

J. DUMONTET

J. VIGNES

Détermination du pas optimal dans le calcul des dérivées sur ordinateur

RAIRO. Analyse numérique, tome 11, n° 1 (1977), p. 13-25

http://www.numdam.org/item?id=M2AN_1977__11_1_13_0

© AFCET, 1977, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Analyse numérique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

DÉTERMINATION DU PAS OPTIMAL DANS LE CALCUL DES DÉRIVÉES SUR ORDINATEUR (*)

par J. DUMONTET ⁽¹⁾ et J. VIGNES ⁽²⁾

Communiqué par J. H. WILKINSON

Résumé. — Calculer sur ordinateur la dérivée d'une fonction, c'est mettre en œuvre une formule d'approximation dépendant du pas h de dérivation. Le résultat obtenu est toujours entaché d'une double erreur. La première due à la formule d'approximation elle-même, est d'autant plus élevée que le pas h est grand. La seconde due à l'arithmétique à précision limitée de la machine est d'autant plus élevée que le pas h est petit.

Nous présentons dans cet article une méthode numérique qui permet de déterminer le pas de dérivation optimal h_p , et ainsi d'obtenir la meilleure approximation de la dérivée.

1. INTRODUCTION

Soit f une fonction de $A \subset \mathbf{R}$ dans \mathbf{R} supposée continue et admettant une dérivée première continue. Sur le plan algébrique, la dérivée de f en un point $x_0 \in A$ est définie par

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0+h) - f(x_0-h)}{2h}. \quad (1)$$

La formule (1) ne donne la valeur exacte de la dérivée, qu'au passage à la limite, c'est-à-dire, lorsque $h \rightarrow 0$. Sur le plan du calcul numérique, ne pouvant passer à la limite et calculant avec un pas $h \neq 0$ on n'obtient qu'une approximation de la dérivée qui est donnée par

$$f'_a(x_0) = \frac{f(x_0+h) - f(x_0-h)}{2h} \quad (h \neq 0), \quad (2)$$

La valeur exacte $f'(x_0)$ de la dérivée au point x_0 et son approximation $f'_a(x_0)$ fournie par le calcul numérique sont liées par

$$f'(x_0) = f'_a(x_0) + e_m. \quad (3)$$

L'erreur e_m , dite erreur de méthode, est d'autant plus faible que le pas h , appelé pas de dérivation, est plus petit.

(*) Reçu juin 1975.

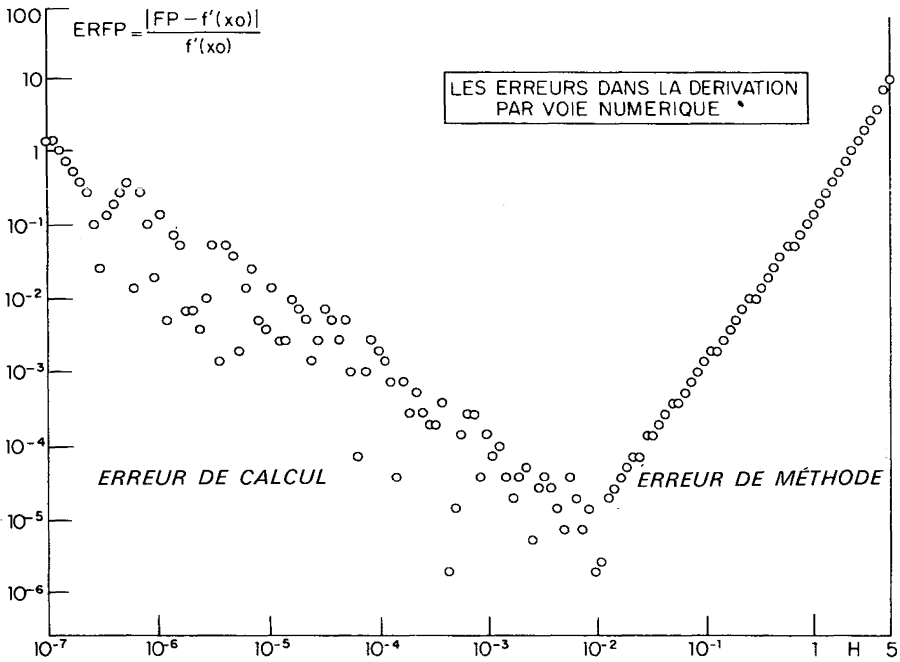
(1) Faculté des Sciences de Poitiers, Service Calcul et Informatique.

