REVUE FRANÇAISE D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. SÉRIE ROUGE

A. GUILLOU
J. L. SOULÉ

La résolution numérique des problèmes différentiels aux conditions initiales par des méthodes de collocation

Revue française d'informatique et de recherche opérationnelle. Série rouge, tome 3, n° R3 (1969), p. 17-44

<http://www.numdam.org/item?id=M2AN_1969__3_3_17_0>

R.I.R.O.
(3e année, No R-3, 1969, p. 17-44)

LA RESOLUTION NUMERIQUE DES PROBLEMES DIFFERENTIELS AUX CONDITIONS INITIALES PAR DES METHODES DE COLLOCATION

par A. GUILLOU et J. L. SOULÉ

Commissariat à l'Énergie Atomique, Département de Calcul Électronique

Résumé. — *On étudie de manière systématique les méthodes de résolution par pas successifs qui consistent à déterminer de proche en proche sur chaque intervalle* $[t_n, t_{n+1}]$ *une approximation polynomiale définie d'une part par des conditions d'enchaînement (égalité avec le polynôme précédent au point* t_n*: formules à pas séparés; ou égalité aux points* t_n, t_{n-1}, ... t_{n-p}*: formules à pas liés) et d'autre part par des conditions de collocation (satisfaction du système différentiel en des nœuds convenablement choisis). L'une des principales propriétés établies pour les formules implicites ainsi obtenues est que l'ordre de troncature général de chacune d'elles est identique à celui de la formule de quadrature correspondante. On étudie diverses généralisations qui conduisent à d'autres formules implicites, dont on caractérise l'erreur de troncature.*

I. UNE CLASSE FONDAMENTALE DE FORMULES IMPLICITES A PAS SEPARES

A. Définition des formules. Equivalence avec la forme traditionnelle

Nous considérons le problème général de conditions initiales pour un système différentiel non linéaire quelconque mis sous la forme canonique

$$\frac{dy}{dt} = F(y) \qquad y(0) = y_0 \tag{1}$$

où y est un vecteur de dimension quelconque, F une fonction vectorielle dont nous serons amenés à supposer qu'elle est continûment différentiable jusqu'à un ordre convenable dans un domaine approprié contenant y_0.

Pour obtenir numériquement une solution approchée de ce problème, nous choisirons une suite de valeurs $t_0 = 0$, t_1, ..., t_n, ... et nous définirons de proche en proche des polynômes (à coefficients vectoriels) $z_0(t)$, ..., $z_n(t)$, ... qui approcheront respectivement $y(t)$ sur les intervalles successifs (t_0, t_1), ..., (t_n, t_{n+1})...

$z_n(t)$ sera défini par les conditions suivantes :

1) $z_n(t_n) = z_{n-1}(t_n) \qquad z_0(0) = y_0$.

2) $z_n(t)$ est de degré maximum q.

3) $\dfrac{\mathrm{d}}{\mathrm{d}t} z_n(t_i) = F[z_n(t_i)]$ pour $t_i = t_n + (t_{n+1} - t_n)\tau_i \qquad i = 1$ à q

Les τ_i sont q valeurs distinctes caractérisant une méthode particulière.

Une définition équivalente est donnée par certains auteurs à partir de la forme intégrale

$$y(t) = y(t_n) + \int_{t_n}^{t} F[y(t')]\,\mathrm{d}t'$$

Nous allons montrer l'équivalence de notre définition avec la définition traditionnelle de certaines formules de Runge-Kutta implicites. Pour cela, nous prendrons pour inconnues les quantités :

$$k_i = \frac{\mathrm{d}}{\mathrm{d}t} z_n(t_i)$$

On sait que le polynôme $\dfrac{\mathrm{d}}{\mathrm{d}t} z_n(t)$, de degré $q - 1$, est parfaitement défini en fonction des k_i par la formule de Lagrange :

$$\frac{\mathrm{d}}{\mathrm{d}t} z_n(t) = \Sigma k_i Q_i(t)$$

De même :

$$z_n(t) = z_{n-1}(t_n) + \Sigma k_i \int_{t_n}^{t} Q_i(t')\,\mathrm{d}t'$$

et par suite

$$z_n(t_i) = z_{n-1}(t_n) + \Sigma k_j \int_{t_n}^{t_i} Q_j(t')\,\mathrm{d}t'$$

Nous aboutissons ainsi à la méthode suivante : déterminer les k_i par le système d'équations implicites

$$k_i = F\left[\bar{y}_n + h \sum_j \alpha_{ij} k_j\right] \qquad i = 1 \text{ à } q$$

puis prendre

$$\bar{y}_{n+1} = \bar{y}_n + h \sum_i \beta_i k_i$$

où l'on a posé

$$\bar{y}_n = z_{n-1}(t_n) = z_n(t_n) \qquad \bar{y}_{n+1} = z_n(t_{n+1})$$

$$h = t_{n+1} - t_n$$

$$\alpha_{ij} = \frac{\int_0^{\tau_i} (\tau - \tau_1) \dots (\tau - \tau_{j-1})(\tau - \tau_{j+1}) \dots (\tau - \tau_q)\, d\tau}{(\tau_j - \tau_1) \dots (\tau_j - \tau_{j-1})(\tau_j - \tau_{j+1}) \dots (\tau_j - \tau_q)}$$

$$\beta_i = \frac{\int_0^{1} (\tau - \tau_1) \dots (\tau - \tau_{i-1})(\tau - \tau_{i+1}) \dots (\tau - \tau_q)\, d\tau}{(\tau_i - \tau_1) \dots (\tau_i - \tau_{i-1})(\tau_i - \tau_{i+1}) \dots (\tau_i - \tau_q)}$$

Les α_{ij} et les β_i sont des constantes qui ne dépendent que des τ_i. Sous la forme obtenue, on montre par le raisonnement habituel que si F satisfait une condition de Lipschitz, la méthode définit, pour une borne supérieure assez petite des pas, une approximation unique de la solution, convergeant vers la solution lorsque cette borne tend vers zéro.

B. Propriétés de l'approximation sur un intervalle

Nous supposerons provisoirement que $z_n(t_n) = y(t_n)$, autrement dit qu'il s'agit du premier intervalle ($n = 0$), et nous supprimerons l'indice n.

[1.] Comparons d'abord $z(t_i)$ à $y(t_i)$ et $\frac{dz}{dt}(t_i)$ à $\frac{dy}{dt}(t_i)$.

Si $y(t)$ possède une dérivée d'ordre $q + 1$ continue entre 0 et h, on démontre aisément l'identité :

$$y(h\tau_i) = y_0 + h \sum_j \alpha_{ij} y'(h\tau_j) + h^{q+1} \int_0^1 K_i(\tau) y^{q(+1)}(h\tau)\, d\tau$$

avec
$$K_i(\tau) = \frac{(\tau_i - \tau)^q}{q!} - \sum_j \alpha_{ij} \frac{(\tau_j - \tau)^{q-1}}{(q-1)!}$$

(chaque puissance étant prise nulle lorsqu'elle porte sur un argument négatif).

Nous pouvons donc écrire :

$$y_i = y_0 + h\Sigma\alpha_{ij} y'_j + h^{q+1}\varepsilon_i(h)$$

avec

$$|\varepsilon_i(h)| \leqslant \int_0^1 |K_i(\tau)|\ \text{dr. sup}\ |y^{(q+1)}(h\tau)| \leqslant C$$

et $\varepsilon_i(h)$ fonction continue de h. Par ailleurs :

$$z_i = y_0 + h\Sigma\alpha_{ij} k_j,$$

soit :

$$z_i = y_i + h\Sigma\alpha_{ij}(k_j - y'_j) - h^{q+1}\varepsilon_i(h)$$

En faisant la différence des égalités $y'_i = F(y_i)$ et $k_i = F(z_i)$ et en posant $k_i - y'_i = \delta' i$, on obtient :

$$\delta'_i = F[y_i + h\Sigma\alpha_{ij}\delta'_j - h^{q+1}\varepsilon_i(h)] - F(y_i)$$

Désignons par L la constante de Lipschitz, par D le maximum des $|\delta_i'|$, par A le maximum de $\sum_j |\alpha_{ij}|$. On a l'inégalité :

$$D \leqslant L(hAD + h^{q+1}C) \qquad \text{ou} \qquad D(1 - hLA) \leqslant h^{q+1}LC$$

On en déduit que, lorsque h tend vers zéro :

$$k_i - y_i' = h^{q+1}\eta_i(h)$$

où $\eta_i(h)$ est une fonction continue en h, de partie principale $-\frac{dF}{dy}[y_0] \cdot \varepsilon_i(0)$. Réciproquement, considérons une formule de Runge-Kutta implicite de rang q, définie par des α_{ij} et des β_i *a priori* quelconques. Posons $\sum_j \alpha_{ij} = \tau_i$ et supposons que pour tout F suffisamment différentiable on ait :

$$k_i = y'(h\tau_i) + 0(h^{q+1})$$

Ceci entraîne qu'on ait

$$F(y_0 + h\Sigma\alpha_{ij}k_j) = F(y_i) + 0(h^{q+1})$$

ou encore

$$F(y_0 + h\Sigma\alpha_{ij}y_j') = F(y_i) + 0(h^{q+1})$$

c'est-à-dire finalement :

$$y_0 + h\Sigma\alpha_{ij}y'(h\tau_j) = y(h\tau_i) + 0(h^{q+1})$$

pour toute fonction $y(t)$ continûment dérivable $q + 1$ fois entre 0 et h, ou encore

$$y_0 + h\Sigma\alpha_{ij}y'(h\tau_j) = y(h\tau_i)$$

pour tout polynôme de degré au plus égal à q.

