CSMA 2017
13ème Colloque National en Calcul des Structures
15-19 Mai 2017, Presqu'île de Giens (Var)

Méthodes par blocs et recyclage pour la résolution de très grands systèmes linéaires

P. Jolivet[1], R. Haferssas[2], F. Nataf[2], P.-H. Tournier[2]

[1] *IRIT–CNRS pierre.jolivet@enseeiht.fr*
[2] *LJLL–UPMC et EPI ALPINES–Inria Paris {haferssas,nataf,tournier}@ljll.math.upmc.fr*

Résumé — Certaines avancées algorithmiques permettent maintenant la résolution de systèmes distribués de très grandes tailles en utilisant des méthodes de Krylov sophistiquées comme le GMRES par blocs, le Gradient Conjugué par blocs sans panne... Nous présenterons certains de ces méthodes, ainsi que des résultats numériques obtenus dans le cadre de diverses applications en mécanique des solides, en dynamique des fluides, ou encore en imagerie tomographique.
Mots clés — recyclage ; méthodes de Krylov par blocs ; décomposition de domaine ; méthodes multigrilles algébriques.

1 Introduction

La discrétisation d'Équations aux Dérivées Partielles pour modéliser des phénomènes physiques entraîne la génération de grands systèmes qui ne peuvent pas être résolus directement. Il faut donc avoir recours à des préconditionneurs performants et à des méthodes itératives efficaces. Depuis plusieurs années, l'étude et la conception des préconditionneurs qui passent à l'échelle a suscité beaucoup d'intérêts. Cependant, on se sert généralement de méthodes itératives "basiques" comme le GMRES ou le Gradient Conjugué (CG) pour la phase de résolution à proprement parler. De nombreuses méthodes plus poussées ont tout de même vu le jour pour pipeliner les communications [1, 2], limiter les synchronisations [3, 4], faire décroître le nombre d'itérations en utilisant plusieurs directions de recherche [5, 6], recycler des espaces de Krylov pour résoudre des systèmes successifs [7, 8]... Des méthodes itératives permettant de résoudre des systèmes avec plusieurs seconds membres ont aussi été proposées [9, 10, 11].

1.1 État de l'art

Il existe plusieurs librairies numériques pour l'algèbre linéaire creuse distribuée. Comme nous allons le mettre en exergue dans la liste ci-dessous, la plupart ne permettent cependant pas de traiter des problèmes avec multiples seconds membres, ou d'utiliser des outils de recyclage d'espaces de Krylov pour faciliter la convergence de résolutions successives.

— *hypre* [12], DUNE [13], et PARALUTION [14] disposent seulement de méthodes traditionnelles comme GMRES ou CG.
— PETSc [15] dispose de quelques méthodes plus avancées comme Loose GMRES [16] ou Deflated GMRES [17, 18], mais telles qu'implémentées dans la librairie, ces méthodes ne peuvent pas être utilisées pour recycler de l'information d'un système à un autre.
— Trilinos [19], à travers sa sous-librairie Belos [20], est une des librairies qui propose certaines des méthodes les plus avancées comme GMRES par blocs ou GCRO-DR. Il existe cependant certaines limitations importantes comme : l'impossibilité d'utiliser des préconditionneurs variables, la difficulté d'utiliser la librairie dans un code n'utilisant pas Trilinos.

1.2 Contributions

Dans cet exposé, nous faisons part de certaines avancées algorithmiques dans la libraire HPDDM [21] qui permettent :

— d'utiliser le recyclage d'espaces de Krylov, éventuellement avec des préconditionneurs variables. Nous détaillerons les performances de la méthode via des cas tests académiques en mécanique des structures et en dynamique des fluides sur plusieurs milliers de processeurs pour résoudre des systèmes avec plusieurs centaines de millions d'inconnues.
— de combiner recyclage et méthodes itératives par blocs, éventuellement avec des préconditionneurs variables. Le cas test sera cette fois-ci tiré d'une application en imagerie tomographique qui nécessitent la résolution d'un problème inverse avec plusieurs seconds membres.
— d'utiliser la librairie dans un grand nombres d'applications, grâce à des interfaces en C, C++, Python et Fortran. En particulier, nous nous servons d'un exemple PETSc (écrit en C), qui tire entièrement profit de certaines des capacités les plus avancées de la librairie comme le solveur multigrille GAMG, tout en utilisant HPDDM pour résoudre les systèmes linéaires.