(2) Institut de Programmation, Université Pierre-et-Marie-Curie, Paris.

Pour obtenir donc sur ordinateur la meilleure approximation de $f'(x_0)$ on pourrait croire qu'il suffit de choisir le pas h aussi petit que possible. Hélas ! il n'en est rien car l'ordinateur, travaillant dans le domaine du discret limité, possède une arithmétique à précision limitée, arithmétique qui elle-même n'est qu'une approximation de l'arithmétique exacte. Notons que l'arithmétique approchée des ordinateurs et la propagation des erreurs ont fait l'objet de plusieurs études [3, 6 à 11]. Ainsi, l'approximation $f'_a(x_0)$ de la dérivée $f'(x_0)$ sera, outre l'erreur de méthode, entachée aussi d'une erreur de calcul.

Cette erreur de calcul est d'autant plus forte que le pas de dérivation h est petit. Ainsi les deux erreurs entachant le résultat jouent-elles en sens contraire. Le but de cette étude est donc, après avoir évalué chacune de ces erreurs, de déterminer le pas optimal h_p qui minimise l'erreur globale. Pour bien montrer l'influence de ces deux erreurs, faisons calculer sur un ordinateur ayant une précision relative de $3 \cdot 10^{-7}$ la dérivée de la fonction $f = e^x$ au point $x_0 = 0.5$.

La figure montre la variation, en fonction du pas de dérivation h , de l'erreur relative de la dérivée numérique FP calculée par (2).



L'erreur relative $ERFP$ est définie par

$$ERFP = \frac{|FP - f'(x_0)|}{f'(x_0)}. \quad (4)$$

Il ressort clairement de la figure 1 que :

- lorsque le pas h est grand, l'erreur de méthode est telle que la valeur de la dérivée numérique est non significative,
- lorsque le pas h est petit, c'est l'erreur de calcul due à la soustraction de deux quantités voisines toutes deux entachées d'erreur qui rend non significative la valeur de la dérivée numérique,
- lorsque le pas h est voisin de $8 \cdot 10^{-3}$ l'erreur globale est minimale et l'on obtient une valeur de la dérivée numérique qui est une très bonne approximation de $f'(x_0)$.

2. LES DIVERSES ERREURS LORS DU CALCUL D'UNE DÉRIVÉE SUR ORDINATEUR

Lorsque l'on veut obtenir sur ordinateur la valeur de la dérivée $f'(x_0)$ c'est en fait la formule (2) qui est mise en œuvre et qui fournit une valeur FP définie par

$$FP = (F(X_0 \oplus H) \ominus F(X_0 \ominus H)) / (2 \star H), \quad (5)$$

avec :

F , image en machine de la fonction f ;

X_0 et H , représentations en machine de x_0 et h ;

$\ominus \oplus \star /$ opérateurs informatiques qui sont les approximations des opérateurs arithmétiques exacts correspondants.

La valeur FP est donc entachée d'erreurs de méthode et de calcul. En appelant e_g l'erreur globale on peut écrire :

$$f'(x_0) = FP + e_g. \quad (6)$$

Cette erreur globale e_g se décompose en plusieurs erreurs qui sont définies par les égalités (7) à (10). En effet, on peut écrire :

$$f'(x_0) = \frac{f(x_0+h) - f(x_0-h)}{2h} + e_1 = f'_a(x_0) + e_1, \quad (7)$$

e_1 est l'erreur de méthode que nous avons déjà appelée e_m :

$$\frac{f(x_0+h) - f(x_0-h)}{2h} = \frac{f(X_0 \oplus H) - f(X_0 \ominus H)}{2H} + e_2, \quad (8)$$

e_2 représente l'erreur absolue sur la valeur de $f'_a(x_0)$ calculée non avec les valeurs exactes des arguments mais avec leur représentation machine.