On vérifie aisément que ceci ne peut être obtenu qu'avec les formules de la classe définie ci-dessus.

Théorème I

Parmi toutes les formules de Runge-Kutta implicites de rang donné q, les formules considérées possèdent la propriété caractéristique suivante : les k_i approchent les valeurs exactes $y'(h\tau_i)$ avec une erreur qui est de l'ordre de h^{q+1}, pourvu que la solution possède une dérivée d'ordre $q + 1$ continue.

[2.] Étudions maintenant le comportement des coefficients de $z(t)$. Soit :

$$z(t) = y_0 + a_1 t + \ldots + a_q t^q$$

Les a_i s'expriment linéairement en fonction des k_i, on peut donc les décomposer en deux parties, l'une correspondant aux y_i', l'autre aux $k_i - y_i'$. En faisant appel à la théorie de l'interpolation, on en déduit aisément :

$$a_k = \frac{1}{k!}y^{(k)}(0) + 0(h^{q+1-k})$$

Lorsque h tend vers zéro, $z(t)$ converge uniformément en t dans tout intervalle fini vers le développement limité de $y(t)$ arrêté au terme en t^q, et un énoncé analogue est valable pour les dérivées successives.

Théorème II

Lorsque h tend vers zéro, chaque coefficient de $z(t)$ converge vers le coefficient correspondant du développement en série de $y(t)$ à l'origine.

[3.] Comparons enfin $z(t)$ et $y(t)$ sur l'intervalle $(0, h)$.

De $k_i = y'_i + h^{q+1}\eta_i(h)$ on déduit que $z'(t)$, entre 0 et h, diffère d'une quantité en h^{q+1} du polynôme $u'(t)$ de degré $q-1$ interpolé sur les y'_i, lequel vérifie (en vertu de la théorie de l'interpolation) :

$$y'(t) - u'(t) = \frac{y^{(q+1)}(\xi)}{q!}(t - h\tau_1) \dots (t - h\tau_q)$$

Par intégration, on obtient :

$$z(h\tau) - y(h\tau) \sim -\frac{h^{q+1}y^{(q+1)}(0)}{q!}\int_0^\tau (\tau' - \tau_1) \dots (\tau' - \tau_q)\, d\tau'$$

C. Etude de l'erreur commise aux extrémités des intervalles

La dernière formule obtenue, prise pour $\tau = 1$, montre que l'ordre de troncature est égal à q, chaque fois que l'ordre de quadrature sur les nœuds τ_i dans [0, 1] ne dépasse pas q. Mais elle ne nous renseigne pas dans le cas contraire où l'ordre de quadrature est plus élevé.

Posons

$$\delta(t) = z(t) - y(t)$$
$$\Delta(t) = z'(t) - F[z(t)]$$

on a par hypothèse $\frac{d}{dt}\delta(t) = z'(t) - y'(t) = \Delta(t) + F[z(t)] - F[y(t)]$ ce qui fait apparaître entre $\delta(t)$ et $\Delta(t)$ la relation différentielle :

$$\frac{d}{dt}\delta(t) - F[y(t) + \delta(t)] + F[y(t)] = \Delta(t)$$

Compte tenu de $\delta(0) = 0$, nous pouvons considérer la transformation qui fait passer de $\Delta(t)$ à $\delta(t)$. Supposons d'abord que $F[y(t) + \delta(t)]$ soit linéaire en $\delta(t)$:

$$F[y(t) + \delta(t)] = F[y(t)] + G(t) \cdot \delta(t)$$

Nous avons alors à inverser le système linéaire

$$\frac{d}{dt}\delta(t) - G(t) \cdot \delta(t) = \Delta(t) \qquad \delta(0) = 0$$

Selon la théorie classique des systèmes différentiels linéaires, pour évaluer $\delta(h)$ nous introduisons le résolvant matriciel $R(t)$ défini par :

$$\begin{cases} \dfrac{dR}{dt} + R(t)G(t) = 0 \\ R(h) = I \end{cases}$$

ce qui permet d'obtenir l'identité

$$[R(t) \cdot \delta(t)]' = R(t)\Delta(t)$$

et par suite

$$\delta(h) = \int_0^h R(t)\Delta(t)\,dt$$

Nous appliquons à la fonction $R(t)\Delta(t)$ la formule de quadrature interpolatoire sur les nœuds $h\tau_i$. Puisque $\Delta(t)$ s'annule par construction pour chacune de ces valeurs, l'intégrale est égale au reste de la formule, c'est-à-dire à

$$h^{r+1} \int_0^1 K(\tau)[R\Delta]^{(r)}(h\tau)\,d\tau$$

où $K(\tau)$ ne dépend que des τ_i et où la dérivée $[R\Delta]^{(r)}$ est supposée exister entre 0 et h, r étant l'ordre de la formule de quadrature sur les nœuds τ_i. R est continue ainsi que ses r premières dérivées s'il en est ainsi de F; de même Δ, à travers z dont nous avons établi plus haut la régularité. Dans ces conditions, on voit que la formule a un ordre de troncature égal à r.

Nous pouvons lever l'hypothèse de linéarité. En effet, dans le cas général nous pourrons écrire :

$$\frac{d}{dt}\delta(t) - \frac{dF}{dy}[y(t)] \cdot \delta(t) = \Delta(t) + 0(\delta^2)$$

Puisque $\delta(t) = 0(h^{q+1})$, nous aurons :

$$\delta(h) = \int_0^h R(t)\Delta(t)\,dt + 0(h^{2q+3})$$

or on sait que r ne peut pas dépassser $2q$, de sorte que le terme complémentaire est d'un ordre négligeable.

On peut aussi établir cette extension en posant :

$$F[y(t) + \delta(t)] = F[y(t)] + \frac{dF}{dy}[y(t) + \rho\delta(t)] \cdot \delta(t)$$

$$= F[y(t)] + G(t) \cdot \delta(t)$$

autrement dit en faisant entrer $\delta(t)$ dans la définition de $G(t)$ et par suite de $R(t)$.

Théorème III

Toute formule de la classe considérée a un ordre de troncature identique à son ordre de quadrature.

Pour préciser l'expression de l'erreur et sa partie principale, nous remplaçons $R\Delta$ par $(R\delta)'$, soit :

$$\delta(h) = h^{r+1} \int_0^1 K(\tau)[R\delta]^{(r+1)}(h\tau)\, d\tau$$

En posant $A = \int_0^1 K(\tau)\, d\tau$ (constante d'erreur de la formule de quadrature)

$$\delta(h) \sim Ah^{r+1}[R\delta]^{(r+1)}(0)$$

on obtiendra une expression plus explicite en fonction de $y(t)$ et $F(y)$ en appliquant à $R\delta$ la formule de Leibnitz et en remarquant que :

1° $$\delta^{(s)}(0) = \begin{cases} 0 & \text{pour} \quad s \leqslant q \\ -y^{(s)}(0) & \text{pour} \quad s > q \end{cases}$$

2° les dérivées successives de R s'obtiennent par récurrence en fonction de celles de $\dfrac{dF}{dy}$ en utilisant la définition

$$\frac{dR}{dt} + R\frac{dF}{dy} = 0$$

Enfin, on peut étudier la propagation de l'erreur d'un intervalle à l'autre selon l'analyse habituelle des formules à pas séparés. Ceci permet en particulier de préciser les propriétés de l'approximation sur un intervalle, compte tenu du fait que la valeur initiale est erronée.

D. Formules particulières. Extension aux nœuds confondus

La classe étudiée dans cette première partie a déjà été considérée par Ceschino et Kuntzmann [1]. Toutefois le théorème III ne semble avoir été énoncé que dans des cas particuliers :

— Gauss-Legendre [1], [2],

— Radau (avec nœud à l'origine) [3].

Par ailleurs, Céa [4] a étudié le cas des nœuds équidistants (Newton-Coates). On peut faire aussi le rapprochement avec la méthode de Lanczos, dans le cas où les nœuds sont les zéros d'un polynôme de Tchebichev.

Les formules d'ordre élevé (Gauss, Radau, Lobatto) paraissent *a priori* les plus intéressantes, dans l'esprit de la méthode.