Une partie de ces résultats a déjà mené à une publication [22].

2 Recyclage pour les méthodes de Krylov

Les travaux présentés dans cette communication sont principalement basés sur la méthode GCRO-DR (pour Generalized Conjugate Residual method with inner Orthogonalization with Deflated Restarting) [23]. Celle-ci a été développé dans le contexte de la modélisation de fractures et pour la fatigue des structures, lorsqu'il est généralement nécessaire de résoudre une séquence de systèmes linéaires :

$$A_i x_i = b_i \qquad i = 1, 2, \ldots$$

La méthode initiale ne gère pas l'utilisation de préconditionneurs variables, mais cette limitation a été levée par la suite. Il a même été montré que sous certaines conditions, la méthode FGCRO-DR (pour Flexible GCRO-DR) est équivalente à FGMRES-DR [24], une variante souple d'une autre méthode itérative qui utilise le recyclage d'informations : GMRES with Deflated Restarting [25]. L'un des avantages certain des méthodes basées sur GCRO-DR au lieu de GMRES-DR est que celle-ci permet le recyclage d'informations lors de la résolution de plusieurs systèmes linéaires. L'information n'est pas seulement réutilisée lors des redémarrages des méthodes tronquées, mais aussi lors de l'initialisation des résidus initiaux.

3 Méthodes par blocs

Il est souvent nécessaire de résoudre des systèmes linéaires de la forme suivante :

$$AX = B,$$

où A est la matrice (généralement carrée) du système linéaire, et le second membre B est également une matrice mais cette fois-ci "tall and skinny" (beaucoup plus de lignes que de colonnes). On rencontre ce type de système lors de la simulation de certains phénomènes ondulatoires ou fréquentiels [26, 27], ou en chromodynamique quantique sur réseau [28] avec plusieurs sources, car chacune de ces sources est susceptible de générer une colonne dans la matrice B. Une des premières méthodes itératives, adaptée au cas des systèmes linéaires à seconds membres multiples, fût le Gradient Conjugué [29]. Une étude initiale du GMRES par blocs a été proposée dans la thèse de Vital [30]. Ce type de méthode n'a pas été testé de manière exhaustive pour des applications massivement parallèles. Nous proposons une étude sur un tel cas test qui montre que la complexité opératoire des méthodes de Krylov par blocs est très supérieure aux méthodes traditionnelles. Une alternative que nous proposons pour limiter cette complexité est de travailler sur des sous-blocs de colonnes de la matrice B et d'utiliser les techniques de recyclage proposées dans le paragraphe précédent pour accélérer la convergence de sous-bloc en sous-bloc.

4 Résultats numériques

Certains des essais numériques qui seront présentés oralement lors de la conférence sont repris dans ce paragraphe du résumé étendu.

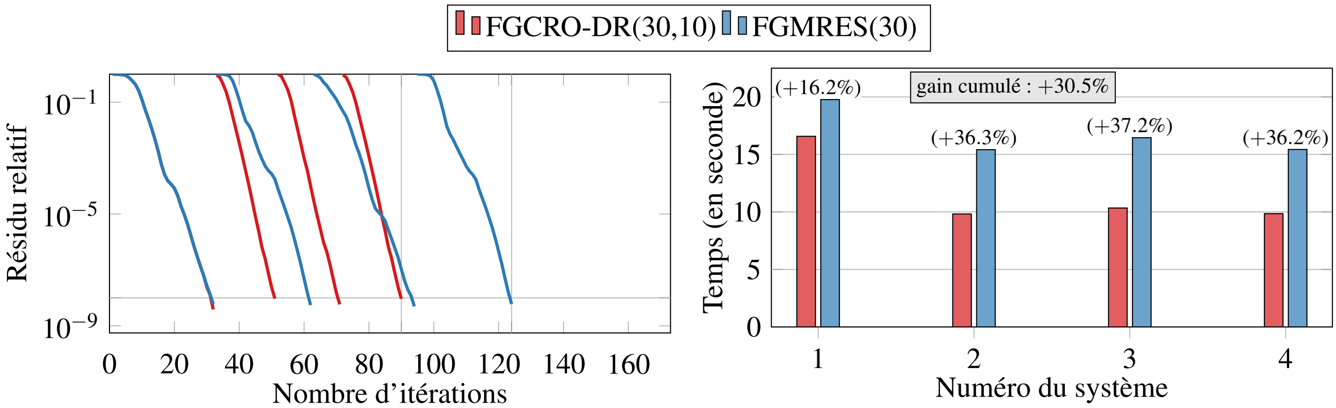
Dans un premier temps, nous résolvons l'équation de Poisson bidimensionnelle avec un schéma différences finies standard à 5 points. Ce cas test provient de l'exemple 32 de la librairie PETSc. Nous l'avons légèrement modifié pour générer une séquence de 4 systèmes linéaires où seul le second membre change. On résout donc :