$$\frac{f(X_0 \oplus H) - f(X_0 \ominus H)}{2H} = \frac{F(X_0 \oplus H) - F(X_0 \ominus H)}{2H} + e_3, \quad (9)$$

e_3 représente l'erreur engendrée par F qui est une image en machine de la fonction f . Nous appelons cette erreur par la suite erreur de donnée et nous la notons par e_4 .

$$\frac{F(X_0 \oplus H) - F(X_0 \ominus H)}{2H} = (F(X_0 \oplus H) \ominus F(X_0 \ominus H)) / (2 \star H) + e_4, \quad (10)$$

e_4 est l'erreur engendrée par les opérateurs informatiques $\ominus /$ et \star . En additionnant membre à membre les égalités (7) à (11) on obtient :

$$e_g = e_1 + e_2 + e_3 + e_4. \quad (11)$$

L'erreur globale est égale à la somme des 4 erreurs définies ci-dessus. Parmi ces erreurs il ressort que :

- e_4 est négligeable par rapport aux autres erreurs;
- e_2 se décompose elle-même en deux erreurs, l'une due au remplacement de x_0 et h par X_0 et H que l'on peut négliger en première approximation, l'autre, provenant de l'opérateur \oplus ou \ominus dans le calcul de $X_0 \oplus H$ et $X_0 \ominus H$. En choisissant convenablement H on peut rendre cette dernière erreur rigoureusement nulle. En effet compte tenu de l'arithmétique binaire en virgule flottante normalisée de la machine, le pas H doit appartenir à un intervalle I délimité par $HINF$ et $HSUP$ tels que :

$$HINF = X_0 \cdot 2^{-n}, \quad HSUP = X_0 \cdot 2^n, \quad (12)$$

n étant le nombre de bits de la mantisse et l'on doit avoir :

$$H \in]X_0 2^{-n}, X_0 2^n[= I. \quad (13)$$

En effet, si $H \leq HINF$, H disparaît totalement devant X_0 , lors du calcul en machine de $X_0 \oplus H$ et $X_0 \ominus H$; si $H \geq HSUP$, c'est X_0 qui disparaît totalement devant H .

Si, de plus, H est calculé sur l'ordinateur par la relation

$$H = (Hp \oplus X_0) \ominus X_0 \quad Hp \in I, \quad (14)$$

H est tel que tous les bits de sa mantisse dénormalisée de rang supérieur à n sont nuls, après égalisation des exposants de H et X_0 .

Et l'on a

$$X_0 \oplus H = X_0 + H \quad \text{et} \quad X_0 \ominus H = X_0 - H, \quad (15)$$

ce qui permet de considérer e_2 comme négligeable.

L'erreur globale se limite donc à : $e_g = e_1 + e_3 = e_m + e_d$.

3. ÉTUDE DES ERREURS DE MÉTHODE ET DE DONNÉES

3.1. Étude de l'erreur de méthode e_m

Nous admettons que f admet des dérivées définies et continues sur $[x_0 - h, x_0 + h]$ jusqu'à l'ordre 3. En développant chaque terme du numérateur de (2) en série de Taylor jusqu'à l'ordre 3 on obtient

$$e_m = -\frac{h^2}{6} f'''(\xi), \quad \xi \in [x_0 - h, x_0 + h]. \quad (16)$$

3.2. Étude de l'erreur de donnée e_d

L'erreur e_d est due à l'image en machine F de f . Cette image F n'est qu'une approximation de f puisqu'elle est calculée avec une arithmétique à précision limitée. L'erreur relative commise sur F , dépend du point où la fonction est calculée, nous la notons $\varepsilon(x)$. Elle est définie par

$$F(x) - f(x) = \varepsilon(x)f(x). \quad (17)$$

Si l'on introduit l'hypothèse que d'une part, il existe un nombre strictement positif P petit devant l'unité tel que pour tout $x \in A$ on ait :

$$|\varepsilon(x)| \leq P \quad (18)$$

et que, d'autre part, on peut, dans ce calcul d'erreur, négliger les termes en εh , ce qui revient à dire :

$$F(x_0 + h) \simeq F(x_0 - h) \simeq F(x_0), \quad (19)$$

on obtient à partir de (9) avec (17) :

$$e_d \simeq \frac{F(X_0)}{2h} (\varepsilon_1 - \varepsilon_2), \quad \varepsilon_1 \in [-P, P], \quad \varepsilon_2 \in [-P, P], \quad (20)$$

avec

$$\varepsilon_1 = \varepsilon(x_0 - h), \quad \varepsilon_2 = \varepsilon(x_0 + h).$$

