Calculs approchés

Ça peut coûter cher…

Pour un chiffre de plus…

Un dispositif de calibrage utilisé dans la fusée Ariane 4 avait été laissé actif, alors qu'il n'était pas utilisé dans Ariane 5.

Les conditions de vol étant différentes entre Ariane 4 et Ariane 5, la valeur d'une donnée traitée par ce dispositif se trouvait, dans Ariane 5, dépasser les limites prévues pour Ariane 4. Ce dépassement avait une probabilité jugée négligeable de survenir dans Ariane 4 et aucune récupération d'erreur n'était prévue pour le traiter.

En conséquence, selon une politique “normale” dans ce cas de figure, le module de navigation fut mis hors service pour erreur irrécupérable. Le module actif et le module de secours étant identiques et contenant le même logiciel, les mêmes causes produisirent les mêmes effets sur les deux modules, donnant lieu à une situation non prévue. En effet, le module de secours était destiné à compenser une erreur transitoire et aléatoire d'origine matérielle, erreur dont la probabilité était jugée suffisamment faible pour exclure dans la pratique une défaillance simultanée des deux modules.

- Plage de mesure : 0 à 9999,9 m
- Erreur: ≤ 0,5%
- Diamètre de la roue : env. 32cm (12.6 ‘’)
- Périmètre de roue: env. 1 mètre
- Longueur totale : env. 101cm (39.8’’)
- Température appropriée: -10 ~ 45 °

Vive le running

Organisateur de courses à pied, vous devez mesurer le plus exactement les 42,195 km d’un marathon à l’aide d’une roue de mesure qu’un catalogue présente ci-contre…

Quelques remarques :

0,5% de 42,195 km font 210,975 m…

12.6’’ font 32,004 cm, bon, mais $32{,}004 \times \pi = 1{,}005435313\ldots$ m cela fait une erreur de 0,54…%

39.8’’ font 1,01092 cm. Quelle est l’épaisseur du caoutchouc?

Question :

On suppose que le diamètre de la roue a été mesuré à 32 cm avec une précision d’un dixième de mm. Peut-on être sûr de la longueur du marathon à 1 m près?

Proportionnalité : fake news

Mon automobile me ment

Comment fabriquer un trou ?

Si la sensibilité des capteurs est 0,02 s et si leur distance est 3,5 m, le calcul de la vitesse avec des temps compris entre 0,12 et 0,14 s donne des vitesses de 105 km/h ou 90 km/h du fait de la **troncature.**

Un trou dans les données

On étudie le comportement de 180 000 véhicules empruntant une bretelle d'autoroute. Pour cela, on a installé deux capteurs magnétiques (très peu éloignés*), mesuré les temps de passage puis calculé les vitesses grâce à $v = \frac{d}{t}$.

Surprise : sur 180 000 véhicules, 40 000 dépassent les 100 km/h, mais on n'en trouve **aucun** entre 90 et 100...

**Précision : l'étude était demandée par une société d'autoroutes (Atlandes) et portait sur les prises à contresens de la bretelle, ce qui explique la petite distance entre les capteurs…*

(www.**scmsa**.eu/archives/BB_Absence_donnees_2019_01_27.pdf)

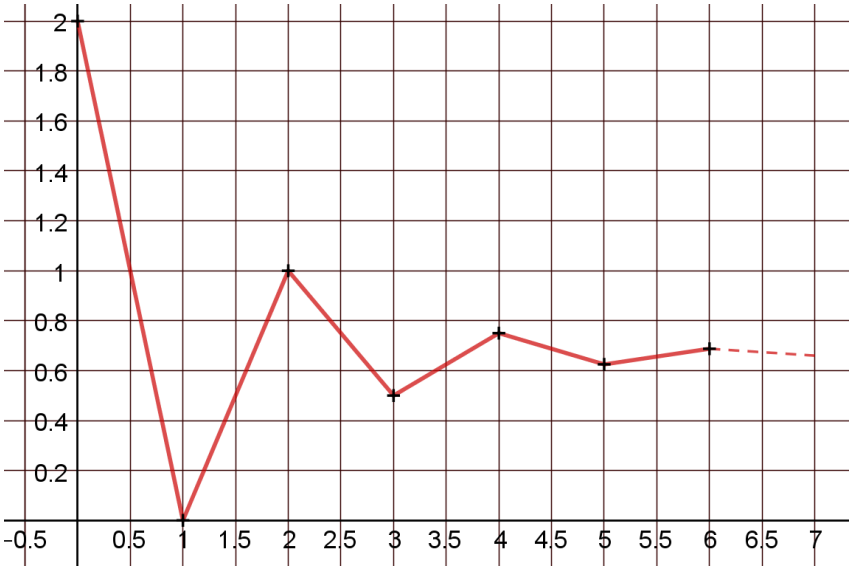
On veut des définitions (1)