En faisant intervenir explicitement les dérivées de F, on peut étendre de façon naturelle la classe considérée au cas où certains nœuds sont confondus. Par exemple, dire que deux nœuds sont confondus en τ_i, c'est compléter la condition de collocation par :

$$\frac{d^2}{dt^2} z(h\tau_i) = \frac{dF}{dy}[z(h\tau_i)] \cdot \frac{d}{dt} z(h\tau_i)$$

On obtient de cette façon des formules qui ne sont plus du type Runge-Kutta implicite, mais auxquelles la théorie précédente s'applique. Parmi elles figure évidemment la méthode de Taylor (q nœuds confondus à l'origine), des méthodes reversibles particulières ne faisant intervenir que les deux extrémités de l'intervalle, etc.

E. Application aux systèmes différentiels à coefficients constants

Après diagonalisation, l'étude de ces systèmes revient essentiellement à celle de l'équation

$$\frac{dy}{dt} = \lambda y \qquad y_0 = 1$$

Nous allons obtenir dans ce cas une expression intéressante de $z(h)$. Par construction $z'(t)$ et $\lambda z(t)$ coïncident aux q points $h\tau_i$. Mais λz est de degré q et z' de degré $q - 1$, de sorte que cela entraîne :

$$\lambda z - z'(t) = C(t - h\tau_1) \dots (t - h\tau_q) = CQ(t)$$

Pour déterminer C, il suffit de multiplier cette égalité par $e^{-\lambda t}$ et d'intégrer de zéro à l'infini (dans une direction convenable), en utilisant $z(0) = 1$:

$$1 = C \int_0^\infty e^{-\lambda t} Q(t)\, dt$$

Maintenant, en intégrant de h à l'infini, on obtient :

$$e^{-\lambda h} z(h) = C \int_h^\infty e^{-\lambda t} Q(t)\, dt$$

D'où le résultat :

$$z(h) = \frac{\displaystyle\int_0^\infty e^{-\lambda t} Q(t+h)\, dt}{\displaystyle\int_0^\infty e^{-\lambda t} Q(t)\, dt} = \frac{\displaystyle\int_0^\infty e^{-\rho\tau} \bar{Q}(\tau+1)\, d\tau}{\displaystyle\int_0^\infty e^{-\rho\tau} \bar{Q}(\tau)\, d\tau}$$

où l'on a posé $\bar{Q}(\tau) = (\tau - \tau_1) \dots (\tau - \tau_q)$, $\rho = \lambda h$

Ceci met en évidence une expression de la fraction rationnelle en ρ qui est fournie par la méthode comme approximation de la solution exacte e^{ρ}. (Les deux intégrales sont des polynômes en $\frac{1}{\rho}$ de degré $q+1$, sans terme constant.) Cette fraction, dont le numérateur et le dénominateur en ρ sont au plus de degré q, a un développement limité à l'origine qui coïncide avec celui de e^{ρ} jusqu'au terme en ρ^r ($r \geqslant q$).

Réciproquement, considérons une fraction rationnelle $\Phi(\rho) = \dfrac{A(\rho)}{B(\rho)}$ vérifiant les conditions suivantes :

1° degré $(A) = p \leqslant r \qquad A(0) = 1$

degré $(B) = p' \leqslant r \qquad B(0) = 1$

2° Le développement en série entière de $\Phi(\rho)$ s'accorde avec celui de e^{ρ} au moins jusqu'au terme en ρ^r inclusivement.

Divisons le numérateur et le dénominateur par ρ^{r+1}, on aura :

$$\Phi(\rho) = \frac{\bar{A}\left(\frac{1}{\rho}\right)}{\bar{B}\left(\frac{1}{\rho}\right)} = \frac{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{P}(\tau)\,d\tau}{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{Q}(\tau)\,d\tau} \qquad \text{degré } (\bar{P}) = \text{degré } (\bar{Q}) = r$$

$$1 - e^{-\rho}\Phi(\rho) = \frac{\displaystyle\int_0^{1} e^{-\rho\tau}\bar{P}(\tau - 1)\,d\tau}{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{Q}(\tau)\,d\tau} + \frac{\displaystyle\int_0^{\infty} e^{-\rho\tau}[\bar{Q}(\tau) - \bar{P}(\tau - 1)]\,d\tau}{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{Q}(\tau)\,d\tau}$$

Or cette quantité doit se comporter comme $c\rho^{r+1}$ pour $\rho \to 0$. Par hypothèse, c'est bien le cas pour le premier terme, quels que soient $\bar{P}$ et $\bar{Q}$. Mais le second numérateur ne restera borné que s'il est nul.

On obtient ainsi l'expression générale des fractions rationnelles considérées

$$\Phi(\rho) = \frac{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{Q}(\tau + 1)\,d\tau}{\displaystyle\int_0^{\infty} e^{-\rho\tau}\bar{Q}(\tau)\,d\tau} \qquad \text{degré } (\bar{Q}) = r$$

Pour avoir $p < r$, il faut et il suffit que les $r - \rho$ derniers termes de $\bar{P}$ soient nuls, autrement dit que $\bar{Q}(\tau)$ contienne $(\tau - 1)^{r-p}$ en facteur. De même pour avoir $p' < r$, il faut et il suffit que $\bar{Q}(\tau)$ contienne $\tau^{r-p'}$ en facteur. Ceci n'est possible que si $p + p' \geqslant r$. A la limite, pour $p + p' = r$, on a nécessaire-

ment $\bar{Q}(\tau) = \tau^p(\tau - 1)^p$ et on obtient une expression générale des fractions figurant dans la table de Padé :

$$\Phi_{p,q}(\rho) = \frac{\int_0^\infty e^{-\rho\tau}(\tau + 1)^p \tau^{p'} \, d\tau}{\int_0^\infty e^{-\rho\tau}\tau^p(\tau - 1)^{p'} \, d\tau}$$

Par ailleurs, nous n'avons pas supposé que $A(\rho)$ et $B(\rho)$ étaient nécessairement premiers entre eux. Ceci a pour conséquence qu'une même fraction irréductible vérifiant $p < r$ et $p' < r$ (r : ordre effectif) a plusieurs représentations distinctes de la forme indiquée.

Par exemple, pour les fractions de la table de Padé, on peut prendre

$$\bar{Q}(\tau) = c_0\tau^p(\tau - 1)^{p'} + c_1 \frac{d}{d\tau}[\tau^p(\tau - 1)^{p'}] + \ldots + c_s \frac{d^s}{d\tau^s}[\tau^p(\tau - 1)^{p'}]$$

avec $s = \min[p, p']$; $c_0, \ldots, c_s$ quelconques.

Nous sommes maintenant en mesure de fixer les τ_i pour obtenir une fraction rationnelle donnée, du moins lorsqu'il lui correspond un $\bar{Q}(\tau)$ dont toutes les racines sont réelles. Par exemple, nous pouvons obtenir de plusieurs façons chaque fraction de Padé; la plus économique nécessite de prendre

$$q = \max(p, p') \qquad \bar{Q}(\tau) = \frac{d^s}{d\tau^s}[\tau^p(\tau - 1)^{p'}] \qquad s = \min(p, p').$$

En prenant pour s toutes les valeurs décroissantes jusqu'à zéro, on augmente q jusqu'à $p + p'$ et on obtient chaque fois des systèmes de nœuds compris entre 0 et 1 (avec des nœuds multiples en 0 ou 1).

Par exemple, pour $p = p'$, on obtient d'abord la formule de Gauss à p nœuds, puis la formule de Lobatto à $p + 1$ nœuds, et en dernier lieu la formule avec p nœuds confondus à l'origine et p nœuds confondus à l'extrémité.

Les formules de Radau s'obtiennent avec $|p - p'| = 1$, $s = \min(p, p')$.

Avec $p' = 0$, on obtient la méthode de Taylor (polynômes); avec $p = 0$, on obtient la méthode de Taylor inverse (inverses de polynômes).

Pour compléter l'étude des systèmes à coefficients constants, il faut considérer ce qui se passe pour les chaînes du type :

$$\frac{dy_1}{dt} = \lambda y_1, \qquad y_1(0) = 1$$

$$\frac{dy_2}{dt} = \lambda y_2 + y_1, \qquad y_2(0) = 0 \qquad \left(\frac{d^2y_2}{dt^2} - 2\lambda\frac{dy_2}{dt} + \lambda^2 y_2 = 0\right)$$

on établit aisément que dans ce cas $z_2(t) = \dfrac{d}{d\lambda} z_1(t)$.