3.3. Étude de la somme des erreurs $e_m + e_d$

Les erreurs ε_1 et ε_2 étant supposées quelconques et indépendantes dans l'intervalle $[-P, +P]$ nous allons estimer la valeur moyenne de $|e_m + e_d|$ c'est-à-dire $\overline{|e_m + e_d|}$.

Nous avons d'après (16) et (20) :

$$|e_m + e_d| = \left| -\frac{h^2}{6} f'''(\xi) + \frac{F(x_0)}{2h} (\varepsilon_1 - \varepsilon_2) \right|. \quad (21)$$

En supposant que dans D la densité de probabilité $p = p(\varepsilon_1, \varepsilon_2)$ est constante, la valeur moyenne de $|e_m + e_d|$ est définie par

$$\overline{|e_m + e_d|} = \frac{1}{4P^2} \iint_D \left| -\frac{h^2}{6} f'''(\xi) + \frac{F(x_0)}{2h} (\varepsilon_1 - \varepsilon_2) \right| d\varepsilon_1 d\varepsilon_2, \quad (22)$$

D étant le domaine rectangulaire de surface $4P^2$ défini par les horizontales d'ordonnées $-P$ et $+P$ et les verticales d'abscisses $-P$ et $+P$. Cette intégrale double dont le calcul délicat est présenté dans [2], fournit comme résultat :

$$\overline{|e_m + e_d|} = \frac{P|F(x_0)|}{3h} + \frac{h^5 f'''(\xi)^2}{36P|F(x_0)|} - \frac{h^8 |f'''(\xi)|^3}{648P^2 F(x_0)^2}. \quad (23)$$

4. DÉTERMINATION DU PAS OPTIMAL h_p

En supposant que pour tout h , $f'''(\xi) \simeq f'''(x_0)$, le minimum de $\overline{|e_m + e_d|}$ s'obtient en annulant la dérivée de (23) par rapport à h , ce qui donne la valeur du pas optimal h_p :

$$h_p = \sqrt[3]{\frac{1,67P|F(x_0)|}{|f'''(x_0)|}}. \quad (24)$$

Pour pouvoir calculer par (24) la valeur du pas optimal h_p il faut évaluer la dérivée troisième de la fonction. Ceci peut paraître un paradoxe; il n'en est rien car $f'''(x_0)$ n'intervient que par sa racine cubique; il suffit donc de connaître une valeur très approximative de la dérivée troisième. L'évaluation de $f'''(x_0)$ va se faire à l'aide de ce que nous appelons la méthode de l'erreur volontaire. Soit k un réel strictement positif et supposons que la dérivée troisième $f'''(x)$ varie suffisamment peu sur $[x_0 - 2k, x_0 + 2k]$. En développant f par la formule de Taylor avec reste de Lagrange aux points $x_0 + k$, $x_0 - k$, $x_0 + 2k$, $x_0 - 2k$ et en combinant ces divers développements on obtient, compte tenu des hypothèses ci-dessus la valeur de $f'''(x_0)$ qui s'exprime par

$$f'''(x_0) = \frac{f(x_0 + 2k) - f(x_0 - 2k) - 2(f(x_0 + k) - f(x_0 - k))}{2k^3}. \quad (25)$$

Le problème consiste à déterminer une valeur convenable pour le pas k . Tout comme nous l'avons montré au paragraphe 2 pour le pas h (13), le pas k doit appartenir à un intervalle borné, inférieurement par $KINF$ et supérieurement par $KSUP$. Ces bornes sont définies par

$$KINF = X_0 \cdot 2^{-n+1}, \quad KSUP = X_0 \cdot 2^{n-1}. \quad (26)$$

Nous allons chercher dans l'intervalle $[KINF, KSUP]$ une valeur de k telle que l'erreur de données soit volontairement élevée sans toutefois dépasser une certaine limite, afin de rendre l'erreur de méthode négligeable, de façon à ce que l'erreur globale soit majorée.

