

Demande de Bourse CIFRE

Description Technique

Composabilité en Algèbre Linéaire Haute Performance - Application à l'Aéroacoustique et à l'Électromagnétisme

Encadrement (sous réserve d'accord de l'école doctorale) : Guillaume SYLVAND (Airbus Central R&T), Luc GIRAUD (CONCACE, INRIA)

Contexte industriel :

Le laboratoire de Recherche et Technologie Airbus Central R&T développe depuis de nombreuses années des outils de modélisation destinés à être utilisés dans les branches du groupe Airbus (Airbus Helicopters, Airbus Commercial, Airbus Defence & Space) ainsi que dans ses filiales et partenaires (ArianeGroup, MBDA). En particulier, le département Virtual Product Engineering (VPE) développe des logiciels de simulation de propagation d'ondes électromagnétiques (la suite ASERIS) et acoustiques (la suite ACTIPOLE) utilisés très activement pour les applications en placement d'antenne, compatibilité électromagnétique, furtivité radar pour ce qui est de l'électromagnétisme, et en modélisation des bruits de réacteurs d'avions civils en ce qui concerne l'acoustique.

Le contexte concurrentiel très fort sur les marchés concernés (défense, espace, aviation civile, etc.), un cadre réglementaire de plus en plus contraignant (normes sur les rayonnements acoustiques et électromagnétisme), la recherche de robustesse dans les calculs (via des techniques de gestions des incertitudes): tous ces facteurs imposent un usage de plus en plus important de la simulation dans le développement des produits chez Airbus. Les outils logiciels se doivent de s'améliorer pour accompagner ces usages, tout en respectant les contraintes environnementales et les évolutions matérielles des ressources de calcul.

Contexte Scientifique :

L'augmentation de la puissance de calcul disponible et l'utilisation généralisée de la simulation numérique dans la R&D ont conduit à l'émergence d'une grande diversité d'algorithmes nouveaux (on peut citer p.ex. les algorithmes *randomized* basés sur des approches statistiques, ceux basés sur des techniques d'apprentissages et d'IA, les approches à arithmétique mixte, les méthodes hiérarchiques, etc.) venant se rajouter aux algorithmes dits classiques. Si on ajoute à cela le choix dans le format de stockage des données (dense, creux, compressé, hiérarchique, etc.), le parallélisme (à mémoire partagée et distribuée), les multiples arithmétiques (simple-, double-, demi-précision), et les architectures de calcul hétérogènes (CPU, GPU, TPU, etc.), on voit que l'on fait face à une combinatoire extrêmement complexe si on veut en tirer la quintessence. Le projet CONCACE, commun à Airbus, Cerfacs et Inria, propose de s'attaquer à ce problème en tirant parti des outils et des langages de développement modernes pour concevoir des expressions de haut niveau des algorithmes parallèles complexes. Alors que l'approche traditionnelle du calcul intensif consiste à exploiter

pleinement le matériel, notre approche complémentaire permettra une composabilité plus riche des méthodes numériques, permettant d'exploiter pleinement les algorithmes numériques existants et nouveaux. Dans le cadre de cette thèse, nous proposons de tenter une première mise en œuvre de cette méthodologie dans le solveur H-matrice exploité par les logiciels ASERIS et ACTIPOLE précités.

La méthode mathématique utilisée dans ces outils est celle des équations intégrales dans le domaine fréquentiel, résolues par éléments finis de frontières [6]. Cette approche permet une grande précision dans les résultats, et une excellente fidélité dans la prise en compte de la géométrie des objets traités. En revanche, elle conduit à devoir résoudre des systèmes linéaires denses, une tâche dont le coût calculatoire est vite prohibitif : il croît comme le cube du nombre d'inconnues du problème à traiter, nombre d'inconnues qui lui même croît comme le carré de la fréquence des ondes simulées et de la dimension des objets traités. On atteint vite des durées de calcul de plusieurs semaines... Pour pallier ces limitations, la communauté scientifique a développé des solveurs performants spécifiques à ce type d'application : dans les années 2000, la méthode multipôle rapide [3] a permis un gain d'un facteur 100 sur le nombre d'inconnues des problèmes que l'on pouvait traiter. Dans les années 2010, la méthode des H-matrices [1] a conduit à une accélération considérable de la résolution des problèmes comportant au plus quelques millions d'inconnues. Toutefois, tous les algorithmes ne se prêtent pas au passage à l'échelle avec autant d'aisance : ainsi, alors que la méthode multipôle rapide peut traiter plusieurs centaines de millions d'inconnues sans soucis, la méthode des H-matrices - bien que beaucoup plus rapide - se limite aujourd'hui à environ 10^7 inconnues.

Sujet de Recherche :

Nous proposons de travailler sur la méthode des H-matrices afin d'augmenter sa performance et son efficacité sur tous les cas d'applications existant, et ainsi d'en faire la méthode de résolution de référence dans le domaine de la propagation des ondes. Pour ce faire, nous exploiterons l'idée de composabilité à travers 4 axes de travail correspondant soit à des points de blocages, soit à des possibilités d'amélioration offertes par des innovations (informatiques ou mathématiques) récentes.

