

Propositions de MFE en Analyse Numérique

SUJETS (accessibles à tous les étudiants)

English version pages 5–8

1 Méthode multi-grilles pour GPU basée sur un format hybride stencil-CSR.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

2 Partitionnement pour méthode multi-grilles algébrique en mode multithread.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

3 Solveur de grille grossière pour méthodes multi-grilles parallèles: grands agrégats vs décomposition de domaine.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

4 GMRES et déflation optimale.

[Yvan Notay (ynotay@ulb.ac.be)]

5 Méthode multi-grilles parallèle pour les problèmes curl-curl.

[Artem Napov (anapov@ulb.ac.be), Yvan Notay (ynotay@ulb.ac.be)]

6 Parallélisation d'une méthode de factorisation de Cholesky incomplète.

[Artem Napov (anapov@ulb.ac.be)]

7 Méthode multi-grilles algébrique pour le calcul des plasmas.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

CONTEXTE

Certains des sujets proposés sont liés aux **méthodes multi-grilles**. Ces méthodes sont utilisées pour la résolution de très grands systèmes linéaires, avec, typiquement, plusieurs millions voire milliards d'inconnues. Le Service de Métrologie Nucléaire dispose d'une expertise mondialement reconnue dans le domaine des méthodes multi-grilles. En particulier, il développe le logiciel AGMG, qui a plusieurs centaines d'utilisateurs de part le monde.

Les méthodes multi-grilles ont pour principe de base d'effectuer la résolution d'un système linéaire donné en utilisant une hiérarchie de systèmes de plus en plus petits. Les sujets proposés dans ce cadre portent en général sur des méthodes *basées sur l'agrégation*, qui construisent progressivement une hiérarchie de systèmes de plus en plus petits en groupant les inconnues en agrégats; il s'agit de variantes multi-grilles à la fois simples et particulièrement efficaces. Comme les agrégats sont formés automatiquement sur base de la matrice du système linéaire, elles sont par ailleurs beaucoup plus flexibles que les méthodes multi-grilles basées sur une hiérarchie de grilles prédéfinie.

Tous les sujets proposés sont directement liés aux activités de recherche qui ont lieu dans le Service de Métrologie Nucléaire.

BRÈVE PRÉSENTATION : pages 2–4

1 Méthode multi-grilles pour GPU basée sur un format hybride stencil-CSR

Traditionnellement, le calcul numérique est géré par les CPUs des ordinateurs. Il y a cependant un intérêt économique et écologique à développer le calcul numérique sur les processeurs graphiques (GPUs). En effet, ces derniers sont relativement bon marché et offrent une puissance de calcul par watt consommé beaucoup plus importantes que les CPUs. Cependant, comme leur nom l'indique, elles sont conçues pour gérer l'affichage graphique et pas le calcul numérique. Celui-ci ne peut en fait pleinement profiter de la puissance de calcul que si la structure de donnée est adaptée. Par exemple, les calculs avec des matrices creuses seront beaucoup plus efficaces si la matrice est stockée en format stencil plutôt qu'en un format général comme le format CSR.

Pour ce travail, on développera d'abord un format hybride stencil-CSR qui combine l'efficacité du format stencil avec la généralité du format CSR. Le but sera alors d'exploiter ce format pour développer une méthode multi-grilles efficace sur une architecture hybride GPU-CPU.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

2 Partitionnement pour méthode multi-grilles algébrique en mode multithread

Le logiciel AGMG existe en différentes versions : purement séquentielle, en version parallèle pour mémoire partagée, et en version parallèle pour mémoire distribuée. La version pour architectures à mémoire partagée est basée sur le «multithreading», et a l'avantage d'avoir la même séquence d'appel que la version séquentielle ; *i.e.* elle permet de bénéficier de l'accélération parallèle offerte par les processeurs modernes sans modification du programme appelant. Elle s'avère cependant quelque peu moins robuste que la version séquentielle parce que, concrètement, le logiciel répartit la matrice du système à résoudre sur les différents «threads», et que cela induit des changements algorithmiques qui peuvent affecter les performances si cette répartition n'est pas appropriée.