Posons

$$F'''(x_0) = \frac{F(x_0 + 2k) - F(x_0 - 2k) - 2(F(x_0 + k) - F(x_0 - k))}{2k^3} \quad (27)$$

et

$$\begin{aligned} T_1 &= F(x_0 + 2k), & T_2 &= F(x_0 - 2k), \\ T_3 &= -2F(x_0 + k), & T_4 &= -2F(x_0 - k), \end{aligned} \quad (28)$$

et désignons par A et B les quantités suivantes :

$$\begin{aligned} A &= \sum_i T_i \quad \text{avec} \quad T_i > 0 \quad i \in [1, 4], \\ B &= \sum_i T_i \quad \text{avec} \quad T_i < 0 \quad i \in [1, 4]. \end{aligned}$$

Si l'on remplace dans (25), $f(x)$ par son expression tirée de (17) faisant apparaître l'erreur sur la fonction, c'est-à-dire :

$$f(x) = \frac{F(x)}{1 + \varepsilon(x)}, \quad (29)$$

et que l'on affecte à $\varepsilon(x)$ suivant les cas les valeurs extrêmes $-P$ ou $+P$, on obtient les limites inférieures et supérieures de $f'''(x_0)$ que nous notons respectivement par $f'''_{\text{inf}}(x_0)$ et $f'''_{\text{sup}}(x_0)$. Ces valeurs extrêmes s'expriment par

$$\begin{aligned} f'''_{\text{inf}}(x_0) &= \frac{[A/(1+P)] + [B/(1-P)]}{2k^3}, \\ f'''_{\text{sup}}(x_0) &= \frac{[A/(1-P)] + [B/(1+P)]}{2k^3}. \end{aligned} \quad (30)$$

De plus P^2 étant négligeable devant l'unité nous pouvons écrire grâce aux égalités (27) et (30) :

$$F'''(x_0) = \frac{f'''_{\text{inf}}(x_0) + f'''_{\text{sup}}(x_0)}{2}. \quad (31)$$

Il est à remarquer que $F'''(x_0)$ est une image de $f'''(x_0)$ entachée de la double erreur de donnée et de méthode. Comme nous l'avons souligné précédemment ces deux erreurs jouent en sens inverse de telle sorte que si l'erreur de donnée devient négligeable, c'est que l'erreur de méthode est trop élevée

et réciproquement. Il faut donc choisir un moyen terme qui se traduit par le fait que le rapport

$$L = \frac{f'''_{\text{sup}}(x_0)}{f'''_{\text{inf}}(x_0)}, \quad (32)$$

doit se situer à une distance raisonnable de l'unité, c'est-à-dire :

- ni trop proche de 1, sinon l'erreur de donnée est négligeable et donc l'erreur de méthode est trop importante. Ceci se produit lorsque k est trop grand;
- ni trop éloigné de 1, sinon l'erreur de donnée est trop importante. Ceci se produit lorsque k est trop petit.

Ces contraintes se traduisent par les conditions suivantes :

$$\begin{aligned} L &\in [L1, L4] \\ L &\notin [L2, L3]. \end{aligned} \quad (33)$$

Il est montré dans [2] que, quelle que soit f , les bornes figurant dans (33) peuvent être prises égales à :

$$\begin{aligned} L1 &= 15, & L3 &= \frac{1}{L2}, \\ L2 &= 2, & L4 &= \frac{1}{L1}. \end{aligned} \quad (34)$$

Ainsi, va-t-on, par dichotomie dans l'intervalle $[KINF, KSUP]$, chercher une valeur de k qui satisfait les conditions (33). La valeur initiale de k sera prise égale à $\sqrt{KINF \cdot KSUP}$. Avec cette valeur on calcule L par (30) et (32). Trois cas peuvent se rencontrer :

a) $L \notin [L1, L4]$. Cela veut dire que l'erreur de donnée est trop grande, on doit alors augmenter le pas k ; on pose $KINF = k$ et le processus dichotomique recommence.

