ÉLÉMENT DE PORTFOLIO 01

Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Connolly, M.P., Higham, N.J., and Mary, T. (2021). Stochastic Rounding and its Probabilistic Backward Error Analysis. SIAM Journal on Scientific Computing, 43(1) : A566–A585.

URL de l'élément : `https://hal.science/hal-02556997`

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Ces travaux répondent à une conjecture vieille de 60 ans et montrent comment l'arrondi stochastique peut améliorer la précision des calculs, ce qui explique son succès en IA et autres applications utilisant les précisions faibles.

3 PRÉSENTATION DE CET ÉLÉMENT

Les calculs en précision finie sont sujets à l'accumulation d'erreurs d'arrondi : si une opération élémentaire individuelle (addition, soustraction, multiplication, etc.) produit une erreur d'ordre u (la précision machine), un calcul plus complexe combinant plusieurs de ces opérations élementaires cumule ces erreurs. En algèbre linéaire, la plupart des algorithmes impliquant des matrices d'ordre n (produit matrice-vecteur, résolution de système linéaire, etc.) peuvent générer une erreur au pire de l'ordre de nu. Pour des précisions faibles et/ou des problèmes de grande taille, cette borne est particulièrement préoccupante puisqu'elle suggère que le résultat est potentiellement entièrement incorrect.

Cependant, cette borne nu est une borne pire-cas : si n erreurs d'arrondi se cumulent effectivement linéairement quand elles sont de même signe, elles peuvent aussi s'annuler mutuellement si elles sont de signe opposé. Intuitivement, on peut espérer avoir environ la moitié des erreurs de chaque signe : on peut modéliser cette intuition en supposant que les erreurs d'arrondi sont des variables aléatoires d'espérance nulle, ce qui permet d'obtenir une borne d'erreur réduite de l'ordre de $\sqrt{n}u$. Cette ligne de pensée a été développée dès le début de l'ère digitale, dans les années 50, et a abouti à une célèbre conjecture de Wilkinson [5] déclarant que les constantes dépendant des dimensions dans les bornes d'erreur peuvent être remplacées par leur racine carrée. Cette observation est restée une conjecture, faute de preuve rigoureuse expliquant précisément sous quelles conditions cette meilleure borne tient, et avec quelle probabilité.

Dans des travaux en collaboration avec Nick Higham (Université de Manchester), nous avons prouvé la conjecture de Wilkinson. Dans une série de trois articles publiés dans *SIAM Journal of Scientific Computing* entre 2019 et 2021 [1,3,4], nous avons construit un modèle probabiliste des erreurs d'arrondis qui donne des conditions précises et suffisantes pour obtenir une borne d'erreur $\lambda\sqrt{n}u$, qui tient avec une probabilité connue $P(\lambda)$ qui se rapproche de 1 à vitesse exponentielle quand λ augmente (ainsi, la borne tient avec une très grande probabilité pour de modestes valeurs de λ). Cette nouvelle analyse d'erreur probabiliste permet aussi d'identifier certaines situations où le modèle n'est pas satisfait, et la borne pire-cas nu est atteinte.

Surtout, dans le troisième article [1], nous avons démontré un lien fort entre cette analyse probabiliste et l'arrondi stochastique. Ce mode d'arrondi, qui connaît un fort intérêt récent notamment en IA, consiste à arrondir chaque opération élémentaire de manière aléatoire ; nous nous intéressons plus particulièrement à une version de cet arrondi où la probabilité d'arrondir dans une direction ou l'autre est inversement proportionnelle à la distance de la quantité à arrondir. Or, dans [1], nous avons démontré que l'utilisation de l'arrondi stochastique force les erreurs d'arrondi à satisfaire le modèle probabiliste, et garantit donc une borne de l'ordre de $\sqrt{n}u$. L'arrondi stochastique est ainsi particulièrement attrayant dans les situations où la borne pire-cas est atteinte, il ainsi permet de réduire l'erreur d'un facteur $\sqrt{n}$ et de préserver un résultat de qualité acceptable, même en précision faible. Ceci explique donc son succès dans de nombreuses applications employant une précision faible, notamment pour l'entraînement de réseaux de neurones [2].

4 RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Michael P Connolly, Nicholas J Higham, and Théo Mary. Stochastic Rounding and its Probabilistic Backward Error Analysis. *SIAM Journal on Scientific Computing*, 43(1) :A566–A585, February 2021.

[2] Suyog Gupta, Ankur Agrawal, Kailash Gopalakrishnan, and Pritish Narayanan. Deep learning with limited numerical precision. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 1737–1746. JMLR.org, 2015.

[3] Nicholas J. Higham and Theo Mary. A new approach to probabilistic rounding error analysis. 41(5) :A2815–A2835, 2019.

[4] Nicholas J Higham and Théo Mary. Sharper Probabilistic Backward Error Analysis for Basic Linear Algebra Kernels with Random Data. *SIAM Journal on Scientific Computing*, 42(5) :A3427–A3446, October 2020.

[5] J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Notes on Applied Science No. 32, Her Majesty's Stationery Office, London, 1963. Also published by Prentice-Hall, Englewood Cliffs, NJ, USA. Reprinted by Dover, New York, 1994.