Méthodes Numériques

MERCI !!

M. Schatzman, Analyse numérique,
Masson.

Eric Goncalves : Méthodes et Analyse Numérique
http://cel.archives-ouvertes.fr/docs/00/55/69/67/PDF/MethodesNumeriques_EricGoncalves.pdf

Paola Goatin : Analyse Numérique
http://team.inria.fr/opale/files/2011/11/Anum.pdf

Hervé Sauer : Calcul numérique, visualisation & programmation avec Matlab®

INSTITUT d'OPTIQUE GRADUATE SCHOOL ParisTech

Buts de ce cours

- Formaliser un problème
  - Du problème au modèle

- Résoudre numériquement
  - Du modèle à la simulation numérique

- Challenges
  - Équations continues vs. formes discrètes
  - Garantir l'obtention d'un résultat
  - Rapidité
  - Précision et validité du résultat

Calcul : numérique vs. Formel

- Calcul formel
  - Manipulation d'expressions
  - Solution analytiques et exact
  - Logiciels : Maple, maxima (xmaxima, wmaxima), ….
  - Bibliothèques : GiNaC, sympy, ….

- Calcul numérique
  - Manipulation de valeurs numériques
  - Solutions approchées de la solution limite
  - Outils : matlab, mathematica, scilab, octave, …..
  - Bibliothèques : NAG, LAPACK, Eigen, …

Ordinateur : un monde fini

- Ressources limitées
  - Mémoire
  - Stockage
  - Calcul

- Un nombre fini de valeurs
  - Des mots de N bits : 2^N valeurs

- Problèmes
  - Comment rester dans les bornes ?
  - Comment représenter et manipuler des valeurs réelles ?
  - Quelle précision ?

Entier et virgule fixe

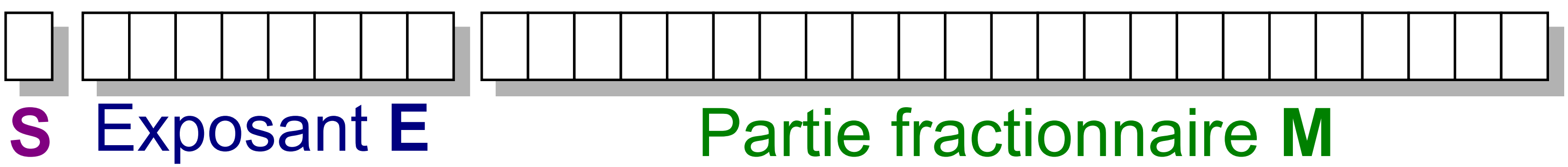
- Entiers signés sur N bits
  - 1 bit de signe pour les entiers signés
  - Débordement : X+ 1= ?
- Réels à virgule fixe sur N bits
  - Un intervalle [a,b] sur $[0, 2^N-1]$
  - Un intervalle [a,b[sur $[0, 2^N-1]$
  - Une arithmétique particulière
    - Calcul entier ramené au bon intervalle
  - Ex : couleurs d'une image

Approximation des réels : virgule flottante

- Écriture scientifique
  - **M**antisse
  - **E**xposant
  - **B**ase
  - **S**igne

$$(-1)^{S} \times M \times B^{E}$$

- En binaire



$$(-1)^{S} \times 1.M \times 2^{E}$$

Codage binaire : virgule flottante

- **Standard IEEE 754 (32 et 64 bits)**

32 bits: 1 bit **S** + 8 bits **E+126** + 23 bits **M**

$2^{-126} \leq |X| \leq (2-2^{-23}) \times 2^{127}$

$1.18 \times 10^{-38} \leq X \leq 3.40 \times 10^{38}$ (approx.)

Exemple: 1 10000001 01000000000000000000000 =

$(-1)^1 \times 2^{129-127} \times 1.(0.01_2) = -1 \times 2^2 \times 1.25 = -5$

64 bits: 1 bit **S** + 11 bits **E+1022** + 52 bits **M**

Quelques sources d'erreurs communes

- Débordement : résultat au delà des limites
  - Addition, multiplication

- Perte de précision
  - Somme de deux valeurs très différents
  - Différences de deux valeurs très proches

- Non commutativité des opérations

La stabilité

- Une petite variation des données
  - En physique : problème chaotique
    - Résultat imprévisible
  - En mathématique : conditionnement
    - Large variation du résultat

- Pour les méthodes numériques
  - Propagation des erreurs numériques
    - Instable = forte propagation

Conditionnement

- Problème :
  - Données d (perturbées $d + \delta d$)
  - Résultats x (perturbées $x + \delta x$)
- Conditionnement relatif

$$K_{\text{rel}}(\boldsymbol{d}) = \lim_{\epsilon \to 0} \sup_{|\delta \boldsymbol{d}| < \epsilon} \left\{ \frac{|\delta \boldsymbol{x}| / |\boldsymbol{x}|}{|\delta \boldsymbol{d}| / |\boldsymbol{d}|} \right\}$$

- Conditionnement absolu

$$K_{\text{abs}}(\boldsymbol{d}) = \lim_{\epsilon \to 0} \sup_{|\delta \boldsymbol{d}| < \epsilon} \left\{ \frac{|\delta \boldsymbol{x}|}{|\delta \boldsymbol{d}|} \right\}$$

Exemple 1D

- Conditionnement du calcul de $x = \mathrm{f}(d)$
- Évaluer laquelle des formulations équivalentes donne la meilleure solution approchées lorsque l'on utilise une valeur $\sqrt{2}$ approchée de

$$\left(\frac{1}{\sqrt{2}+1}\right)^6 \qquad \left(3-2\sqrt{2}\right)^3 \qquad \left(\frac{1}{3+2\sqrt{2}}\right)^3 \qquad 99-70\sqrt{2} \qquad \frac{1}{99+70\sqrt{2}}$$

- Moralité ?

Stabilité

- **Problème** :
  - Données numérique $\boldsymbol{d}_n$ (perturbées $\boldsymbol{d}_n + \delta \boldsymbol{d}_n$)
  - Résultats $\boldsymbol{x}_n$ (perturbées $\boldsymbol{x}_n + \delta \boldsymbol{x}_n$)

$$\begin{array}{c} \forall \mu > 0 \\ \exists M > 0 \quad \text{tel que} \quad |\delta \boldsymbol{x}_n| < M\, |\delta \boldsymbol{d}_n| \\ \forall\, |\delta \boldsymbol{d}_n| < \mu \end{array}$$

- Condition nécessaire pour la **convergence**
  - L'approximation se réduit avec
    - La finesse de la discrétisation
    - Le nombre de pas d'itération

Consistance

- Qualité de l'approximation numérique
  - passage du continue au discret
- Exemples
  - Intégration de Riemman (consistant)

$$\lim_{n \to +\infty} \frac{b-a}{n} \sum_{i=1}^{n} \mathrm{f}\left(a+i\,\frac{b-a}{n}\right) = \int_a^b \mathrm{f}(x)\,\mathrm{d}x$$

- Stabilité + Consistance → Convergence