b) $L \in [L1, L4]$ mais $L \in [L2, L3]$. Cela veut dire que l'erreur de donnée est trop petite, on doit alors diminuer le pas k ; on pose $KSUP = k$ et le processus dichotomique recommence.

c) $L \in [L1, L4]$ et $L \notin [L2, L3]$. Les conditions (33) sont satisfaites; cette valeur de k notée k_c est correcte. Ayant k_c on calcule $f'''(x_0)$ par (25) puis enfin h_p par (24).

5. MISE EN ŒUVRE PRATIQUE DE LA MÉTHODE

Les formules établies ci-dessus fournissant le pas h_p qui permet d'obtenir la meilleure approximation de la dérivée doivent donc être mises en œuvre sur ordinateur. Il faudrait donc les présenter dans ce paragraphe en notations informatiques \oplus , \ominus , \star , $/$, définies au paragraphe 2. Afin de simplifier l'écriture nous les présentons avec les notations mathématiques classiques.

5.1. Détermination de P

Pour estimer la valeur de P , c'est-à-dire, la valeur de l'erreur relative de donnée sur la fonction F , on peut, soit utiliser la méthode de permutation-perturbation exposée dans [9], soit prendre des valeurs estimatives établies dans [2].

Méthode de permutation-perturbation

Cette méthode consiste à faire exécuter plusieurs fois le calcul de la fonction F en changeant d'une part l'ordre d'exécution des opérations élémentaires (permutation) et d'autre part, en prenant comme résultat de chaque opération élémentaire sa valeur par défaut ou par excès (perturbation).

A partir des divers résultats obtenus on déduit par un calcul élémentaire la précision sur F et par conséquent P .

Valeurs estimatives

En première approximation la valeur de P peut être prise égale à

$$P = 2^{-n/\alpha}, \quad (34)$$

- n , étant le nombre de bits de la mantisse du mot mémoire;
 $\alpha = 4$, si la fonction se calcule mal c'est-à-dire par exemple s'il y a, en cours de calcul, une soustraction entre deux quantités voisines toutes deux entachées d'erreurs;
 $\alpha = 2$, si la fonction se calcule normalement;
 $\alpha = 4/3$, si la fonction se calcule avec une grande précision.

Notons que l'on a toujours intérêt à avoir une fonction F aussi numériquement stable que possible, c'est-à-dire une fonction dont l'expression présente le moins possible de soustractions. Une telle expression s'obtient en chassant, par voie algébrique comme il est présenté dans [5], les soustractions réductibles.

5.2. Évaluation de $F'''(X_0)$

Pour pouvoir calculer $F'''(X_0)$ il faut estimer au préalable le pas K_c .

Évaluation de K_c

L'évaluation de K_c se fait en utilisant la méthode de dichotomie dans l'intervalle $[KINF, KSUP]$ comme expliqué au paragraphe 4. Dès que les conditions définies par (33) sont satisfaites on a obtenu la valeur de K_c .

Calcul de $F'''(X_0)$

Connaissant K_c , on obtient $F'''(X_0)$ par la relation

$$F'''(X_0) = \frac{F(X_0 + 2K_c) - F(X_0 - 2K_c) - 2(F(X_0 + K_c) - F(X_0 - K_c))}{2K_c^3} \quad (35)$$

5.3. Calcul de $F'(X_0)$

Détermination du pas H_p

Comme il a été montré ci-dessus le pas H_p est donné par

$$H_p = \sqrt[3]{\frac{1,67 P |F(X_0)|}{|F'''(X_0)|}}. \quad (36)$$

Enfin pour rendre l'erreur e_2 négligeable le pas H utilisé pour obtenir la meilleure approximation de la dérivée est donné par

