Institut de Recherche
en Informatique de Toulouse

Equipe APO
IRIT-CNRS
ENSEEIHT
2 rue C. Camichel
31071 TOULOUSE CEDEX

Approximations de rang faible randomisées en précision mixte

Contexte

Les applications modernes an analyse de données et calcul scientifique sont caractérisées par des volumes de données conséquents et une complexité opératoire extrêmement élevée qui rendent l'utilisation des supercalculateurs inévitable. Suite aux récentes évolutions technologiques, l'architecture des calculateurs à hautes performances devient de plus en plus hétérogène, c'est à dire que les processeurs sont souvent accompagnés par des unités de calcul *spécialisées* et donc capables d'atteindre des performances élevées mais seulement sur des applications particulières. Une forme de spécialisation est représentée par les unités de calcul à précision faible telle que la *half precision* permettant d'effectuer des calculs de manière plus efficace (moins de temps, de mémoire et d'énergie) mais avec une précision plus faible. Dans ce contexte, il est extrêmement important de concevoir des algorithmes avec une complexité opératoire faible et, en même temps, capables de tirer profit de la puissance des calculateurs modernes en exploitant leur hétérogénéité. Ce stage s'intéressera à l'étude d'algorithmes d'algèbre linéaire approchés à précision mixte, c'est à dire, des algorithmes efficaces, à complexité opératoire faible reposant sur l'utilisation simultanée de plusieurs arithmétiques en virgule flottante afin de tirer profit des unités de calcul à précision faible tout en fournissant des garanties sur la précision de la solution.

Sujet du stage

Les méthodes d'approximation de rang faible[3] sont un outil fondamental en calcul scientifique et analyse de données de par leur capacité à réduire la consommation de mémoire ainsi que la complexité des calculs dans de nombreux algorithmes. Ces techniques tirent profit de la décroissance rapide des valeurs singulières de certains types de matrices pour représenter celles-ci de manière compacte sous forme d'un produit de matrices de petit rang:

$$B \approx XY^T, \quad B \in \mathcal{R}^{n\times n}, \quad X,Y \in \mathcal{R}^{n\times r}, \quad r \ll n, \quad \|B - XY^T\| < \varepsilon \tag{1}$$

Bien que la décomposition en valeurs singulières (SVD) soit la méthode la plus précise pour calculer une telle représentation, en pratique d'autres algorithmes moins précis mais

moins coûteux sont utilisé comme la factorisation QR avec pivotage des colonnes (QR-CP). Des travaux récents [1] ont montré que la consommation de mémoire et la complexité calculatoire peuvent être réduits d'avantage en utilisant un format de stockage en virgule flottante en précision mixte. Dans ce format, p précisions différentes peuvent être utilisées en même temps: une précision élevée (par exemple, la précision double) est utilisée seulement pour représenter les données portant le plus d'information ($i = 1$) et des précisions plus faibles (par exemple, la précision simple ou half) sont employées pour les données les moins importantes:

$$B \approx \sum_{i=1}^{p} X_i Y_i^T, \qquad X_i, Y_i \in \mathcal{R}^{n \times r_i}, \qquad \sum_{i=1}^{p} r_i = r \tag{2}$$

Une analyse d'erreur rigoureuse montre que l'utilisation de la précision mixte ne dégrade pas la qualité de la représentation par rapport au cas en précision uniforme.

Objectif du stage

La représentation de rang faible en précision mixte peut être calculée en deux étapes: d'abord on calcule la représentation en précision uniforme (équation (1)) à l'aide d'une factorisation QR-CP en précision élevée et, ensuite, les matrices X et Y sont partitionnées et chaque part est convertie dans le précision correspondante. Cependant cette approche, relativement facile à implémenter, est peu efficace et parallélisable principalement parce que la factorisation QR-CP repose sur des opérations BLAS-2. De plus, cette factorisation est entièrement calculée en précision élevée. L'objectif de ce stage est de développer un algorithme à précision mixte pour le calcul de l'approximation (2) directement à partir de la matrice B initiale et sans passer par le format à précision uniforme (1). Cette approche permettra de bénéficier d'une réduction de la complexité opératoire et de la consommation de mémoire dans l'algorithme d'approximation. Pour atteindre cet objectif, une approche prometteuse consisterait à utiliser des techniques de randomisation[2] au sein de la factorisation QR-CP; ces techniques permettent de réduire considérablement l'utilisation d'opérations BLAS-2 et de changer plus facilement la précision utilisée au cours de la factorisation. Les travaux du stage seront concernés par la conception de l'algorithme, son analyse afin d'en évaluer la robustesse et la précision et son implémentation efficace pour des processeurs classiques et, possiblement, des GPUs.

Organisation

Affectation	:	equipe IRIT-APO
Encadrement	:	en collaboration avec le laboratoire LIP6 (Paris)
Durée	:	5 à 6 mois
Rémunération	:	environ 600€ / mois
Date d'embauche prévue	:	Fevrier/Mars 2023
Lieu	:	ENSEEIHT-IRIT, 2 rue Claude Camichel, 31000 Toulouse
Contact	:	Alfredo Buttari <alfredo.buttari@irit.fr>

References

[1] Patrick Amestoy, Olivier Boiteau, Alfredo Buttari, Matthieu Gerest, Fabienne Jézéquel, Jean-Yves L'Excellent, and Theo Mary. Mixed precision low-rank approximations and their application to block low-rank LU factorization. *IMA Journal of Numerical Analysis*, 08 2022. drac037.

[2] Per-Gunnar Martinsson, Gregorio Quintana OrtÍ, Nathan Heavner, and Robert van de Geijn. Householder qr factorization with randomization for column pivoting (hqrrp). *SIAM Journal on Scientific Computing*, 39(2):C96–C115, 2017.

[3] Theo Mary. *Block Low-Rank multifrontal solvers: complexity, performance, and scalability*. PhD thesis, EDMITT, Université Paul Sabatier, Toulouse, France, November 2017.