Notons enfin l'expression générale de la fraction rationnelle obtenue par les formules de Runge-Kutta implicites :

$$\Phi(\rho) = \frac{\begin{vmatrix} 1 & -\rho\beta_1 & \dots & -\rho\beta_q \\ 1 & 1-\rho\alpha_{11} & \dots & -\rho\alpha_{1q} \\ 1 & -\rho\alpha_{21} & \dots & -\rho\alpha_{2q} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 1 & -\rho\alpha_{q1} & \dots & 1-\rho\alpha_{qq} \end{vmatrix}}{\begin{vmatrix} 1-\rho\alpha_{11} & \dots & -\rho\alpha_{1q} \\ & & \\ & & \\ & & \\ -\rho\alpha_{q1} & \dots & 1-\rho\alpha_{qq} \end{vmatrix}}$$

F. Formules modifiées

La démonstration du théorème III suggère une variante intéressante des formules considérées jusqu'ici. Supposons que les τ_i soient tous différents de un. Après avoir abouti à la formule

$$\delta(h) = \int_0^h R(t)\Delta(t)\,\mathrm{d}t$$

nous pouvons appliquer à $R\Delta$ la formule de quadrature interpolatoire sur $q+1$ nœuds : les $h\tau_i$ et h. Profitant de $R(h) = I$, nous obtenons

$$\delta(h) = h\gamma\Delta(h) + h^{r'+1}\int_0^1 K(\tau)[R\Delta]^{(r')}(h\tau)\,\mathrm{d}\tau$$

où r' est cette fois l'ordre de quadrature sur les nœuds $\tau_1, \dots, \tau_q, 1$. γ est un coefficient connu et $\Delta(h) = z'(h) - F[z(h)]$ peut s'évaluer par un calcul explicite.

Ceci nous conduit à approcher $y(h)$ non plus par $z(h)$ mais par $z(h) - h\gamma\Delta(h)$, ce qui nous permet d'atteindre l'ordre r'.

On peut présenter autrement la modification effectuée en disant qu'on a remplacé $z(t)$ par un autre polynôme d'approximation de degré $q+1$, soit $\bar{z}(t)$, défini de la façon suivante :

$$\bar{z}(0) = y_0 \qquad \frac{\mathrm{d}\bar{z}}{\mathrm{d}t} = F(z) \qquad \text{pour } t = h\tau_1, \dots, h\tau_q, h.$$

Le polynôme $\bar{z}(t)$ a tous ses coefficients qui tendent vers ceux du développement limité de $y(t)$ jusqu'à l'ordre t^{q+1} lorsque h tend vers zéro.

On voit qu'on atteint de cette façon le même ordre de troncature que si on avait effectué la collocation directement sur les $q+1$ nœuds $h\tau_1, \ldots, h\tau_q, h$. L'avantage est que le système d'équations implicites à résoudre comporte une équation en moins.

La partie principale de l'erreur est cependant différente, du fait que $\delta^{(s)}(0) = -y^{(s)}(0)$ à partir de $s = q+1$ et non de $s = q+2$.

Le cas particulier des formules de Radau à point terminal et des formules de Lobatto a déjà été étudié par Butcher [3].

Dans le cas où F est linéaire à coefficients constants, les conditions imposées à $\dfrac{d\bar{z}}{dt}$ entraînent $\dfrac{d\bar{z}}{dt} = F(z)$. En particulier pour l'équation $y' = \lambda y (y_0 = 1)$, on aura : $\dfrac{d\bar{z}}{dt} = \lambda z = z'(t) + cQ(t)$ et par suite :

$$\bar{z}(t) = z(t) + c\int_0^t Q(t')\,dt' \qquad \lambda\bar{z} - \bar{z}'(t) = \lambda c \int_0^t Q(t')\,dt'$$

ce qui conduit à :

$$\bar{z}(h) = \frac{\displaystyle\int_0^\infty e^{-\lambda t} Q_1(t+h)\,dt}{\displaystyle\int_0^\infty e^{-\lambda t} Q_1(t)\,dt} \qquad Q_1(t) = \int_0^t Q(t')\,dt'$$

L'identité :

$$\int_0^t \frac{dt}{t-1} \frac{d^p}{dt^p} t^{p+1}(t-1)^{p+1} = \frac{1}{p+2} \frac{d^p}{dt^p} t^{p+2}(t-1)^p$$

montre que les formules de Lobatto modifiées conduisent à un décalage dans la table de Padé qui est d'une case à 45° vers le haut et la droite.

Il en est de même pour les formules de Radau à point terminal modifiées; autrement dit, on retombe sur les fractions obtenues avec les formules de Radau à point initial (non modifiées).

On pourrait obtenir de nouvelles formules modifiées en partant de formules à point final multiple. Par exemple, lorsqu'on sait calculer le jacobien $\dfrac{dF}{dy}$, on peut construire une formule complètement explicite d'ordre 4 sur le point initial et le point final.

G. Systèmes différentiels d'ordre supérieur

Considérons par exemple l'équation différentielle d'ordre m :

$$\frac{d^m y}{dt^m} = f\left[y, \frac{dy}{dt}, ..., \frac{d^{m-1}y}{dt^{m-1}}\right], \qquad y(0) = y_0, ..., \frac{d^{m-1}y}{dt^{m-1}}(0) = y_0^{(m-1)}$$

on la ramène classiquement à la forme canonique (1) en introduisant le vecteur $\left[y(t), y_1 = \frac{dy}{dt}, ..., y_{m-1} = \frac{d^{m-1}}{dt^{m-1}} y\right]$. Par conséquent, les formules que nous avons étudiées s'appliquent à ce problème. On vérifie qu'elles conduisent à approcher y et ses $m-1$ premières dérivées par des polynômes de même degré q et qu'on obtient le même ordre de troncature pour toutes ces fonctions.

Dans un tel cas, cependant, on peut appliquer différemment le principe de collocation : on approchera y par un polynôme z de degré maximum $q + m - 1$ défini par :

$$z(0) = y_0, ..., \frac{d^{m-1}z}{dt^{m-1}}(0) = y_0^{(m-1)}$$

$$\frac{d^m z}{dt^m}(t_i) = f\left[z(t_i), \frac{dz}{dt}(t_i), ..., \frac{d^{m-1}}{dt^{m-1}} z(t_i)\right] \qquad \text{en } q \text{ points}$$

Les propriétés des formules obtenues s'étudient de façon analogue. Les z_i approchent les y_i avec une erreur en $0(h^{q+m})$, les $\frac{dz}{dt}(t_i)$ approchent les $\frac{dy}{dt}(t_i)$ avec une erreur en $0(h^{q+m-1})$, etc. Seules les dérivées d'ordre $m-1$ et m sont approchées aux points t_i avec une erreur en $0(h^{q+1})$.

L'étude de l'erreur de troncature peut se faire à nouveau à l'aide du résolvant matriciel, en notant en particulier qu'une seule composante de Δ est non nulle. Si r est l'ordre de quadrature sur les nœuds utilisés, l'ordre de troncature pour y est $\max[r, q + m - 1]$, celui pour y' est $\max[r, q + m - 2]$, etc.

On peut traiter de même le cas de plusieurs équations différentielles couplées, d'ordres égaux ou différents [6].

II. UNE CLASSE FONDAMENTALE DE FORMULES IMPLICITES A PAS LIES

A. Définition des formules. Equivalence avec la forme traditionnelle

Pour obtenir numériquement une solution approchée du problème (1), nous choisirons maintenant une suite de valeurs équidistantes

$$t_{-p} = -ph, ..., t_1, t_0 = 0, t_1, ..., t_n = nh, ...$$

et des approximations $z_0, z_{-1}, \ldots, z_{-p}$ de $y_0, y_1, \ldots, y_{-p}$; puis nous définirons de proche en proche des polynômes $z_0(t), \ldots, z_n(t), \ldots$ qui approcheront respectivement $y(t)$ sur les intervalles successifs $(t_0, t_1), \ldots (t_n, t_{n+1}), \ldots$ $z_n(t)$ sera défini par les conditions suivantes :

1) $z_n(t_n - jh) = z_{n-1}(t_n - jh) \qquad j = 0 \text{ à } p \qquad z_0(-jh) = z_{-j}$

2) $z_n(t)$ est de degré maximum $p + q$

3) $$\frac{\mathrm{d}}{\mathrm{d}t} z_n(t_n + h\tau_i) = F[z_n(t_n + h\tau_i)] \qquad i = 1 \text{ à } q$$

Les τ_i sont q valeurs distinctes caractérisant (avec p) une méthode particulière. Certains des τ_i peuvent prendre l'une des valeurs $0, \ldots, -p$. De toutes façons, nous supposerons le système $(-j, \tau_i)$ régulier [1].

Nous allons montrer l'équivalence de notre définition avec la définition traditionnelle de certaines formules hybrides (ou mixtes) implicites. Pour cela, nous prenons comme inconnues les quantités :

$$k_i = \frac{\mathrm{d}}{\mathrm{d}t} z_n(t_n + h\tau_i) \qquad i = 1 \text{ à } q$$

Puisque le système $(-j, \tau_i)$ est régulier (voir annexe), le polynôme $z_n(t)$ est défini de façon unique en fonction des k_i et des $z_{n-1}(t_n - jh)$. Donc, on a des expressions de la forme :

$$z_n(t_n + h\tau_i) = \sum_j \alpha_{ij} z_{n-1}(t_n - jh) + h \sum_k \gamma_{ik} k_k$$

Les k_i sont donc déterminés par le système d'équations implicites :

$$k_i = F\left[\sum_j \alpha_{ij} z_{n-j} + h \sum_k \gamma_{ik} k_k\right] \qquad i = 1 \text{ à } q$$

puis

$$z_{n+1} = \Sigma \delta_j z_{n-j} + h \Sigma \beta_i k_i$$

L'existence et la convergence pour h tendant vers zéro s'étudient par les méthodes usuelles.