$$-\Delta u_i = f_i \qquad i = 1, 2, 3, 4,$$

avec

$$f_i(x,y) = \frac{1}{\nu_i} e^{-\frac{(1-x)^2}{\nu_i}} e^{-\frac{(1-y)^2}{\nu_i}},$$

où $\{\nu_i\} = \{0.1, 10, 0.001, 100\}$. Les systèmes sont résolus soit avec FGMRES via PETSc, soit avec FGCRO-DR via HPDDM. Le préconditionneur utilisé est GAMG, avec comme choix de lisseur trois itérations de GMRES. Les cycles multigrilles sont ainsi non-linéaires et il est indispensable d'utiliser une variante souple (qui n'est pas disponible dans Belos). La figure suivante présente les courbes de convergence des deux méthodes, ainsi que les temps de calcul sur 8 192 processeurs (le système linéaire a 280 millions d'inconnues et le préconditionneur est assemblé en deux minutes).



(a) Courbes de convergence—lisseur GMRES(3).

(b) Temps pour atteindre la convergence. Gains relatifs entre parenthèses.

Dans un second temps, nous résolvons l'équation de l'élasticité linéaire tridimensionnelle avec des éléments finis $\mathbb{Q}_1$. Ce cas test provient de l'exemple 56 de la librairie PETSc. Nous l'avons légèrement modifié pour générer une séquence de 4 systèmes linéaires où la matrice et le second membre changent. On résout donc :

$$-\nabla \cdot \boldsymbol{\sigma} = \mathbf{f},$$

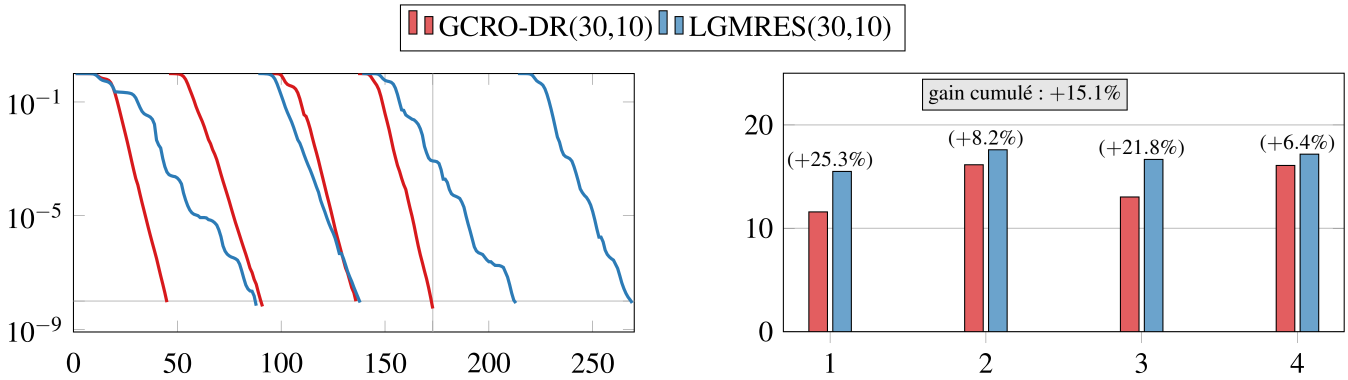
où $\boldsymbol{\sigma}$ est le tenseur des contraintes et $\mathbf{f}$ représente les forces volumiques. Pour obtenir quatre systèmes évoluants, nous utilisons cinq paramètres indexés par $i \in [\![1, \dots, 4]\!]$:

$$\begin{aligned} \{s_i\} &= \{30, 0.1, 20, 10\} & \{r_i\} &= \{0.5, 0.45, 0.4, 0.35\} \\ \{x_i\} &= \{0.5, 0.4, 0.4, 0.4\} & \{y_i\} &= \{0.5, 0.5, 0.4, 0.4\} \\ \{z_i\} &= \{0.5, 0.45, 0.4, 0.35\} \end{aligned}$$

pour définir la paramétrisation d'une petite inclusion sphérique en déplacement :

$$\forall (x,y,z) \in [0;1]^3, (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 < r_i^2.$$

