IUT de Belfort - Montbéliard
DUT Informatique 1er année
2008-2009 ALUSTWANI Husam

Arithmétique numérique

Initiation, Erreurs numériques

A. Introduction :

L'*Analyse Numérique*, est une discipline des mathématiques. On vise à résoudre, par des calculs purement numériques, des problèmes d'analyse mathématique. Plus exactement, l'objectif est le passage d'une représentation implicite à une représentation explicite.

Explicite : les nombres entiers (2) et décimaux (3/2 =1.5) sont connus <u>mais pas</u> les réels ($\sqrt{3}$). On considère par contre que le nombre réel est connu si l'on connaît sa précision. Par exemple $e = 2{,}7183 \pm 10^{-4}$.

Implicite : la connaissance du nombre est liée à une propriété. Par exemple :

- racine d'une équation (ou d'un système d'équations): $x^2 - 4 = 0$
- limite d'une suite : $\frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots = e$

Par extension, une fonction est connue de manière explicite si l'on peut donner la valeur explicite en tout point de son domaine de définition.

- Explicite : $f(x) = \frac{x^3 + 6}{x^2 + 1}$, $f(x) = \frac{x^3 + 6}{x - 1}$, ou bien un tableau des valeurs de la fonction

*	0	1
0	0	0
1	0	1

$f(x, y) : x, y \in \{0,1\}$: la multiplication binaire

- Implicite : $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

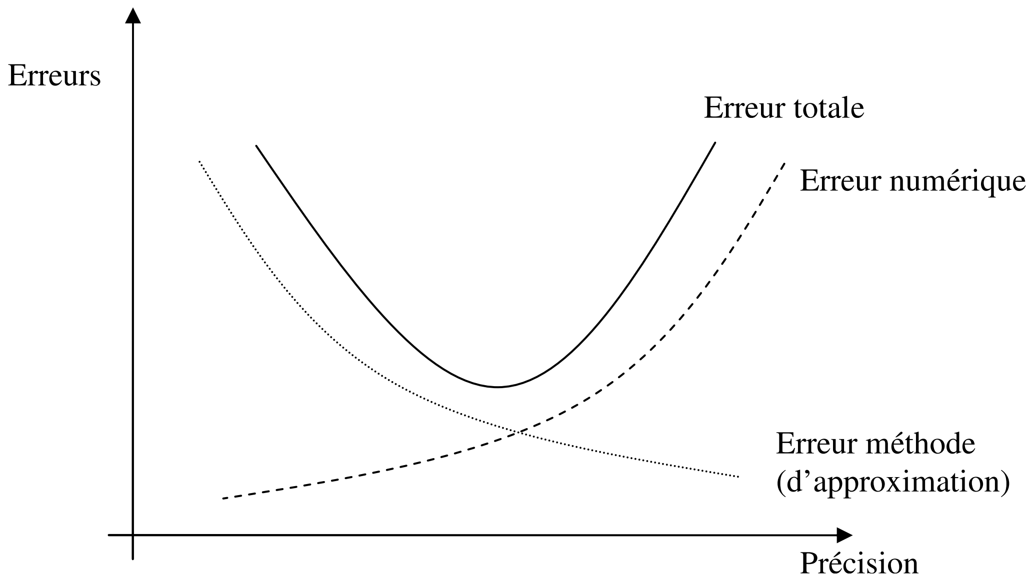
L'avantage de l'analyse numérique est de pouvoir donner une réponse approximative à un problème dont on ne connaît pas de solution directe (ou la résolution du problème prendrait un temps phénoménal : un problème NP complet). C'est-à-dire là où l'analyse mathématique classique est incapable de résoudre le problème.

Les méthodes (algorithmes) numériques : Les données et les résultats sont des nombres (le classement par ordre alphabétique d'un ensemble de mots n'est pas un algorithme numérique). Les algorithmes numériques sont de deux types principaux :

- Méthode directe (exacte) : Elles se définissent sous la forme d'équations de récurrence et conduisent à la solution en un nombre fini d'opérations arithmétiques élémentaires (factoriel, Fibonacci , etc).
- Méthodes itératives : Elles évaluent la solution exacte du problème par des approximations successives de celle-ci. Dans une telle méthode, on démarre depuis une valeur devinée ou estimée grossièrement, puis on trouve des approximations successives qui devraient converger (après avoir effectué un nombre infini d'itérations) vers la solution sous certaines conditions. Dans la pratique, il est impossible de procéder à l'infini. On s'arrête donc après un nombre fini d'itérations (après s'être approché ''suffisamment'' de la solution). Pour avoir plus de précision, on fait plus d'itérations (plus de calcul). Une étude de **l'erreur d'approximation** est donc nécessaire.

L'utilisation de l'analyse numérique (ayant pour objectif d'évaluer des nombres) est grandement facilitée par les ordinateurs notamment avec l'accroissement de la disponibilité et de la puissance de ceux-ci. Ces ordinateurs, ne travaillant qu'avec un nombre fini de *chiffres significatifs*, engendrent des **erreurs numériques**. L'erreur totale est donc la somme des erreurs d'approximations et des erreurs numériques.

- L'erreur d'approximation <u>décroît</u> donc avec la précision demandée (avec l'augmentation du nombre d'itérations).
- L'erreur numérique <u>croît</u> donc avec le nombre de calculs à effectuer.



B. Objectif à atteindre dans les (5) TP de l'analyse numérique :

Fournir une idée des méthodes numériques sans présenter les détails de démonstrations mathématiques de ces méthodes. Survoler certains problèmes numériques élémentaires (pour les approfondir ou découvrir d'autres méthodes et d'autres problèmes, reportez vous à la section: *Pour en savoir plus*) :

- Résolution des équations f(x) = 0 (méthode ''bisection'')
- Résolution des équations f(x) = 0 (méthode de Lagrange)
- Résolution des équations f(x) = 0 (méthode du Point fixe)
- l'intégration numérique des fonctions (méthode de rectangle, méthode de trapèze).