B. Propriétés de l'approximation sur un intervalle. Erreur de troncature

1. Pour étudier ces propriétés on supposera provisoirement que :

$$z_n(t_n - jh) = y(t_n - jh) \qquad j = 0 \text{ à } p$$

et on supprimera l'indice n ($t_n = 0$).

---

(1) Voir annexe.

Par un raisonnement analogue à celui utilisé pour les formules à pas séparés on obtient les relations :

$$k_i - y_i' = F[y_i + h\Sigma\gamma_{ij}(k_j - y_j') - h^{p+q+1}\varepsilon_i(h)] - F[y_i]$$

d'où l'on déduit $k_i - y_i' = h^{p+q+1}\eta_i(h)$ et l'on établit que cette propriété est caractéristique parmi les formules implicites à pas liés faisant intervenir p valeurs antérieures de la solution et q valeurs de la dérivée. On en déduit aussi le comportement des coefficients de $z(t)$ pour h tendant vers zéro, et celui de $z(h\tau) - y(h\tau)$.

2. Si l'on pose à nouveau

$$\delta(t) = z(t) - y(t) \qquad \Delta(t) = z'(t) - F[z(t)]$$

on a toujours $\dfrac{\mathrm{d}}{\mathrm{d}t}\delta(t) - F[y(t) + \delta(t)] + F[y(t)] = \Delta(t)$

avec

$$\delta(-jh) = 0 \qquad j = 0 \text{ à } p$$

En introduisant le résolvant matriciel $R(t)$, on obtient une fonction $R(t) \cdot \delta(t)$ qui possède les propriétés suivantes :

a) elle est égale à $\delta(h)$ pour $t = h$,

b) elle s'annule pour $t = -jh$ $j = 0$ à p,

c) sa dérivée vaut $R(t)\Delta(t)$ dans le cas linéaire et $R(t)\Delta(t) + 0(h^{2p+2q+2})$ dans le cas général, de sorte qu'elle est nulle ou négligeable (1) pour $t = h\tau_i$ $i = 1$ à q,

d) ses dérivées d'ordre supérieur sont régulières avec celles d'y.

En appliquant à $R(t)\delta(t)$ la formule de quadrature à pas liés donnant la valeur pour $t = h$ en fonction de la valeur pour $t = -jh$ et de la dérivée pour $t = h\tau_i$, on obtient :

$$\delta(h) = 0(h^{2p+2q+3}) + h^{r+1}\int K(\tau)[R\delta]^{(r+1)}(h\tau)\,\mathrm{d}\tau$$

soit

$$\delta(h) \sim ch^{r+1}[R\delta]^{(r+1)}(0)$$

avec

$$\delta^{(s)}(0) = \begin{cases} 0 & \text{pour} \quad s \leqslant p + q \\ -y^{(s)}(0) & \text{pour} \quad s > p + q \end{cases}$$

D. Etude de la stabilité

L'étude de la propagation de l'erreur s'effectue selon l'analyse habituelle des formules à pas liés. En particulier, on peut appliquer la théorie de la stabilité de Dahlquist. Selon cette théorie, la stabilité asymptotique

(1) Voir Annexe.

pour h tendant vers zéro est fonction des racines de l'équation caractéristique :

$$s^{p+1} - \Sigma\delta_j s^{p-j} = 0$$

et plus précisément du fait qu'il existe ou non des racines de module supérieur à un.

E. Formules particulières

Les formules dans lesquelles τ_i ne peut prendre que les valeurs $-p, \ldots, -1$, 0 et éventuellement $+1$ sont classiques. Au contraire, les formules hybrides implicites (à points auxiliaires) ne semblent pas avoir été considérées jusqu'ici, sauf en tant que formules de quadrature à pas liés (Krylov).

Celles qui paraissent les plus intéressantes sont les formules à ordre de troncature maximum qui ont été définies plus haut. Nous compléterons en Annexe les indications de Krylov concernant la stabilité de ces formules. On pourrait, ici encore, étendre la théorie au cas des nœuds confondus.

F. Formules modifiées

On peut à nouveau bâtir des formules modifiées faisant intervenir de façon purement explicite la valeur $\tau_i = 1$.

Le cas où il n'y a pas de τ_i différent de $-j$ est classique. Les autres cas sont apparemment nouveaux.

III. NOUVELLES CLASSES ETENDUES

A. Définition d'une nouvelle classe de formules à pas séparés. Equivalence avec la forme traditionnelle

1. Résoudre le problème (1) consiste à rechercher deux fonctions $y(t)$ et $y'(t)$ liées par les deux relations :

$$y'(t) = F(y(t)) \qquad y(t) = y_0 + \int_0^t y'(t')\,dt'$$

Nous définirons une nouvelle classe de formules en approchant respectivement $y'(t)$ et $y(t)$ sur $[0, h]$ par les polynômes $k(t)$ et $l(t)$ de même degré $q - 1$, définis par deux groupes de q conditions :

$$k(h\tau_i) = F[1(h\tau_i)] \qquad i = 1 \text{ à } q \qquad \tau_i \text{ distincts}$$

$$1(h\xi_j) = y_0 + \int_0^{h\xi_j} k(t')\,dt' \qquad j = 1 \text{ à } q \qquad \xi_j \text{ distincts}$$

Les $2q$ valeurs τ_i et ξ_j caractérisent une formule particulière. Pour finir, on prendra $\bar{y}(h) = y_0 + \int_0^h k(t)\,dt$

Posons $Q(t) = (t - h\xi_1) \ldots (t - h\xi_q) = h^q \bar{Q}\left(\frac{t}{h}\right)$.

Le polynôme $l(t)$ et le polynôme $z(t) = y_0 + \int_0^t k(t')\,dt'$ sont imposés égaux en q points. Or $z(t)$ est de degré q. Il en résulte la condition équivalente : $l(t) = z(t) - CQ(t)$ où C est le coefficient de t^q dans $z(t)$.

Sous cette forme, on peut étendre la classe considérée en prenant pour $\bar{Q}(\tau)$ un polynôme quelconque de degré q. Par contre, les τ_i continueront à être supposés distincts.

2. L'équivalence avec la forme traditionnelle des formules de Runge-Kutta implicites s'établit comme suit. Le coefficient du terme en t^{q-1} dans $k(t)$ s'exprime linéairement en fonction des k_i; il en est donc de même pour C, à savoir

$$C = \frac{1}{h^{q-1}} \Sigma \gamma_i k_i \qquad \gamma_i = \frac{1}{q} \frac{1}{(\tau_i - \tau_1) \ldots (\tau_i - \tau_{i-1})(\tau_i - \tau_{i+1})(\tau_i - \tau_q)}$$

D'autre part $z(h\tau_i) = y_0 + k\Sigma\alpha_{ij}k_j$.

D'où

$$l_i = l(h\tau_i) = y_0 + h\Sigma\alpha'_{ij}k_j \qquad \text{avec} \qquad \alpha'_{ij} = \alpha_{ij} - \gamma_j \bar{Q}(\tau_i)$$

3. La classe de formules ainsi définie contient comme cas particuliers d'une part la classe fondamentale définie au chapitre I (prendre $\xi_i = \tau_i$ pour tout i), d'autre part les formules modifiées définies en $I - F$. Pour obtenir ces dernières, il faut prendre :

$$Q(t) = q \int_0^t (t' - h\tau_1) \ldots (t' - h\tau_{q-1})\,dt' \qquad (\text{et } \tau_q = 1)$$

autrement dit $Q(0) = 0 \cdot \frac{dQ}{dt}(h\tau_i) = 0$ pour $i = 1$ à $q - 1$.

En effet puisque $\frac{dl}{dt} = \frac{dm}{dt} - C\frac{dQ}{dt}$ on aura

$$l(0) = y_0 \qquad \text{et} \qquad \frac{dl}{dt}(h\tau_i) = k(h\tau_i) \qquad i = 1 \text{ à } q - 1$$

B. Propriétés de l'approximation sur un intervalle

1. La quantité $h\Sigma\gamma_i y'_i$ est égale au coefficient du terme du plus haut degré du polynôme d'interpolation des y'_i, multiplié par qh^q. Cette quantité est donc d'ordre h^q. Par suite, on a :

$$y_i = y_0 + h\Sigma\alpha'_{ij}y'_j + h^q\varepsilon_i(h)$$

$$k_i - y'_i = F[y_i + h\Sigma\alpha'_{ij}(k_j - y'_j) - h^q\varepsilon_i(h)] - F[y_i]$$

D'où

$$k_i = y'_i + h^q\eta_i(h)$$

$$l_i = y_i + h^q\xi_i(h)$$

Réciproquement, toute formule de Runge-Kutta implicite de rang q vérifiant l'une ou l'autre de ces conditions appartient à la classe considérée (à moins qu'elle ne se ramène à une formule du chapitre I de rang $q - 1$ par confusion de deux des nœuds).