Valeur approchée : on dit que a est une valeur approchée de x à ε près ssi $|x - a| \leq \varepsilon$.

Par exemple, pour la fonction représentée (partiellement) ci-contre, on peut établir que pour tout ε positif, il existe M positif tel que

$$x > M \Rightarrow \left| f(x) - \frac{2}{3} \right| < \varepsilon$$

La valeur approchée est liée à la notion de *distance*, fondamentale pour la suite



On veut des définitions (2)

Les nombres sont écrits dans la numération de base q, dont les chiffres sont 0, 1, 2, etc. Si l'écriture de N (éventuellement illimitée) est : $N = \alpha q^n + \beta q^{n-1} + \cdots + \sigma q^{-m+1} + \tau q^{-m} + \omega q^{-m-1} + \cdots$

(où m et n sont des entiers naturels),

La **troncature** de N au rang $-m$ est $T_m(N) = \alpha q^n + \beta q^{n-1} + \cdots + \sigma q^{-m+1} + \tau q^{-m}$.

L'**arrondi** au rang $-m$ de N est $A_m(N) = T_m(N)$ si $\omega < \frac{q}{2}$ et $A_m(N) = T_m(N) + q^{-m}$ si $\omega \geq \frac{q}{2}$.

La propagation des erreurs

On cherche à effectuer le produit 1×1. On suppose qu'on commet une erreur ε sur chacun des deux facteurs. Le produit effectué est donc
$(1+\varepsilon)(1+\varepsilon) = 1 + 2\varepsilon + \varepsilon^2$.
L'erreur relative est de 2ε. Mais si on « entre » dans le calculateur on peut trouver pire.

Exemple 1 La méthode dite de Babylone pour calculer la racine carrée de a : $x = \frac{1}{2}\left(x + \frac{a}{x}\right)$ consiste à itérer le calcul. Pour $a = 2$, en partant de $x = 1$, on obtient rapidement…

Exemple 2 On fait la même chose avec une équation dont la solution connue est 1, et qui s'écrit : $x = 4\,916{,}1x - 4\,915{,}1$ et un logiciel de calcul scientifique.
On obtient les valeurs : x = 1.0000000000005, x = 1.0000000018631, x = 1.0000076314441 , x = 1.0312591580864, x = 129.04063743776, x = 524468.25500881, etc.

Mais alors, que peut-on faire?

Limiter les erreurs humaines

Augustin Louis Cauchy (1789-1857)

Mémoire de 1840 : « Sur les moyens d'éviter les erreurs dans les calculs numériques »

Un exemple :

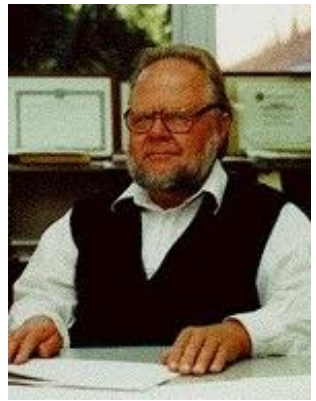
	4	2	3	$\overline{4}$	$\overline{2}$	C'est 42 258
+	1	$\overline{5}$	$\overline{4}$	5	$\overline{3}$	C'est 4 647
=	5	$\overline{3}$	$\overline{1}$	1	$\overline{5}$	C'est 46 905

Autre exemple : $\frac{1}{7} =$

$0{,}142857\ 142857\ 142587\ \ldots$

S'écrit aussi $\frac{1}{1\overline{3}} = 0,143\ \overline{1}\ \overline{4}\ \overline{3}\ \ldots$

Avizienis redécouvre sans le savoir le système de Cauchy. Ses travaux sont utilisés dans les processeurs. Ces systèmes évitent la propagation de retenues.