Pour chacun des problèmes présentés, on procèdera en deux étapes :

- Positionner le problème
- Programmer une petite application pour résoudre le problème

(Le choix du langage est libre)

Il existe des logiciels dédiés au calcul numérique comme MATLAB et Scilab. Ces logiciels intègrent des centaines (voire des milliers) de fonctions mathématiques et d'analyse numérique optimisées. L'utilisateur a la possibilité de créer ses propres fonctions. Ils permettent une visualisation graphique (2D, 3D) et beaucoup d'autres choses encore. La programmation avec un logiciel comme Scilab devient intéressante lors de l'utilisation des matrices car il intègre les notations matricielles (il effectue les calculs sur les matrices comme sur les nombres).

C. Erreurs numériques

La place réservée en mémoire pour ranger un nombre est un nombre fixe d'octets.
Il en résulte que les nombres sont rangés sous forme tronquée donc approchée. Ceci conduit à une erreur sur le nombre appelée **erreur d'arrondi**. Cette erreur est très faible et peu significative si l'on considère le nombre individuellement.

I. Erreurs dues au changement de base

L'Homme fournit à l'ordinateur, en général, les nombres sous forme décimale. A l'intérieur de l'ordinateur, ces nombres sont transformés et manipulés sous forme binaire (avec les erreurs que cela comporte). Une fois le calcul terminé, ces nombres sont reconvertis en base de 10 ce qui introduit une nouvelle erreur. L'affichage à l'écran ne fournit donc pas nécessairement la valeur exacte contenue en mémoire.

- Comment expliquez-vous les résultats suivants ?

Opération	Résultat affiché
$\frac{1}{3}+\frac{1}{3}+\frac{1}{3}$	1
$1-\frac{1}{3}-\frac{1}{3}-\frac{1}{3}$	$x \neq 0$

La première opération donne un résultat apparemment exact. Si ce résultat était rigoureusement exact on devrait avoir 0 comme résultat de la seconde opération. Le résultat de la première opération n'est pas 1 (elle vaut soit 1 + ε, soit 1 - ε selon la technique d'arrondissement utilisé par l'ordinateur), mais après l'avoir converti en base 10, cette valeur est arrondi à 1. Ceci n'est pas le cas dans la seconde opération.

II. Erreur de l'addition

- Trouver à l'aide de l'ordinateur le nombre r suffisamment grand pour que :

Opération	Résultat donné par l'ordinateur
$1+10^{-r}$	1
$1+10^{-r}-1$	0
$1-1+10^{-r}$	10^{-r}
$\dfrac{1+10^{-r}-1}{1-1+10^{-r}}$	0 (au lieu de 1)

- Sachant que les opérations sont effectuées de la gauche vers la droite, que remarquez-vous ?

Quand on effectue l'addition $1+10^{-r}$, l'opération de troncature sur 10^{-r} transforme ce nombre en 0.

- Comparer la valeur de s après l'exécution des instructions suivantes :

s = 0 ;
for (i = 0 ; i < 100 ; i++)
    s += 10^{-r} ;
s+= 1 ;
	s = 1 ;
	for (i = 0 ; i < 100 ; i++)
	    s += 10^{-r} ;

- Qu'en déduisez vous ?

Dans une somme il sera toujours préférable d'additionner d'abord les termes les plus petits en valeur absolue pour que leur cumul ne soit pas négligé face aux termes les plus grands.

III. Erreur de soustraction de deux quantités voisines

Généralement, l'erreur (relative) lors de l'opération $x-y$ s'accroît avec un facteur de $\dfrac{|x|+|y|}{|x-y|}$. Plus x et y sont proches l'un de l'autre, plus la valeur de ce facteur est grande.

On évitera donc en général d'utiliser des formules d'analyse faisant intervenir des différences de nombre de même signe. On leur préféra des formules calculant des sommes ou des produits. On peut (dans certains cas) obtenir ces formules par transformation directe.

- Soient a = 1.02 et b = 1.01. Comparez les valeurs (a-b)(a+b) et $a^2 - b^2$. Pourquoi la première formule est-il plus précise ?

En calculant les valeurs on remarque que $(a^2 - b^2) \succ (a - b)$. Cela explique l'erreur plus accentuée dans la seconde formule.

IV. Accumulation et amplification des erreurs

Une erreur répétée un grand nombre de fois ou multipliée par un grand nombre peut conduire à une erreur globale non négligeable.

- Testez les instructions suivantes :

```
a = 1/3 ;
s = 0 ;
for (i = 0 ; i < 30000 ; i++)
        s+=a ;
erreur = s - 10000 ;
```

- Comment expliquez vous que la valeur de erreur n'est pas 0 ?

L'erreur de troncature effectuée lors de l'enregistrement en mémoire de la valeur 1/3 s'accumule au fur et à mesure que la valeur de s augment.

D. Pour en savoir plus :

- Sites Internet :

- http://rfv.insa-lyon.fr/~jolion/ANUM/node1.html
- http://www.iecn.u-nancy.fr/~sokolows/support/support.html
- http://www.eudil.fr/eudil/jbeuneu/index.html

- Livres:

- Mathématique sur ordinateur, François COTTET-EMARD et Pierre GOETGHELUCK.
- Algorithmique numérique, C. BREZINSKI.
- Initiation à l'analyse numérique, P. LASCAUX.
- Méthodes numériques appliquées, A. GOURDIN, M.BOUMAHART
- Introduction à l'analyse numérique, Jacques BARANGER