La démonstration de cette réciproque est basée sur le fait que $\Sigma\gamma_i y'_i$ est à un facteur près la seule combinaison linéaire d'ordre h^{q-1} pour toute fonction régulière (à condition que les τ_i soient distincts).

2. De ce qui précède, on déduit que si on pose :

$$l(t) = a_0 + a_1 t + \ldots + a_{q-1}t^{q-1}$$

$$k(t) = b_0 + b_1 t + \ldots + b_{q-1}t^{q-1}$$

$$z(t) = y_0 + b_0 t + \ldots + b_{q-1}\frac{t^q}{q}$$

on a

$$a_i = \frac{1}{i!}y^{(i)}(0) + 0(h^{q-i})$$

$$\frac{b_{i-1}}{i} = \frac{1}{i!}y^{(i)}(0) + 0(h^{q+1-i})$$

Par suite, il y a convergence uniforme de chacun des polynômes vers un développement limité de $y(t)$ ou $y'(t)$.

$z(t) - y(t)$ est de l'ordre de h^{q+1} dans tout l'intervalle $[0, h]$, mais la partie principale de l'écart n'est pas aussi simple que pour les formules du chapitre I.

3. Considérons momentanément l'ensemble des formules de Runge-Kutta implicites (ou explicites) de rang q, en excluant toutefois les formules à nœuds confondus :

$$k_i = F[y_0 + h\Sigma\alpha_{ij}k_j] \qquad \tau_i = \sum_j \alpha_{ij} \qquad (\tau_i \text{ distincts})$$

on peut les interpréter à l'aide des polynômes $k(t)$ et $l(t)$ de degré $q-1$:

$$k(h\tau_i) = k_i \qquad l(h\tau_i) = l_i = y_0 + h\Sigma\alpha_{ij}k_j$$
$$k(t) = b_0 + b_1 t + \ldots + b_{q-1}t^{q-1}$$
$$l(t) = a_0 + a_1 t + \ldots + a_{q-1}t^{q-1}$$

Au lieu de caractériser une méthode particulière par la matrice α_{ij}, on peut la caractériser par les τ_i et par la matrice qui fait passer des b_i aux a_i. Il est facile de voir qu'elle est de la forme :

$$a_0 = y_0 + \gamma_{01}b_1h^2 + \ldots + \gamma_{0,q-1}h^q b_{q-1}$$
$$a_1 = b_0 + \gamma_{11}b_1h + \ldots + \gamma_{1,q-1}h^{q-1}b_{q-1}$$
$$a_2 = \gamma_{21}b_1 + \ldots + \gamma_{2,q-1}h^{q-2}b_{q-1}$$
$$a_{q-1} = \gamma_{q-1,1}b_1h^{3-q} + \ldots + \gamma_{q-1,q-1}hb_{q-1}$$

(La classe de formules considérées dans ce chapitre est caractérisée par $\gamma_{i,i-1} = \frac{1}{i}$ pour $i = 1$ à $q-1$; $\gamma_{i,q-1} \neq 0$ pour $i = 0$ à $q-1$; $\gamma_{i,j} = 0$ autrement.)

Sous cette forme, on peut caractériser les formules « régulières », à savoir celles qui conduisent à des polynômes dont tous les coefficients convergent lorsque h tend vers zéro : ce sont celles qui vérifient $\gamma_{ij} = 0$ pour $i-j > 1$. (En particulier : toutes les formules de rang $\leqslant 3$ sont régulières; les formules explicites de rang $\geqslant 4$ ne sont pas régulières.)

C. Etude de l'erreur commise aux extrémités des intervalles

1. Nous considérons d'abord l'application à deux types particuliers de systèmes différentiels.

Dans l'application aux quadratures, l'erreur commise est l'erreur de quadrature sur les nœuds $h\tau_i$.

Dans l'application aux systèmes linéaires à coefficients constants, la condition $k(h\tau_i) = F[l(h\tau_i)]$ avec même degré $q-1$ pour k et l entraîne l'identité $k(t) = F[l(t)]$ pour tout t. Par conséquent $z'(h\xi_j) = F[z(h\xi_j)]$. Autrement dit, on obtient la même approximation que par les formules du premier chapitre en remplaçant les nœuds $h\tau_i$ par les nœuds $h\xi_j$ (et on peut transposer tous les résultats de I — E). On pourra donc parler respectivement de l'ordre de quadrature r et de l'ordre de « troncature linéaire » r' de la formule.

2. Posons

$$\delta(t) = z(t) - y(t)$$
$$\Delta(t) = k(t) - F[l(t)]$$
$$\varepsilon(t) = l(t) - z(t) = -CQ(t)$$

on a

$$\frac{d}{dt}\delta(t) - F[y(t) + \delta(t) + \varepsilon(t)] + F[y(t)] = \Delta(t) \qquad \delta(0) = 0$$

Dans le cas où F est linéaire

$$\frac{d}{dt}\delta(t) - G(t)\delta(t) = G(t)\varepsilon(t) + \Delta(t)$$

$$[R(t)\delta(t)]' = R(t)G(t)\varepsilon(t) + R(t)\Delta(t)$$

$$\delta(h) = \int_0^h R(t)G(t)\varepsilon(t)\,dt + \int_0^h R(t)\Delta(t)\,dt$$

$$= h^{r'+1}\int_0^1 K'(\tau)[RG\varepsilon]^{(r')}(h\tau)\,d\tau + h^{r+1}\int_0^1 K(\tau)[R\Delta]^{(r)}(h\tau)\,d\tau$$

Dans le cas général où F n'est pas linéaire, on devra ajouter à Δ un terme en $0(\delta + \varepsilon)^2$, soit une contribution à $\delta(h)$ en $0(h^{2q+1})$, ou modifier la définition de G.

Théorème : Toute formule de la classe considérée a un ordre de troncature égal au minimum de son ordre de quadrature et de son ordre de troncature linéaire.

Pour préciser la partie principale de l'erreur on remplacera $R\Delta$ par $[R\delta]' + R'\varepsilon$. Le premier terme de $\delta(h)$ est équivalent à :

$$-A'h^{r'+1}[R'\varepsilon]^{(r')}(0) = A'h^{r'+1}C_{r'}^{q}R^{(r'+1-q)}(0)y^{(q)}(0)$$

Le second terme est équivalent à :

$$Ah^{r+1}\{[R\delta]^{(r+1)}(0) - C_r^q R^{(r+1-q)}(0)y^{(q)}(0)\}$$

avec

$$\begin{cases} \delta^{(s)}(0) = 0 & \text{pour} \quad s \leqslant q \\ \delta^{(s)}(0) = -y^{(S)}(0) & \text{pour} \quad s > q \end{cases}$$

D. Nouvelle classe de formules à pas liés

On peut définir de façon analogue une nouvelle classe de formules à pas liés en faisant intervenir simultanément le polynôme $z(t)$ de degré $p + q$ et le polynôme $l(t)$ de degré $p + q - 1$, liés par les conditions suivantes :

$$z_n(-jh) = z_{n-1}(-jh) \qquad j = 0 \text{ à } p$$

$$z_n'(h\tau_i) = F[l_n(h\tau_i)] \qquad i = 1 \text{ à } q$$

$$z_n(h\xi_j) = l_n(h\xi_j) \qquad j = 1 \text{ à } p + q$$

ou encore en définissant $z_n(t)$ par :

$$z_n(-jh) = z_{n-1}(-jh) \qquad j = 0 \text{ à } p$$

$$z_n'(h\tau_i) = F[z_n(h\tau_i) - CQ(h\tau_i)] \qquad i = 1 \text{ à } q$$

C étant le coefficient de t^{p+q} dans $z_n(t)$.

La théorie de ces formules s'établit selon des lignes analogues.

ANNEXE. FORMULES DE QUADRATURE A PAS LIES

A. Généralités

Les formules de quadrature ordinaires (formules à pas séparés) sont des formules d'approximation du type

$$y(t_n + h) = y(t_n) + h\sum_i \beta_i y'(t_n + h\tau_i) + R$$

Les formules de quadrature à pas liés sont des formules qui utilisent plusieurs valeurs déjà calculées de y; à pas constant, elles sont donc du type

$$y(t_n + h) = \sum_{j=0}^{p} \alpha_j y(t_n - jh) + h\sum_{i=1}^{q} \beta_i y'(t_n + h\tau_i) + R$$

Krylov [5, chap. 16] a donné à leur sujet un certain nombre de résultats, que nous reprenons et étendons ci-dessous.

