Aabu

Limites de suites et précision limitée - Partie 3

22 mars 2019

Table des matières

% LIMITES DE SUITES ET PRÉCISION LIMITÉE - PARTIE 3 % Aabu % 20 avril 2018

1. Introduction

Dans un billet précédent et sa suite , j'explorais comment était influencée la limite d'une suite lorsqu'on calculait en précision limitée.

Ces résultats ont servi à justifier le design d'un estimateur embarqué sur lequel j'ai pu travailler.

2. Présentation du problème

Pour chaque pas de calcul n, on souhaite calculer un paramètre $p(n)$ définit par la suite suivante :

$$p(n+1) = ap(n) + b(n)$$
$$p(0) = p_0$$

où $b(n)$ est une valeur a priori nouvelle à chaque pas de calcul et a est une constante telle que $0 < a < 1$.

Un cas particulier courant est celui où $b(n)$ est constant et on a $b(n) = b$. La suite se simplifie alors en :

$$p(n+1) = ap(n) + b$$
$$p(0) = p_0$$

On aimerait que dans le calculateur embarqué, le calcul de la suite p simplifiée converge comme la suite idéale. Plus précisément, on se donne les exigences suivantes :

— Le paramètre a peut être tel que $0.5 \leq a < 1$.
— Le paramètre b peut être tel que $0 \leq b \leq 200$.

— La limite doit pouvoir être calculée tant qu'elle est dans l'intervalle $[0, 300]$.
— La valeur de la suite embarquée doit être éloignée de celle de la suite idéale de 1 au plus à chaque instant.
— Tous les types utilisés pour les calculs et le stockage de valeurs sont à virgule fixe et sur 32 bits.

3. Notes sur les types à virgule fixe

En informatique, le calcul sur les nombres à virgule se fait le plus souvent en utilisant des flottants, qui sont un type généraliste assez pratique. Cependant, de nombreux calculateurs embarqués n'offrent par défaut que du calcul à virgule fixe, assez différent.

Dans ce type de stockage, on traite un nombre à virgule comme un entier divisé par une puissance de deux donnée (ce qui revient à placer la virgule à une place fixe dans le nombre binaire). Par exemple, si on souhaite stocker la valeur 0.5, on peut la stocker comme 1 divisé par 2. Dans ce même format, pour stocker la valeur 8, il faudra stocker l'entier 16 (parce que 16 divisé par deux donne 8).

Chaque format est donné par la puissance de deux qu'on utilise pour obtenir la valeur réelle à partir de l'entier stocké. La résolution d'un type à virgule fixe correspond à l'inverse de la puissance de deux correspondante. La gamme de valeur stockable est d'autant plus réduite que la puissance de deux est grande.

Le tableau ci-dessous résume différents choix possible pour stocker une valeur en virgule fixe.

Décalage de la virgule	Puissance de deux	Résolution	Valeur minimale	Valeur maximale
1	2	$2^{-1} = 0{,}5$	0	2147483647,5
2	4	$2^{-2} = 0{,}25$	0	1073741823,75
8	256	$2^{-8} = 0{,}00390625$	0	16777215,99609375
16	65536	$2^{-16} = 0{,}000015259$	0	65535,999984741
24	16777216	$2^{-24} = 5{,}9604... \times 10^{-8}$	0	255,99999994
32	4294967296	$2^{-32} = 2{,}32830... \times 10^{-10}$	0	0.99999999976...

4. Précision sur les paramètres a et b

Quelle précision peut-on obtenir sur a et b avec des types à virgule fixe ?

Pour a, la plus grande précision au regard de l'intervalle $]0.5, 1[$ est de 2^{-32}, obtenue en décalant la virgule de 32 bits, ce qui permet un maximum de tout juste $1\text{-}2^{-32}$. C'est parfait puisque 1 est exclu de notre intervalle.

Pour b, la plus grande précision atteignable pour représenter des valeurs dans l'intervalle]0, 200] est 2^{-24}, parce que le maximum du type correspondant est 255,9 et des poussières.

Dans le billet précédent , je donnais une formule pour estimer l'erreur sur la limite due à l'erreur sur les paramètres initiaux, tant que celle-ci est faible :

$$\epsilon_u = \frac{1}{1-a}\epsilon_b + \frac{b}{(1-a)^2}\epsilon_a$$

On peut la calculer dans un cas défavorable. Pour se placer dans un cas défavorable, on prend les erreurs maximales possible sur les arrondis de a et b : $\epsilon_a = 2^{-32}$ et $\epsilon_b = 2^{-24}$. Ensuite, il s'agit de prendre a assez proche de 1, mais qui reste réaliste. La physique du système suggère que la valeur maximale en pratique pour a sera de 0,999999 (six fois le chiffre 9). Pour rester à notre limite maximale, on prend $b = (1-a) \times 300$.

En faisant le calcul, on trouve :

$$\epsilon_u \approx 0{,}13$$

Ce qui est amplement suffisant !

5. Conclusion

Ce calcul analytique permet de donner une indication sur la performance obtenue pour la limite de la suite seulement. Bien que ce billet ne montre qu'un exemple, il est possible de calculer sur toute la gamme de valeurs possible en balayant les intervalles pour a et b.

Ce calcul ne résout pas la précision du calcul pendant la convergence, qui est aussi importante en pratique. Sur le cas réel, cela a été fait par simulation en comparant une simulation en haute précision avec le même calcul en précision limitée.

Le travail d'analyse et de vérification a permis de s'assurer que l'estimateur embarqué calculait avec la précision souhaitée.