- *Parallélisme distribué à base de tâches* [5] : L'implémentation actuelle des H-matrices se heurtent à la difficulté de paralléliser efficacement le calcul sur machine à mémoire distribuée. Hors, cela est crucial afin de diminuer les temps de calcul et d'augmenter la taille des problèmes pouvant être traités. Nous nous proposons d'investiguer cette difficulté en travaillant sur les algorithmes au cœur de la méthode (notamment la façon de gérer les très gros blocs de matrices, et de factoriser les blocs diagonaux), sur les outils logiciels utilisés pour réaliser les communications [7], et sur la façon de distribuer le stockage des données entre les nœuds.
- *Arithmétique mixte* [4] : Les architectures modernes de calcul sont de plus en plus conçues pour traiter efficacement les problèmes de type Intelligence Artificielle (IA) et Machine Learning (ML). Cela a conduit à l'introduction d'unités de calcul (FPU) travaillant sur 16 bits, c'est-à-dire sur des nombres représentés avec très peu de chiffres significatifs (3 ou 4 en pratique, IA et ML se contentant parfaitement de ce faible niveau

de précision). Nous nous proposons d'étudier la possibilité d'utiliser de manière conjointe les différentes précisions de calcul disponibles (16, 32, 64 bits) dans les processeurs modernes afin de minimiser les temps de calcul (les calculs en 16 bits étant bien évidemment les plus rapides) tout en garantissant la précision du résultat final (au niveau exigé par l'application et par l'utilisateur).

- *Algorithmes Randomized* : Cette classe d'algorithme [2] a le vent en poupe actuellement, car elle est très efficace dans le domaine du machine learning, et de plus elle est parfaitement adaptée aux architectures de calcul actuelles (*via* son utilisation intensive de produits matrice-matrice). L'utilisation de ces méthodes dans les H-matrices interviendrait aux niveaux des briques de base de l'algorithme, et permettrait de gérer de façon plus efficace les gros blocs de matrices qui apparaissent lorsque l'on traite des cas de grande taille. *In fine*, cela conduirait à l'amélioration des performances sur ces cas, et à la possibilité de traiter des cas de plus grande taille.
- *Calcul hétérogène* [5] : L'approche à base de tâche présente dans le solveur H-matrice permet d'aborder relativement simplement le calcul sur architecture hétérogène (CPU, GPU, TPU, processeurs à cœurs hétérogènes comme le M1). Il conviendra de voir dans quelle mesure cette approche peut permettre d'améliorer la performance globale de l'exécution (mesure suivant différentes métriques : temps de calcul ou énergie consommée).

Objectifs :

Au travers de ces 4 thèmes, nous souhaitons tirer la quintessence des avancées récentes en matière d'algorithmes, de sciences du numérique et d'informatiques, au service des outils Airbus reposant sur la méthode des H-matrices. En termes de taille, notre ambition est de pouvoir traiter avec ceux-ci des cas comportant *au moins* 100 millions d'inconnues, ce qui pourrait correspondre à une simulation électromagnétique à 10 GHz sur un avion de chasse. En termes d'empreinte carbone, nous souhaitons gagner un facteur 10 sur un problème à 5 millions d'inconnues.

Planification des travaux :

Les 4 thèmes ci-dessus suivront la planification suivante (susceptible d'être adaptée dynamiquement en fonction de l'avancement des travaux) :

- T0 : Parallélisme distribué à base de tâches
- T0+8 : Arithmétique mixte
- T0+16 : Algorithmes Randomized
- T0+24 : Calcul hétérogène
- T0+32 : rédaction finale (manuscrit et soutenance)
- T0+36 : fin de thèse

Organisation :

La thèse se ferait dans le cadre d'un financement CIFRE (contrat Airbus), et serait hébergée chez Airbus Central R&T à Issy-Les-Moulineaux, ou éventuellement dans le projet CONCACE

(Inria Bordeaux Sud-Ouest, ou Cerfacs à Toulouse). L'encadrement serait assuré conjointement par un chercheur de CONCACE (L. Giraud) et un ingénieur d'Airbus (G. Sylvand).

Candidat :

Le candidat devra avoir une formation en HPC (calcul haute performance), algorithmie, calcul parallèle avec une connaissance de base des approches de type éléments finis. Une bonne connaissance des langages de programmation et des outils utilisés dans ces disciplines (tels que C, C++, fortran, python, shell, jupyter, docker, ...) sera nécessaire, de même que des connaissances en algèbre linéaire et résolution de systèmes linéaires (solveurs directs et itératifs). Le candidat devra savoir concevoir un plan d'expérience et mener une étude scientifique, faire des présentations scientifiques devant un auditoire, et rédiger des articles scientifiques et des rapports techniques (langage LaTeX).

Bibliographie :

[1] Lizé, B. "Fast direct solver for the boundary element method in electromagnetism and acoustics : H-Matrices. Parallelism and industrial applications." Theses, Université Paris-Nord - Paris XIII, June 2014.

[2] Per-Gunnar Martinsson. Randomized methods for matrix computations. Mathematical Institute. The University of Oxford. January 2018.

[3] Sylvand, G. "La méthode multipôle rapide en électromagnétisme : performances, parallélisation, applications". Ecole Nationale des Ponts et Chaussées. 2002.

[4] Abdelfattah et al. "A Survey of Numerical Methods Utilizing Mixed Precision Arithmetic", International Journal of High Performance Computing Applications, 2021. <https://arxiv.org/abs/2007.06674> .

[5] Augonnet et al. "StarPU: A Unified Platform for Task Scheduling on Heterogeneous Multicore Architectures" Concurrency and Computation: Practice and Experience, Special Issue: Euro-Par 2009, 23:187-198, February 2011.

[6] Nedelec "Acoustic and Electromagnetic Équations", Springer, 2001.

[7] Beauchamp, "Portage de StarPU sur la bibliothèque de communication NewMadeleine", PhD Université de Bordeaux, 2017. <https://hal.inria.fr/hal-01587584>