Définition 1. — Nous dirons que le système de nœuds $(t_0, t_1, ..., t_p; u_1, ..., u_q)$ est régulier si tout polynôme de degré au plus égal à $p + q$ vérifiant

$$P(t_0) = ... = P(t_p) = P'(u_1) = ... = P'(u_q) = 0$$

est identiquement nul.

Théorème I. — Une condition nécessaire et suffisante pour que le système $(t_0, t_1, ..., t_p, u_1, ..., u_q)$ soit régulier est qu'il vérifie

$$\begin{vmatrix} 1 & t_0 & t_0^2 & \dots & t_0^{p+q} \\ \dots & & & & \\ 1 & t_p & t_p^2 & & t_p^{p+q} \\ 0 & 1 & 2u_1 & & (p+q)u_1^{p+q-1} \\ \dots & & & & \\ 0 & 1 & 2u_q & & (p+q)u_q^{p+q-1} \end{vmatrix} = 0$$

Démonstration : le système d'équations homogènes satisfait par les coefficients de $P(t)$ doit être régulier.

Théorème II. — Un ensemble de conditions suffisantes pour que le système $(t_0, t_1, ..., t_p, u_1, ..., u_q)$ soit régulier est le suivant : les t_j sont distincts, les u_i sont distincts, chaque u_i est soit égal à un des t_j, soit extérieur au plus petit intervalle contenant les t_j.

Démonstration : Si $P(t)$ s'annule en $p + 1$ points distincts t_j, $P'(t)$ s'annule au moins en p points situés entre les t_j (et distincts des t_j). Donc $P'(t)$, avec les hypothèses formulées, s'annule au moins en $p + q$ points distincts; comme il est de degré au plus $p + q - 1$, il est identiquement nul, et P aussi.

Théorème III. — Si le système (t_j, u_i) est régulier, il existe un polynôme et un seul de degré $p + q$ au plus prenant des valeurs données en t_j et tel que sa dérivée prenne des valeurs données en u_i.

Corollaire. — *Définition 2.* — Si (t_j, u_i) est régulier, il existe, quel que soit t, une formule unique du type :

$$P(t) = A_0P(t_0) + ... + A_pP(t_p) + B_1P'(u_1) + ... + B_qP'(u_q)$$

qui est satisfaite exactement par tout polynôme de degré $p + q$ au plus. Nous l'appellerons « formule interpolatoire en t sur (t_j, u_i) ».

Définition 3. — Nous appellerons « ordre » d'une formule interpolatoire la valeur maximum de r telle que la formule soit satisfaite exactement par tout polynôme de degré r au plus ($r \geqslant p + q$).

Théorème IV. — Si la formule interpolatoire en t sur (t_j, u_i) est d'ordre r, toute fonction $y(t)$ dérivable $r + 1$ fois satisfait identiquement la formule :

$$y(t) = A_0y(t_0) + ... + A_py(t_p) + B_1y'(u_1) + ... + B_qy'(u_q) + \int K(t')y^{(r+1)}(t')\,dt'$$

où $K(t')$ est un noyau continu, nul en dehors du plus petit intervalle contenant t, t_j, u_i, avec $\int K(t')\,dt' = C \neq 0$.

Démonstration : on combine les développements de Taylor de $y(t_0)$, ..., $y'(u_q)$ autour du point t, avec des restes intégraux en $y^{(r+1)}$.

Corollaire. — Si la formule interpolatoire en t sur (t_j, u_i) est d'ordre r, et si la fonction $y(t)$ est $r + 1$ fois continûment différentiable dans un voisinage de ξ, le reste de la formule interpolatoire en $\xi + ht$ sur $(\xi + ht_j, \xi + hu_i)$ tend vers zéro comme $Ch^{r+1}y^{(r+1)}(\xi)$.

B. Formules d'ordre maximum

Nous allons maintenant étudier une classe de formules qui généralise les formules ordinaires de Gauss, Radau et Lobatto.

Théorème V. — L'ordre r d'une formule interpolatoire, pour $t \neq t_j$, ne peut pas dépasser $p + q + s$, si s est le nombre de nœuds u_i qui sont distincts des t_j et de t.

Démonstration : Il suffit de construire un contre-exemple, à savoir un polynôme de degré $p + q + s + 1$ vérifiant $P(t) \neq 0$ et $P(t_j) = P'(u_i) = 0$. Dans le cas où aucun des u_i n'est égal à t, on prendra :

$$P(t') = (t' - t_0) \dots (t' - t_p)(t' - u_1) \dots (t' - u_q)(t' - v_1) \dots (t' - v_s)$$

où $v_1, \dots, v_s$ sont les nœuds u_i distincts des t_j; de cette façon chaque facteur $t' - u_i$ figure en fait deux fois.

Dans le cas où par exemple $u_q = t$, on posera d'abord

$$Q(t') = (t' - t_0) \dots (t' - t_p)(t' - u_1) \dots (t' - u_{q-1})(t' - v_1) \dots (t' - v_s)$$

puis on prendra $P(t') = Q(t')[Q'(t)(t' - t) - Q(t)]$.

Théorème VI. — Si la formule

$$P(t) = A_0 P(t_0) + \dots + A_p P(t_p) + B_1 P'(u_1) + \dots + B_q P'(u_q)$$

(où $t \neq t_j$), est satisfaite exactement par tout polynôme de degré $p + q + s$ (où s est le nombre de nœuds u_i qui sont distincts des t_j et de t), le système (t_j, u_i) est régulier.

Corollaire : la formule écrite est alors la formule interpolatoire (unique) sur (t_j, u_i) et son ordre est exactement égal à $p + q + s$.

Démonstration : Soit $v_1, \dots, v_s$ les u_i distincts de t_j et t. Considérons un polynôme Q de degré $p + q$ au plus vérifiant $Q(t_j) = 0$, $Q'(u_i) = 0$. Si P est de la forme $P = Q'S$ avec degré $S \leqslant s$, on a par hypothèse :

$$Q(t)S(t) = \Sigma A_j Q(t_j)S(t_j) + \Sigma B_i[Q'(u_i)S(u_i) + Q(u_i)S'(u_i)]$$

Pour $S \equiv 1$, cela entraîne $Q(t) = 0$. Puis, d'autre part :

$$0 = \sum_{i=1}^{s} B_i Q(v_i) S'(v_i) \qquad \text{degré } S' \leqslant s - 1$$

D'où nécessairement $B_i Q(v_i) = 0$.

Aucun B_i n'est nul sinon le v_i correspondant ne figurerait pas dans la formule, donc $Q(v_i) = 0$. Il en résulte que Q est identiquement nul.

Théorème VII. — t étant donné, si l'on choisit arbitrairement $p + 1$ valeurs t_j distinctes, situées du même côté de t, puis si l'on sélectionne arbitrairement parmi les t_j et $tq - s$ premières valeurs u_i, il est toujours possible de compléter ce système par s valeurs u_i de façon à obtenir un système régulier (t_j, u_i) fournissant en t une formule interpolatoire d'ordre $p + q + s$. Il y a exactement autant de choix possibles distincts (à une permutation d'indices près) que de façons de répartir les s valeurs entre les $p + 1$ intervalles formés par $t, t_0, \dots, t_p$.

Pour chacune de ces formules, le noyau K du terme d'erreur est de signe constant; le coefficient C est de module minimum lorsqu'on place les s nœuds u_i complémentaires entre t et t_0 (valeur t_i la plus proche de t).

Démonstration : Supposons que le système (t_j, u_i) fournisse en t une formule interpolatoire d'ordre maximum $p + q + s$. Considérons le système obtenu en adjoignant aux t_j les s valeurs de v_i (c'est-à-dire les u_i distincts de t et t_j) et en laissant les u_i inchangés. Les nouvelles valeurs de p, q, s sont $p' = p + s$, $q' = q$, $s' = 0$. Donc $p' + q' + s' = p + q + s$. L'ordre maximum possible d'une formule en t sur le nouveau système est le même que sur l'ancien. Or, nous connaissons une formule atteignant cet ordre, à savoir la même formule que la formule initiale, interprétée comme une formule où les A_j relatifs aux nouveaux t_j sont nuls. Réciproquement, à partir d'une formule interpolatoire sur un système vérifiant $s' = 0$, on obtient une formule interpolatoire d'ordre maximum pour le système obtenu en supprimant des t_j ceux qui ont des A_j nuls.

On est donc conduit à étudier dans quelles conditions on peut ajouter les mêmes s nœuds aux t_j et aux u_i de telle façon que le nouveau système soit régulier et que les A_j relatifs aux nouveaux nœuds soient tous nuls. Cette étude est basée sur les résultats suivants de la théorie de l'interpolation.