Le but du travail est de développer une version qui effectue un partitionnement de la matrice adapté en fonction du problème. On développera une approche originale qui utilise une version séquentielle simplifiée d'AGMG pour guider le partitionnement. Les résultats seront comparés à ceux obtenus en utilisant un logiciel de partitionnement standard comme METIS. Il s'agit d'un travail à portée essentiellement pratique.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

3 Solveur de grille grossière pour méthodes multi-grilles parallèles: grands agrégats vs décomposition de domaine

Dans sa version massivement parallèle, AGMG résout le système sur la grille la plus grossière avec une méthode itérative à deux niveaux basée sur des très grands agrégats (une grille super grossière est formée en agrégeant ensemble toutes les inconnues sur un même processeur). Cette méthode est en fait assez proche d'une méthode de décomposition

de domaine. Le travail proposé consiste à explorer cette analogie, et en particulier à analyser comment la convergence est affectée lorsqu'on utilise des agrégats qui se chevauchent (une approche standard pour les méthodes de décomposition de domaine).

Le travail comporte donc deux aspects : un aspect numérique, qui consiste à développer des programmes (Matlab/Octave/Python) permettant l'évaluation pratique des différentes stratégies ; et un aspect théorique, qui nécessite de bien assimiler les différentes théories de convergence afin de pouvoir les comparer entre elles et avec les résultats pratiques.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

4 GMRES et déflation optimale

La déflation est une procédure algébrique qui accélère les méthodes de résolution de systèmes linéaires basées sur les sous-espaces de Krylov comme GMRES. Elle revient à ajouter au sous-espace de Krylov des vecteurs connus comme étant des «bonnes» directions de recherche. Algébriquement, le processus est assez proche d'une correction de grille grossière, à ceci près que les matrices de restriction et de prolongation ne contiennent qu'un petit nombre de lignes et colonnes, respectivement.

En général, sont considérées comme «bonnes» directions de recherche les approximations des vecteurs propres associées aux valeurs propres les plus basses. Cependant, des résultats récents montrent que, dans le cadre des méthodes multi-grilles, ce type de choix est loin d'être optimal lorsque les matrices ne sont pas symétriques.

Le but du travail est d'étendre ces résultats à la déflation, et, d'une manière plus générale, de déterminer quels sont les vecteurs qui maximisent l'efficacité de ce processus.

[Yvan Notay (ynotay@ulb.ac.be)]

5 Méthode multi-grilles parallèle pour les problèmes curl-curl

Les problèmes aux limites de type curl-curl résultent de l'approximation des régimes quasi stationnaires des équations de Maxwell. Ils sont typiquement utilisés pour modéliser des machines électriques (moteurs, transformateurs, etc.). Une nouvelle méthode multi-grilles basée sur l'agrégation a récemment été proposée pour la résolution numérique des problèmes discrets correspondants.

Le but de ce travail est d'explorer la parallélisation de la nouvelle méthode, c'est-à-dire d'explorer la possibilité de réduire le temps d'exécution en effectuant des tâches différentes sur des processeurs différents. La difficulté principale attendue est liée à l'usage du cycle K, dont la parallélisation sur des machines à grand nombre de processeurs n'est pas triviale. Néanmoins, une méthode basée sur le cycle K pour le problème de Poisson a déjà été parallélisée avec succès, et nous allons utiliser la démarche correspondante comme point de départ. Dans un premier temps, le travail sera restreint au cas d'un algorithme modèle appliqué à un problème modèle. Notons qu'une mise en oeuvre parallèle de la méthode n'est pas obligatoire mais, si elle est faite, le code résultant pourra être testé sur des machines massivement parallèles.

[Artem Napov (anapov@ulb.ac.be), Yvan Notay (ynotay@ulb.ac.be)]

6 Parallélisation d'une méthode de factorisation de Cholesky incomplète

Une factorisation incomplète est typiquement utilisée dans la résolution des grands systèmes linéaires creux lorsqu'aucun preconditionneur plus spécialisé (comme multi-grilles) n'est disponible. En particulier, si la matrice du système est symétrique et définie positive (SPD), il est commun d'utiliser une factorisation incomplète de Cholesky. Malheureusement, certaines factorisations de Cholesky incomplètes peuvent ne pas exister pour certains problèmes et valeurs du seuil d'approximation.