$$H = (H_p + X_0) - X_0. \quad (37)$$

Finalement la dérivée $F'(X_0)$ est donnée par

$$F'(X_0) = \frac{F(X_0+H) - F(X_0-H)}{2H}. \quad (38)$$

5.4. Estimation de l'erreur sur cette dérivée

L'erreur relative sur la dérivée $F'(X_0)$ est définie par

$$e_r = \frac{|F'(X_0) - f'(x_0)|}{|f'(x_0)|}. \quad (39)$$

L'estimation moyenne de cette erreur relative est donnée par

$$ERM = \frac{EAM}{|F'(X_0)|}. \quad (40)$$

EAM étant la moyenne de l'erreur absolue qui s'exprime par

$$EAM = \overline{|F'(X_0) - f'(x_0)|}, \quad (41)$$

c'est-à-dire :

$$EAM = \frac{P |F(X_0)|}{3H} + \frac{H^5 F'''(X_0)^2}{36 P |F(X_0)|} - \frac{H^8 |F'''(X_0)|^3}{648 P^2 F(X_0)^2}. \quad (42)$$

Remarque : Il peut arriver dans certains cas que la dérivée ne puisse pas être correctement évaluée sur l'ordinateur utilisé. Il en est ainsi : lorsque

$$F'''(X_0) = 0,$$

lorsque

$$F(X_0) = 0 \quad \text{et} \quad F(X_0+H) = 0, \quad (43)$$

lorsque k_c n'existe pas (échec de la méthode dichotomique). On pose alors $ERM = 1$, ce qui signifie que la dérivée ne peut pas être évaluée.

6. ACCORD ENTRE THÉORIE ET PRATIQUE

Nous avons appliqué cette méthode sur des fonctions classiques pour lesquelles il est aisé de calculer la dérivée analytiquement, puis sur des fonctions compliquées, enfin sur des fonctions pour lesquelles il est impossible de calculer numériquement la dérivée. Les résultats sont donnés ci-après :

6.1. Fonctions classiques

Pour chacune des fonctions figurant dans le tableau, on a calculé pour 100 valeurs de $X_0 \in [0.1, 12.5]$:

- $F'(X_0)$ avec la méthode exposée ci-dessus;
- $f'(x_0)$ valeur de la dérivée analytique exacte;
- e_r erreur relative exacte définie par (39);
- ERM erreur relative estimée par notre méthode;
- $(ERM - e_r)/e_r$ différence relative de l'erreur estimée et de l'erreur exacte;
- N_F nombre de fois que F est calculée pour obtenir la dérivée.

Sur l'ordinateur utilisé, les fonctions F sont calculées avec une précision relative de l'ordre de 3.10^{-7} .

Le tableau présente la moyenne des 100 calculs.

TABLEAU

Fonction	Erreur relative exacte e_r	Erreur relative estimée ERM	$\frac{ERM - e_r}{e_r}$	N_F
e^x	$1,642.10^{-5}$	$1,662.10^{-5}$	0,012	15
Log x	$2,189.10^{-5}$	$2,278.10^{-5}$	0,041	17
\sqrt{x}	$2,375.10^{-5}$	$2,388.10^{-5}$	0,005	15
Arc tg x	$5,494.10^{-5}$	$5,361.10^{-5}$	- 0,024	20
sin x	$2,478.10^{-5}$	$2,351.10^{-5}$	- 0,051	15

Les résultats présentés dans le tableau mettent en évidence :

- d'une part, l'efficacité de la méthode (erreur peu élevée, très bon accord entre la valeur de l'erreur ERM estimée par la méthode et l'erreur exacte e_r);
- d'autre part, la nécessité de calculer la fonction entre 15 et 20 fois pour obtenir la meilleure dérivée au sens informatique.

6.2. Fonctions compliquées

Un grand nombre de dérivées de fonctions analytiquement compliquées et pour lesquelles il était quand même possible de calculer la valeur exacte de la dérivée ont été calculées par la méthode exposée ci-dessus. Il ressort

de cette étude que :

- d'une part, la méthode a toujours donné satisfaction et n'a jamais pu être mise en défaut;
- d'autre part, que cette méthode nécessite le calcul d'une vingtaine de fois la fonction F pour obtenir la meilleure dérivée au sens informatique.