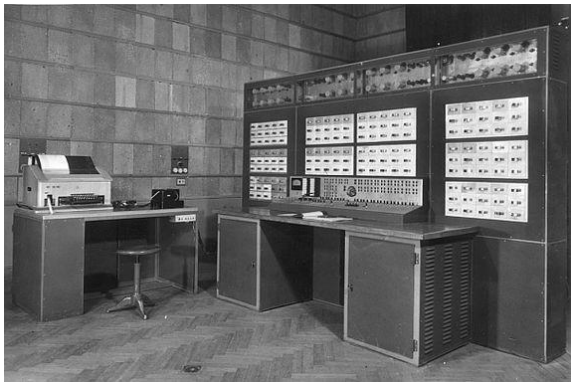


Algirdas Antanas Avizienis (né en 1932)

Arrondir les troncatures ?

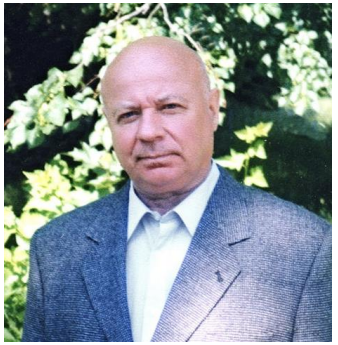
Dans le système ternaire dit « équilibré » les chiffres sont $1, 0$ et $\overline{1}$ comme chez Cauchy, mais il y en a moins… Pour passer d'un entier écrit en base 3, on remplace simplement les 2 qui apparaissent dans l'écriture par $\overline{1}$, en ajoutant 1 aux chiffres situés à leur gauche.
Par exemple : le nombre $1\,789$ s'écrit $2\,110\,021$ en base 3 et $1\overline{1}\,110\,1\overline{1}\,1$ dans le système ternaire équilibré.
Dans ce système, l'arrondi et la troncature coïncident !

En Union soviétique, on a amorcé à la fin des années 1950 la construction d'ordinateurs obéissant à une logique ternaire. Elle fut peu à peu abandonnée en tant que voie industrielle.



Ordinateur Setun
URSS 1958

Supprimer des multiplications



Anatoli A. Karatsuba
(1937 – 2008)

Pour multiplier deux nombres entiers de deux chiffres, il faut 4 produits d'un chiffre par un chiffre, deux sommes et éventuellement deux sommes et l'ajout de trois retenues.

Mais : $(10a + b)(10c + d) = 100 \times ac + 10 \times \big(ac + bd - (a - b)(c - d)\big) + bd$

Ce calcul ne demande que **trois** produits de nombres de un chiffre (souvent inférieurs aux chiffres de départ).

Heureusement, cette façon de faire se généralise à des nombres plus grands, selon le principe « diviser pour régner »

$$(10^n a + b)(10^n c + d) = 10^{2n} \times ac + 10^n \times \big(ac + bd - (a - b)(c - d)\big) + bd$$

Le produit par une puissance de la base (que ce soit 10 ou un autre entier n'a pas d'importance) n'entraîne qu'un déplacement de virgule, ne coûte rien en puissance ou temps

