

Utilisation de compression Block Low-Rank multiprécision pour résoudre efficacement des systèmes linéaires

Patrick AMESTOY, MUMPS Technologies - Lyon
Olivier BOITEAU, EDF R&D - Palaiseau **Alfredo BUTTARI**, IRIT - Toulouse
Matthieu GEREST, EDF R&D / LIP6 - Paris
Fabienne JÉZÉQUEL, LIP6 - Paris
Jean-Yves L'EXCELLENT, MUMPS Technologies - Lyon
Théo MARY, LIP6 - Paris

La compression Block Low-Rank (BLR) tire parti d'une propriété de certaines matrices, dont les blocs extra-diagonaux sont majoritairement de rang numérique faible. On peut alors représenter ces blocs de façon moins coûteuse, sous forme d'une approximation de rang faible telle qu'une SVD tronquée ou une décomposition QR tronquée, tout en contrôlant l'erreur introduite. Une telle méthode a permis de réduire le nombre d'opérations et le coût mémoire de l'algorithme de factorisation LU d'une matrice [5, 4]. Par la suite, la compression BLR a été introduite dans des noyaux de calcul de solveurs directs pour la résolution de systèmes creux de grande taille, notamment MUMPS [3, 2].

Une nouvelle variante de la compression BLR est proposée ici, utilisant à bon escient plusieurs formats de précision pour stocker les coefficients. La plupart des entrées et des opérations peuvent être converties en précision faible (simple et demi-précision), au lieu de la précision double plus couramment utilisée. En effet, une analyse théorique ainsi que des expérimentations permettent de justifier que l'erreur de compression reste du même ordre de grandeur [1]. En intégrant cette variante dans une factorisation LU, on peut ainsi améliorer encore ses performances en termes de coût de stockage et de temps de calcul, en tirant parti de la vitesse d'exécution accrue des opérations en précisions faibles.

- [1] P. Amestoy, O. Boiteau, A. Buttari, M. Gerest, F. Jézéquel, J.-Y. L'Excellent, T. Mary. *Mixed Precision Low Rank Approximations and their Application to Block Low Rank LU Factorization*, 2021. Preprint. <https://hal.archives-ouvertes.fr/hal-03251738>, submitted at IMA Journal of Numerical Analysis.
- [2] P. Amestoy, A. Buttari, J.-Y. L'Excellent, T. Mary. *Performance and Scalability of the Block Low-Rank Multifrontal Factorization on Multicore Architectures*. ACM Transactions on Mathematical Software, **45**, 2 :1-2 :26, 2019.
- [3] P. Amestoy, I. S. Duff, J. Koster, J.-Y. L'Excellent. *A fully asynchronous multifrontal solver using distributed dynamic scheduling*. SIAM Journal on Matrix Analysis and Applications, **23(1)**, 15–41, 2001.
- [4] T. Mary. *Block low-rank multifrontal solvers : complexity, performance, and scalability*. PhD thesis, Université de Toulouse, 2017.
- [5] C. Weisbecker. *Improving multifrontal solvers by means of algebraic block low-rank representations*. PhD thesis, Institut National Polytechnique de Toulouse, 2013.