6.3. Fonction non dérivable numériquement sur ordinateur

Considérons la fonction

$$f(x) = x^2 + 10^{100}. \quad (44)$$

Essayons de calculer la dérivée de cette fonction au point $x_0 = 1$. La dérivée exacte de cette fonction en x_0 est :

$$f'(x_0) = 2. \quad (45)$$

Par la méthode exposée ci-dessus quel que soit $H \in [HINF, HSUP]$ on trouve toujours :

$$F'(X_0) = 0. \quad (46)$$

En effet, la fonction $F(X)$ apparaît toujours comme constante et égale à 10^{100} car, $(X_0 + H)^2$ et $(X_0 - H)^2$ disparaissent toujours devant 10^{100} .

La dérivée $F'(X_0)$ étant trouvée égale à zéro, ERM est mise égale à 1 signifiant que la fonction n'est pas dérivable numériquement sur ordinateur.

7. CONCLUSION

Lorsque l'on veut calculer sur ordinateur la dérivée d'une fonction f en un point x_0 , on met en œuvre une formule d'approximation qui fournit une valeur entachée d'une part d'une erreur dite de méthode qui résulte du fait que l'on ne peut passer à la limite, et d'autre part d'une erreur dite de donnée qui résulte de l'arithmétique à précision limitée de l'ordinateur. Si l'erreur de méthode diminue quand le pas de dérivation h décroît, l'erreur de donnée, au contraire, croît. On a montré que l'erreur globale passait par un minimum et que, par conséquent, il existait un pas h_p optimal qui minimisait l'erreur globale. Nous avons exposé ici une méthode qui permet quel que soit f d'évaluer ce pas optimal et d'en déduire la meilleure valeur de la dérivée au sens informatique. De plus, cette méthode permet de détecter les cas de fonctions qui ne sont pas numériquement dérivables sur ordinateur. Cette étude va ainsi plus loin que celles qui ont été faites précédemment [1 et 4].

La méthode exposée ici a été utilisée pour calculer les dérivées d'un très grand nombre de fonctions simples, compliquées et même de fonctions non numériquement dérivables sur ordinateur. Elle a toujours donné satisfaction et n'a jamais été mise en défaut. Il faut toutefois remarquer que cette méthode est coûteuse en temps machine puisqu'elle nécessite en moyenne une vingtaine

de calculs de la fonction pour obtenir la meilleure approximation de la dérivée sur l'ordinateur considéré. Une étude en cours se propose de montrer qu'avec un nombre de calculs de la fonction beaucoup plus réduit, on peut obtenir une approximation de la dérivée, qui, sans être la meilleure, est néanmoins une approximation acceptable.

BIBLIOGRAPHIE

1. G. DAGMAR, *Extrapolation Method for Numerical Calculation of the Derivative of the Analytical Function and its Error Estimate*. Applikace Matematici, vol. 16, n° 2, Ceskoslovenska, 1971.
2. J. DUMONTET, *Algorithme de dérivation numérique. Étude théorique et mise en œuvre sur ordinateur*. Thèse 3^e cycle, Technip, Paris, 1973.
3. P. HENRICI, *Error Propagation for Difference Methods*. J. Willey and Sons, 1963.
4. Z. KOPAL, *Numerical Analysis*. Chapman and Hall, 1965.
5. M. LA PORTE, *Une formulation numériquement stable*. Rapport I.F.P., réf. 21516, août 1973.
6. M. LA PORTE et J. VIGNES, *Algorithmes Numériques, analyse et mise en œuvre*, tome I, Technip, Paris, 1974.
7. M. LA PORTE et J. VIGNES, *Étude statistique des erreurs dans l'arithmétique des ordinateurs; application au contrôle des résultats d'algorithmes*. Numerische Mathematik, vol. 23, 1974.
8. R. E. MOORE, *Interval Analysis*. Prentice Hall, 1966.
9. J. VIGNES et M. LA PORTE, *Error Analysis in Computing*. Proceeding of I.F.I.P. Congress, Stockholm, August 1974.
10. J. H. WILKINSON, *Error Analysis of Floating-Point Computation*. Numerische Mathematik, vol. 2, 1960.
11. J. H. WILKINSON, *Rounding Error in Algebraic Processes*. Prentice Hall, 1963.