Dans ce travail nous considérons une factorisation de Cholesky incomplète basée sur les approximations orthogonales de faible rang. Cette factorisation existe pour toute matrice SPD et toute valeur du seuil (elle est *breakdown-free*), et se compare favorablement à l'état de l'art. Une implémentation de cette factorisation en langage C a déjà été développée. Le travail consiste à développer une version parallèle pour machines à mémoire partagée. La stratégie de parallélisation suggérée est semblable à celle utilisée pour une factorisation bloc classique – utiliser le parallélisme de l'arbre de l'élimination et paralléliser la factorisation incomplète des grands blocs. La place de l'approximation dans la séquence des opérations est un élément additionnel à prendre en considération.

[Artem Napov (anapov@ulb.ac.be)]

7 Méthode multi-grilles algébrique pour le calcul des plasmas

Le code de calcul TOKAM3X simule la physique des tokamaks, et est utilisée dans le développement du futur réacteur de fusion ITER. Environ 50% du temps de calcul est passé à résoudre des systèmes linéaires. Dans le cadre du centre d'excellence européen EoCoE («Energy oriented center of excellence»), le logiciel AGMG a été intégré à ce code afin d'accélérer cette partie des calculs. Cela conduit à une amélioration notable, mais on observe qu'AGMG reste sous performant à cause de la structure particulière des systèmes linéaires. Cette dernière provient des équations physiques du modèle, caractérisées par une forte anisotropie non alignée avec la grille de discrétisation.

Le but du travail est d'explorer différentes idées pour améliorer AGMG dans ce contexte particulier, notamment en exploitant la connaissance de la direction de couplage dominant.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

Master theses in Numerical Analysis (English version)

SUBJECTS

1 Multigrid method for GPU based on a hybrid stencil-CSR matrix format.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

2 Partitiobning for an algebraic multigrid method in a multi-thread mode.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

3 Coarse grid solver for parallel multigrid methods: large aggregates vs. domain decomposition.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

4 GMRES and optimal deflation.

[Yvan Notay (ynotay@ulb.ac.be)]

5 Parallel aggregation-based multigrid for curl-curl problem.

[Artem Napov (anapov@ulb.ac.be)]

6 Parallelisation of an incomplete Cholesky factorization method.

[Artem Napov (anapov@ulb.ac.be)]

CONTEXT

Some of the master thesis subjects listed above are related to **multigrid methods**. These methods are used for the solution of large linear systems of equations, with typically several millions to several billions of unknowns. The *Service de Métrologie Nucléaire* team has an internationally recognized expertise in the field of multigrid methods. In particular, the team develops the AGMG code, which has hundreds of users around the world.

The basic principle behind multigrid methods is to solve a given linear system using a hierarchy of decreasing in size linear systems. The above subjects within this framework are in general related to *aggregation-based* multigrid methods, which construct a hierarchy of decreasing in size linear systems by grouping unknowns into aggregates; this yields simple yet efficient multigrid variants. Because the aggregates are automatically formed on the basis of the linear system matrix, these methods are on the other hand much more flexible than the multigrid methods based on a predefined hierarchy of grids.

All master thesis subjects are directly connected to the research activities of the team.

SHORT DESCRIPTION : page 6–8

1 Multigrid method for GPU based on a hybrid stencil-CSR matrix format

Traditionally, numerical computation is managed by the computer CPUs. There is, however, an economic and ecological interest in developing numerical computation on graphics processing units (GPUs). Indeed, the latter are relatively cheap and offer much higher computing power per watt consumed than CPUs. However, as their name suggests, they are designed to handle graphical display and not numerical computation. This latter can actually fully benefit from the computing power only if the data structure is adapted. For example, sparse matrix computations are considerably more efficient if the matrix is stored in stencil format rather than in a general format such as the CSR format.

For this work, one will first develop a hybrid stencil-CSR format that combines the efficiency of the stencil format with the generality of the CSR format. The goal is then to exploit this format to develop an effective multigrid method for a hybrid GPU-CPU architecture.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

2 Partitiobning for an algebraic multigrid method in a multithread mode

The AGMG software exists in different versions : purely sequential, parallel for shared memory, parallel for distributed memory. The version for shared memory architectures is based on multithreading, and it has the advantage of having the same calling sequence as the sequential version ; i.e., it allows one to benefit from the parallel acceleration offered by modern processors without modification of the calling program. This version is, however, somehow less robust than the sequential version because, concretely, the software distributes the system matrix on the different threads, and this induces some algorithmic changes that may affect performance if this distribution is not appropriate.