Dans cette inclusion, le module de Young E_i est défini comme $E_i = \frac{E}{s_i}$, où E est le module partout sauf dans l'inclusion. Les systèmes sont résolus soit avec Loose GMRES via PETSc, soit avec GCRO-DR via HPDDM. Nous nous servons à nouveau du préconditionneur multigrille GAMG (en l'appliquant par la droite) avec comme lisseur la méthode de Chebyshev, en utilisant les modes de corps rigide pour définir le quasi-noyau des opérateurs.



(a) Courbes de convergence—lisseur de Chebyshev.

(b) Temps pour atteindre la convergence. Gains relatifs entre parenthèses.

Pour finir, nous proposerons une étude comparative entre certaines méthodes blocs pour la résolution de l'équation de Maxwell tridimensionnelle avec plusieurs seconds membres [31].

Références

[1] Pieter Ghysels, Thomas J. Ashby, Karl Meerbergen, and Wim Vanroose. Hiding Global Communication Latency in the GMRES Algorithm on Massively Parallel Machines. *SIAM Journal on Scientific Computing*, 35(1) :C48–C71, 2013.

[2] Patrick Sanan, Sascha M. Schnepp, and Dave May. Pipelined, Flexible Krylov Subspace Methods. *SIAM Journal on Scientific Computing*, 2016.

[3] Anthony T Chronopoulos. *s*-Step Iterative Methods for (Non)Symmetric (In)Definite Linear Systems. *SIAM Journal on Numerical Analysis*, 28(6) :1776–1789, 1991.

[4] Mark Hoemmen. *Communication-avoiding Krylov subspace methods*. PhD thesis, University of California, Berkeley, 2010.

[5] Paul Soudais. Iterative Solution of a 3-D Scattering Problem from Arbitrary Shaped Multidielectric and Multiconducting Bodies. *IEEE Transactions on Antennas and Propagation*, 42(7) :954–959, 1994.

[6] Allison H. Baker, John M. Dennis, and Elizabeth R. Jessup. On Improving Linear Solver Performance : A Block Variant of GMRES. *SIAM Journal on Scientific Computing*, 27(5) :1608–1626, 2006.

[7] Luc Giraud, Serge Gratton, and Emeric Martin. Incremental spectral preconditioners for sequences of linear systems. *Applied Numerical Mathematics*, 57(11) :1164–1180, 2007.

[8] Pierre Gosselet, Christian Rey, and Julien Pebrel. Total and selective reuse of Krylov subspaces for the resolution of sequences of nonlinear structural problems. *International Journal for Numerical Methods in Engineering*, 94(1) :60–83, 2013.

[9] Valeria Simoncini and Efstratios Gallopoulos. Convergence properties of block GMRES and matrix polynomials. *Linear Algebra and its Applications*, 247 :97–119, 1996.

[10] Julien Langou. *Solving large linear systems with multiple right-hand sides*. PhD thesis, INSA de Toulouse, 2003.

[11] Mickaël Robbé and Miloud Sadkane. Exact and inexact breakdowns in the block GMRES method. *Linear Algebra and its Applications*, 419(1) :265–285, 2006.

[12] Robert Falgout and Ulrike Yang. *hypre* : A library of high performance preconditioners. *Computational Science—ICCS 2002*, pages 632–641, 2002.

[13] Peter Bastian, Felix Heimann, and Sven Marnach. Generic implementation of finite element methods in the distributed and unified numerics environment (DUNE). *Kybernetika*, 46(2) :294–315, 2010.

[14] Dimitar Lukarski and Nico Trost. PARALUTION web page. `http://www.paralution.com`, 2016.