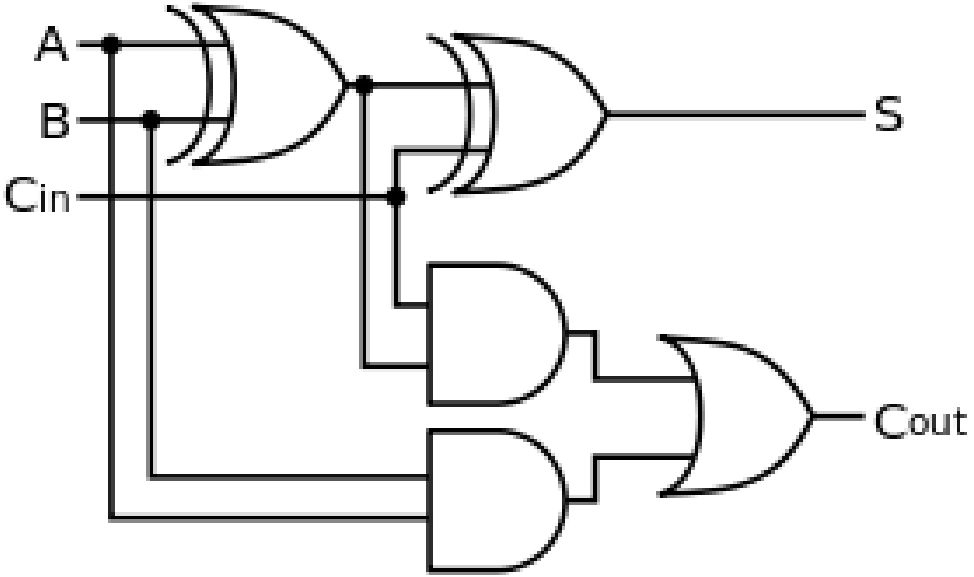
Maîtriser les retenues (1)

Un simple **additionneur binaire** doit à chaque pas gérer une retenue (carry)

Les portes XOR donnent 1 en sortie si et seulement si les entrées sont différentes.

Les portes ET donnent 1 en sortie si et seulement si les deux entrées sont 1, 0 sinon.

Il y a bien d'autres façons d'anticiper ou retarder l'apparition des retenues.



George Boole (1815 – 1864) et Claude Shannon (1916 – 2001)

Maîtriser les retenues (2)

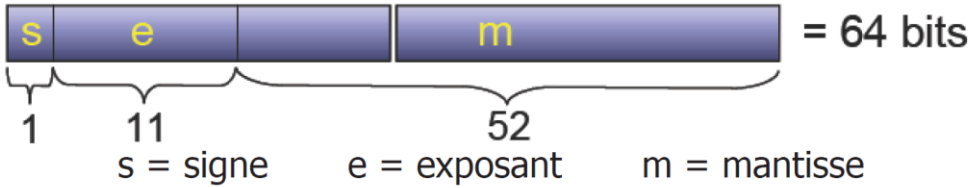
Une vieille histoire…

Calculer par référendum

Avant les calculateurs électroniques, [les calculs] étaient réalisés par des équipes, essentiellement des femmes, réunies dans de grandes salles de calcul qui résolvaient à la chaîne des problèmes arithmétiques simples. Les calculs complexes étaient décomposés et exécutés par deux équipes. Si le résultat était identique, il était validé, sinon, ils répétaient le processus. Les superviseurs archivaient les réponses correctes.

La solution virgule flottante

Double précision = 2 X 32 bits



Nombre = $(-1)^s \times 1.m \times 2^{e-1023}$

Valeur absolue min = $2,225... 10^{-308}$

Valeur absolue max = $1,797... 10^{308}$

Erreur relative = $2,220446049 \times 10^{-16}$

soit 16 chiffres maximum de précision

Les nombres « possibles » avec ce système sont appelés des flottants. Un changement de l'exposant modifie la distance entre deux flottants voisins. Il y a un plus petit *nombre machine*, un plus grand et un *epsilon machine* (Ne jamais faire un test sous la forme $= 0$?)
Les ***arrondis*** sont eux aussi des nombres machine…

Document issu de la présentation de Claude Gomez Pépinière avril 2011

… et la norme IEEE 754

Pour les curieux… qui ont raison de l'être

. La présentation de Claude Gomez :

https://euler.ac-versailles.fr/IMG/pdf/pepiniere_secondes_2010_2011_calcul_numerique.pdf

. Le site https://interstices.info en choisissant les articles par niveau

. À l'adresse http://perso.ens-lyon.fr/jean-michel.muller/goldberg.pdf l'article

« What every computer scientist should know about floating point arithmetics »

Sans oublier **Sylvie Boldo**