The aim of the proposed thesis is to develop a version that makes a partitioning of the matrix adapted to the problem at hand. One will develop an original approach that uses a simplified sequential version of AGMG to guide the partitioning. The results will be compared with those obtained using standard partitioning software like METIS. This work is mainly practical.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

3 Coarse grid solver for parallel multigrid methods: large aggregates vs. domain decomposition

In its massively parallel version, AGMG solves the system on the coarsest grid using a two-level iterative method based on very large aggregates (a super coarse grid is formed by aggregating together all unknowns on a same processor). This method is in fact somewhat similar to a domain decomposition method. The work consists in exploring this analogy, in particular in analyzing how the convergence is affected when one uses overlapping aggregates (a standard approach for domain decomposition methods).

The work has two components : a numerical component, which requires to develop programs (Matlab/Octave/Python) for the practical assessment of the considered strategies ; and a theoretical component, which requires a proper study of the different convergence theories, in order to compare these theories between them and with practical observations.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]

4 GMRES and optimal deflation

Deflation is an algebraic procedure that accelerates Krylov subspace methods for solving linear systems, such as GMRES. It amounts to adding to the Krylov subspace vectors known as “good” search direction. Algebraically, the process is quite close to a coarse grid correction, except that the restriction and the prolongation matrices contain only a small number of rows and columns, respectively.

In general, “good” search directions are approximations of the eigenvectors associated with the lowest eigenvalues. However, recent results show that, in the context of multigrid methods, this type of choice is far from optimal when the matrices are nonsymmetric.

The aim of the work is to extend these results to deflation, and, more generally, to determine which vectors maximize the efficiency of this process.

[Yvan Notay (ynotay@ulb.ac.be)]

5 Parallel aggregation-based multigrid for curl-curl problem

The curl-curl boundary value problems arise when eddy current approximation is applied to Maxwell equations. They are typically used when modelling electrical machines (engines, transformers, etc.). A new aggregation-based multigrid method has been recently proposed for the numerical solution of the corresponding discrete problems.

The aim of the work is to explore the parallelization of the new method, i.e., to explore the possibility to reduce the execution time by running tasks on different processors. The main difficulty is related to the use of the K-cycle on machines with a large number of processors. However, methods based on the K-cycle have already been successfully parallelized for the Poisson problem, and we shall use the corresponding approach as a starting point. At first, we shall restrict ourselves to the model algorithm for a model problem. Note that the actual parallel implementation of the method is not mandatory, but if it is realised, the resulting code could be tested on massively parallel machines.

[Artem Napov (anapov@ulb.ac.be), Yvan Notay (ynotay@ulb.ac.be)]

6 Parallelisation of an incomplete Cholesky factorization method

An incomplete factorization is typically used for the solution of the large sparse linear systems when more specialized preconditioners (such as multigrid) are not available. In particular, if the system matrix is symmetric and positive definite (SPD), an incomplete Cholesky factorization is a common choice. Unfortunately, for some problems and some

values of the approximation threshold, the Cholesky factorization may fail to exist, a phenomenon known as break-down.

In this work we consider an incomplete Cholesky factorization based on orthogonal low-rank approximations. Such a factorization is breakdown-free (it exists for any SPD matrix and for any approximation threshold) and compares favorably with the state of the art. An implementation in C language of this factorization has already been developed. The work consists in developing a parallel version for shared-memory machines. The suggested parallelisation strategy is similar to that of the classical block factorizations – use the parallelism of the elimination tree and extract parallelism from the factorization of large blocks. The place of approximation in the overall operation sequence is an additional element to be considered.

[Artem Napov (anapov@ulb.ac.be)]

7 Algebraic multi-grid method for plasmas computation

The TOKAM3X code simulates the physics of the tokamaks, and is used in the development of the future ITER fusion reactor. About 50% of the computing time is spent in solving linear systems. Within the framework of the European Center of Excellence EoCoE (“Energy oriented center of excellence”), the AGMG software has been integrated into this code in order to speed up this part of the computation. This leads to a significant improvement, but it is observed that AGMG remains underperforming because of the particular structure of the linear systems. The latter comes from the physical equations of the model, characterized by a strong anisotropy not aligned with the discretization grid.

The aim of the work is to explore different ideas for improving AGMG in this particular context, notably by exploiting the knowledge of the dominant coupling direction.

[Yvan Notay (ynotay@ulb.ac.be), Artem Napov (anapov@ulb.ac.be)]