[15] Satish Balay, Shrirang Abhyankar, Mark F. Adams, Jed Brown, Peter Brune, Kris Buschelman, Lisandro Dalcin, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Karl Rupp, Barry F. Smith, Stefano Zampini, and Hong Zhang. PETSc web page. `http://www.mcs.anl.gov/petsc`, 2016.

[16] Allison H. Baker, Elizabeth R. Jessup, and Thomas Manteuffel. A Technique for Accelerating the Convergence of Restarted GMRES. *SIAM Journal on Matrix Analysis and Applications*, 26(4) :962–984, 2005.

[17] Jocelyne Erhel, Kevin Burrage, and Bert Pohl. Restarted GMRES preconditioned by deflation. *Journal of Computational and Applied Mathematics*, 69(2) :303–318, 1996.

[18] Désiré Nuentsa Wakam and François Pacull. Memory efficient hybrid algebraic solvers for linear systems arising from compressible flows. *Computers & Fluids*, 80 :158–167, 2013.

[19] Michael A. Heroux, Roscoe A. Bartlett, Vicki E. Howle, Robert J. Hoekstra, Jonathan J. Hu, Tamara G. Kolda, Richard B. Lehoucq, Kevin R. Long, Roger P. Pawlowski, Eric T. Phipps, et al. An overview of the Trilinos project. *ACM Transactions on Mathematical Software (TOMS)*, 31(3) :397–423, 2005.

[20] Eric Bavier, Mark Hoemmen, Sivasankaran Rajamanickam, and Heidi Thornquist. Amesos2 and Belos : Direct and Iterative Solvers for Large Sparse Linear Systems. *Scientific Programming*, 20(3) :241–255, 2012.

[21] Pierre Jolivet and Frédéric Nataf. Hpddm.

[22] Pierre Jolivet and Pierre-Henri Tournier. Block iterative methods and recycling for improved scalability of linear solvers. In *Proceedings of the 2016 International Conference on High Performance Computing, Networking, Storage and Analysis*, SC16. IEEE, 2016.

[23] Michael L. Parks, Eric de Sturler, Greg Mackey, Duane D. Johnson, and Spandan Maiti. Recycling Krylov Subspaces for Sequences of Linear Systems. *SIAM Journal on Scientific Computing*, 28(5) :1651–1674, 2006.

[24] Luc Giraud, Serge Gratton, Xavier Pinel, and Xavier Vasseur. Flexible GMRES with Deflated Restarting. *SIAM Journal on Scientific Computing*, 32(4) :1858–1878, 2010.

[25] Ronald B Morgan. GMRES with Deflated Restarting. *SIAM Journal on Scientific Computing*, 24(1) :20–37, 2002.

[26] Ru Yu, Eric de Sturler, and Duane D. Johnson. A Block Iterative Solver for Complex Non-Hermitian Systems Applied to Large-Scale Electronic-Structure Calculations. Technical Report UIUCDCS-R-2002-2299, University of Illinois at Urbana-Champaign, Department of Computer Science, 2002.

[27] Henri Calandra, Serge Gratton, Julien Langou, Xavier Pinel, and Xavier Vasseur. Flexible Variants of Block Restarted GMRES Methods with Application to Geophysics. *SIAM Journal on Scientific Computing*, 34(2) :A714–A736, 2012.

[28] Tetsuya Sakurai, Hiroto Tadano, and Yoshinobu Kuramashi. Application of block Krylov subspace algorithms to the Wilson–Dirac equation with multiple right-hand sides in lattice QCD. *Computer Physics Communications*, 181(1) :113–117, 2010.

[29] Dianne P. O'Leary. The Block Conjugate Gradient Algorithm and Related Methods. *Linear Algebra and its Applications*, 29 :293–322, 1980.

[30] Brigitte Vital. *Étude de quelques méthodes de résolution de problemes linéaires de grande taille sur multiprocesseur*. PhD thesis, Université Rennes 1, 1990.

[31] Pierre-Henri Tournier, Iannis Aliferis, Marcella Bonazzoli, Maya de Buhan, Marion Darbas, Victorita Dolean, Frédéric Hecht, Pierre Jolivet, Ibtissam El Kanfoud, Claire Migliaccio, Frédéric Nataf, Christian Pichot, and Serguei Semenov. Microwave Tomographic Imaging of Cerebrovascular Accidents by Using High-Performance Computing. 2016.