Considérons $m + n + 1$ points distincts x_i; il existe un polynôme et un seul de degré $2m + n + 1$ au plus prenant des valeurs données en $x_i (i = 0$ à $m + n)$ et tel que sa dérivée prenne des valeurs données en $x_i (i = 0$ à $m)$. Si ces valeurs sont celles que prend une fonction $y(x)$ et sa dérivée $y'(x)$, le coefficient de x^{2m+n+1} dans ce polynôme est par définition la « différence divisée » de $y(x)$ sur les nœuds $(x_0, x_0, \ldots, x_m, x_m, x_{m+1}, \ldots, x_{m+n})$
on a :

$$D(x_0, x_0, \ldots, x_m, x_m, x_{m+1}, \ldots, x_{m+n})y = \sum_{i=0}^{m+n} A_i y(x_i) + \sum_{i=0}^{m} B_i y'(x_i)$$

avec :

$$(x_i - x_0)^2 \ldots (x_1 - x_{i-1})^2 (x_i - x_{i+1})^2 \ldots (x_i - x_m)^2 (x_i - x_{m+1}) \ldots$$

$$(x_i - x_{m+n}) A_i = -\left[\frac{2}{x_i - x_0} + \ldots + \frac{2}{x_i - x_{i-1}} + \frac{2}{x_i - x_{i+1}} \ldots \right.$$

$$\left. + \frac{2}{x_i - x_m} + \frac{1}{x_i - x_{m+1}} + \ldots + \frac{1}{x_i - x_{m+n}} \right] \qquad i \leqslant m$$

$$(x_i - x_0)^2 \ldots (x_i - x_{i-1})^2 (x_i - x_{i+1})^2$$

$$\ldots (x_i - x_m)^2 (x_i - x_{m+1}) \ldots (x_i - x_{m+n}) B = 1$$

$$(x_i - x_0)^2 \ldots (x_i - x_m)^2 (x_i - x_{m+1})$$

$$\ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots (x_i - x_{m+n}) A_i = 1 \qquad i > m$$

Si y possède une dérivée d'ordre $2m + n + 1$ continue sur le plus petit intervalle I contenant les x_i on a

$D(x_0, x_0, \ldots, x_m, x_m, x_{m+1}, \ldots, x_{m+n})$

$$y = \frac{y^{(2m+n+1)}(\xi)}{(2m+n+1)!} = \int \bar{K}(x') y^{(2m+n+1)}(x')\, dx'$$

où $\bar{K}(x')$ est un noyau indépendant de y nul, à l'extérieur de I et aux extrémités de I et positif à l'intérieur de I ; ξ est un point intérieur de I, dépendant de y.

La relation $\sum_{i=0}^{m+n} A_i P(x_i) + \sum_{i=0}^{m} B_i P'(x_i) = 0$ est (à un facteur près) la seule relation satisfaite simultanément par tous les polynômes de degré $2m + n$ au plus.

On obtient donc un système régulier en prenant comme t_j tous les x_i sauf un, et comme u_i les $m + 1$ premiers x_i, sous la condition nécessaire et suffisante que le A_{i0} relatif au x_{i0} supprimé ne soit pas nul. Cette condition est toujours remplie si $i_0 > m$, ou si x_{i0} est un des points extérieurs.

Dans ces conditions, on obtient pour $t = x_{i0}$ la formule interpolatoire d'ordre $2m + n$:

$$y(t) = -\sum_{i \neq i_0} \frac{A_i}{A_{i0}} y(x_i) - \sum \frac{B_i}{A_{i0}} y'(x_i) + R$$

avec

$$R = \frac{y^{(2m+n+1)}(\xi)}{A_{i0}(2m+n+1)!} = \int \frac{\bar{K}(x')}{A_{i0}} y^{(2m+n+1)}(x')\, dx'$$

Supposons maintenant que tous les x_i soient fixés, sauf s d'entre eux parmi les $m + 1$ premiers (distincts de t) ; identifions $q = m + 1$ et $p = m + n - 1 - s$. Les formules annoncées dans le théorème VI seront établies dans la mesure où on peut choisir les s valeurs x_i restantes de telle façon que leurs coefficients A_i soient nuls (avec $A_{i0} \neq 0$).

Ceci conduit au système de s équations :

$$\sum_{i=0}^{m} \frac{2}{x_k - x_i} + \sum_{i=m+1}^{m+n} \frac{1}{x_k - x_i} = 0$$

où $i \neq k$ et k prend s valeurs.

Il suffit de noter que les solutions de ce système sont les valeurs qui rendent stationnaire la fonction :

$$2 \sum_K \sum_{i=0}^{m} \log |x_k - x_i| + \sum_K \sum_{i=m+1}^{m+n} \log |x_k - x_i|$$

et d'utiliser le fait que dans chaque domaine défini par un ordre des x_i, cette fonction est convexe comme chacun de ses termes, et tend vers $-\infty$ quand

un des arguments s'annule, pour démontrer l'existence d'un maximum unique dans chacun de ces domaines qui est compact (c'est-à-dire pour lequel aucun x_k n'est un point extérieur).

Enfin, on voit que si t est la valeur la plus grande, le coefficient C et le noyau K seront négatifs ou positifs selon que t figure ou ne figure pas parmi les nœuds u_i. La valeur de C se déduit en effet de celle de A_{i0} dont on a une expression explicite. Cette expression permet aussi de montrer aisément que $|C|$ est le plus petit possible quand les nœuds libres sont le plus près possible de t.

Dans le cas $p = 0$, on retrouve les formules de Gauss, Radau et Lobatto, en imposant 0, 1 ou 2 nœuds u_i en t_0 ou t.

C. Résultats numériques (1)

Nous nous plaçons dans le cas où les premiers u_i coïncident avec les t_j, $j = 0, ..., p$ formant une grille de nœuds « doubles » de pas h, soit

$$t = h \quad , \quad t_j = u_{j+1} = -(p-j)h \qquad 0 \leqslant j \leqslant p$$

Nous déterminons les s nœuds u_i suivants qui assurent une formule d'ordre maximum, ces nœuds « auxiliaires » étant situés sur l'intervalle $[0, h]$ soit

$$u_{p+1+j} = h\tau_j \qquad 1 \leqslant j \leqslant s \qquad 0 < \tau_j < 1$$

1) $q = p + s + 2$ avec $u_q = h$.

Nous avons des formules du type Lobatto.

a) $s = 1$ 1 nœud auxiliaire.

Les formules sont stables jusqu'à 6 nœuds doubles.

Nombre de nœuds doubles	τ_1
1	0,5
2	0,5773502
3	0,6180339
4	0,6444328
5	0,6634465
6	0,6780375

b) $s = 2$ deux nœuds auxiliaires.

Les formules sont stables jusqu'à 12 nœuds doubles.

(1) Ces résultats sont dus à M^me^ Bruyère, CEA, Cadarache.

NOMBRE DE NŒUDS DOUBLES	τ_1	τ_2
1	0,2763932	0,7236068
2	0,3283560	0,7613687
3	0,3600704	0,7813830
4	0,3829082	0,7945223
5	0,4006886	0,8040093
6	0,4151913	0,8115140
7	0,4273964	0,8175115
8	0,4379029	0,8225060
9	0,4471037	0,8267600
10	0,4552711	0,8304474
11	0,4626008	0,8336892
12	0,4692387	0,8365723

2) $q = p + s + 1$.

Nous avons des formules du type Radau.

a) $s = 1$ un nœud auxiliaire.

Les formules sont stables jusqu'à 4 nœuds doubles.

NOMBRE DE NŒUDS DOUBLES	τ_1
1	0,6666666
2	0,7403124
3	0,7743178
4	0,7948015

b) $s = 2$ deux nœuds auxiliaires.

Les formules sont stables jusqu'à 9 points antérieurs.

NOMBRE DE NŒUDS DOUBLES	τ_1	τ_2
1	0,3550510	0,8449489
2	0,4207572	0,8717519
3	0,4589917	0,8851468
4	0,4855023	0,8936140
5	0,5055280	0,8996167
6	0,5214634	0,9041746
7	0,5346000	0,9077978
8	0,5457111	0,9107740
9	0,5552946	0,9132799

BIBLIOGRAPHIE

[1] F. CESCHINO et J. KUNTZMANN, *Problèmes différentiels de conditions initiales*, Dunod, (1963).

[2] J. C. BUTCHER, *Implicit Runge-Kutta processes, Mathematics of computation*, **18** (1964), 50-64.

[3] J. C. BUTCHER, *Integration processes based on Radau quadrature formulas, Mathematics of computation*, **18** (1964), 233-244.

[4] J. CEA, Équations différentielles : Méthode d'approximation discrète *p*-implicite, *Revue Française de traitement de l'information*, **8** (1965) 179-194. CEA J. *p*-implicit methods for ordinary differential equations, Proceedings of IFIP Congress 1965, 563-564, Spartan books (1966).

[5] KRYLOV, *Approximate calculation of integrals*, Mac Millan (1962).

[6] J. H. VERNER, *The order of some implicit Runge-Kutta methods*, Numérish Mathematics, **13**, 14-23, (1969).