* = y_0$; $\tilde{\boldsymbol{y}}_0 = \boldsymbol{y}_0$

$$\tilde{\boldsymbol{y}}_{n+1} = \tilde{\boldsymbol{y}}_n + (h\theta(x_n))\tilde{\phi}(x_n, \tilde{\boldsymbol{y}}_n, h\theta(x_n))^*,\ n = 0,1,2,\ldots$$

Théorème 3. Soit $\phi(x, y; h\theta(x))$ la fonction de croissance continue, L – Lipschitz pour la variable vectorielle. Supposons que ε soit une constante positive tel que $\|\varepsilon_n\| < \varepsilon$, $n = 0,1,2,\ldots$

Alors, l'erreur d'arrondi est bornée:

$$\|\boldsymbol{r}_n\| \le \begin{cases} \dfrac{(1+h\theta(x_n))^n - 1}{h\theta(x_n)} & , h \ne 0 \\ n\varepsilon & , h = 0 \end{cases}$$

Demonstration. Nous avons

$$\tilde{\boldsymbol{y}}_{n+1} = \tilde{\boldsymbol{y}}_n + h\theta(x_n)\phi(x_n, \tilde{\boldsymbol{y}}_n;\ h\theta(x_n)) + h\theta(x_n)\tilde{\phi}(x_n, \tilde{\boldsymbol{y}}_n; h\theta(x_n))^* -$$

$$-h\theta(x_n)\widetilde{\phi}(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))+h(\theta(x_n))\widetilde{\phi}(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))-h\theta(x_n)\widetilde{\phi}(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))=$$
$$=\widetilde{\boldsymbol{y}}_n+h\theta(x_n)\phi(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))+\eta_{n+1}+\rho_{n+1}\,.$$

Alors,

$$\widetilde{\boldsymbol{y}}_{n+1}=\widetilde{\boldsymbol{y}}_n+h\theta(x_n)\phi(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))+\boldsymbol{\varepsilon}_{n+1}$$
$$\boldsymbol{y}_{n+1}=\boldsymbol{y}_n+h\theta(x_n)\phi(x_n,\boldsymbol{y}_n;h\theta(x_n)).$$

Donc,

$$\boldsymbol{r}_{n+1}=\boldsymbol{r}_n+h\theta(x_n)[\phi(x_n,\widetilde{\boldsymbol{y}}_n;h\theta(x_n))-\phi(x_n,y_n;h\theta(x_n))]+\boldsymbol{\varepsilon}_{n+1}\,,$$
$$\|\boldsymbol{r}_{n+1}\|\le\|\boldsymbol{r}_n\|+h\theta(x_n)L\|\boldsymbol{r}_n\|+\varepsilon=(1+h\theta(x_n))\|\boldsymbol{r}_n\|+\varepsilon$$

et on applique la lemme.

Conclusion

Par la méthode du pas variable, on obtient une amélioration pour l'éstimation des erreurs de troncature et d'arrondi.

B I B L I O G R A P H I E

1. *B. Démidovitch*, Éléments de calcul numérique, Editions Mir, Moscou, 1973.
2. *P. Henrici*, Error Propagation for Difference Methods, John Willey and Sons, Inc., New-York and London, 1963.
3. *P. Henrici*, Discrete Variable Methods in Ordinary Differential Equations, John Willey and Sons, Inc., New-York and London, 1962.
4. *D. Larionescu*, Metode numerice, Editura Tehnică, Bucureşti, 1989.