Arithmétique des ordinateurs – TD 06 : Division multiplicative

{ christoph.lauter, matthieu.gallet } @ens-lyon.fr

21 avril 2008

Durant cette séance, nous allons examiner la seconde grande famille de méthodes pour calculer des divisions ou des racines carrées : les méthodes multiplicatives.

1 Calculs d'inverse

1. Rappelez l'algorithme de calcul de l'inverse utilisant la méthode de Newton-Raphson.
2. Effectuez les différentes étapes de l'algorithme de Newton-Raphson pour calculer l'inverse de $d = 29/256$. Pour le résultat initial, prenez la troncature aux 4 premiers bits de $R[0] = 2-d$. On recherche une précision d'au moins 2^{-12} pour tout d, tel que $1/2 \leq d < 1$.
3. Rappelez l'algorithme de calcul de l'inverse utilisant la normalisation multiplicative.
4. Effectuez les différentes étapes de la normalisation multiplicative pour calculer l'inverse de $d = 29/256$. On part de la même approximation initiale, et on recherche la même précision que dans la question 2.
5. Rappelez l'intérêt de la normalisation multiplicative sur la méthode de Newton-Raphson
6. Montrez que l'on peut accélérer la vitesse de convergence de l'algorithme en ajoutant des termes de puissances plus grandes dans la récurrence. Cette méthode est-elle intéressante ?

2 Racine carrées

7. Calculez la racine carré de $x = 0,125$ avec une précision d'au moins 2^{-12}. pour tout x compris entre $1/4$ et 1, avec la méthode de Newton-Raphson.
8. Refaire la question précédente, mais avec la méthode de la normalisation multiplicative.

3 Approximations initiales

Les méthodes de Newton-Raphson et de la normalisation multiplicative reposent toutes deux sur une approximation initiale du résultat. Évidemment, une méthode simple consiste à prendre une constante indépendante de d, mais peut donner lieu à une erreur assez importante, il faudra donc plus d'itérations. Dans le cours, Florent vous a parlé de méthodes tabulaires : une table indexée par les premiers bits de d va donner l'approximation initiale.

9. Montrer qu'utiliser k bits de d permet d'avoir une approximation de $k+1$ bits et une erreur maximum de 2^{-k}.

Cependant, il existe bien d'autres méthodes ; on peut par exemple faire une approximation linéaire du résultat, en prenant $R[0] = a - bd$, où a et b sont des constantes judicieusement choisies. On peut également combiner constantes prédéfinies et approximations linéaires. Notons d sous la forme $d = d_1 + d_2 2^{-k} + d_3 2^{-n}$. Partant de là, on va utiliser les k premiers bits de d pour trouver des coefficients a et b, afin d'obtenir $R[0] = a + bd_1 2^{-p}$. Au prix d'une recherche dans la table, d'une petite multiplication et d'une addition, on a peut obtenir avec une table de $g/2$ bits une précision de 2^{-g}.

Autre méthode intéressante, la méthode bipartie, qui, partant des premiers bits de d va chercher deux constantes dans une table et obtient l'approximation initiale avec une addition de ces deux constantes. On coupe d en 3 parties égales : $d = d_1 + d_2 2^{-k} + d_3 2^{-2k}$ (on suppose que l'on a $n = 3k$), vérifiant $0 \leq d_i \leq 1 - 2^{-k}$. On recherche une table f_A indexée par d_1 et d_2 et une table f_B indexée par d_1 et d_3 telle que $f_A + f_B$ approche $1/d$ avec une bonne précision.

10. Trouvez une méthode pour calculer de telles tables f_A et f_B.

11. Calculez ces tables pour $1 \leq x < 2$ et $n = 6$.

4 Calculs d'erreur

Dans cette partie, nous allons nous attarder sur les erreurs effectuées pendant les calculs.

12. Dans le cas de la méthode de Newton-Raphson, calculez la précision requise pour stocker complètement les divers résultats utilisés par l'algorithme, en supposant que $R[0]$ est codé sur a bits et que d est codé sur n bits. Quel est le problème ?

Nous chercherons dans cette partie à avoir une erreur relative ε_T inférieure à 2^{-s}. Cette erreur est la somme d'une erreur due à la méthode — cette erreur algorithmique sera notée ε_A — et une erreur générée par l'implémentation, notée ε_G. À l'étape j de l'algorithme, on aura $\varepsilon_T[j] = \varepsilon_A[j] + \varepsilon_G[j]$.

13. on suppose $\varepsilon_G[j]$ connue. Donnez $\varepsilon_T[m]$ en fonction des $\varepsilon_G[j]$ et de $\varepsilon_A[0]$ pour la méthode de Newton-Raphson.

14. de la même façon, établissez la même formule, mais cette fois pour la méthode de normalisation multiplicative.