

RTRA Digiteo
Recherche en Sciences et
Technologie de l'Information
et de la Communication

Appel à projets 2007 pour le DIM VI
Logiciels et Systèmes Complexes

Proposition d'un projet de recherche
16 mai 2007

Demande d'allocation post doctorale (18 mois) sur le thème :

**Calcul parallèle pour la propagation et la quantification
d'incertitudes dans les modèles**

Laboratoire partenaire n° 1 :

CEA-DEN DM2S/SFME

Centre d'Etudes de Saclay

Département de Modélisation des Systèmes et Structures

91191 Gif sur Yvette Cedex

Laboratoire partenaire n° 2 :

LIMSI

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

CNRS, UPR 325

Université Paris Sud

91403 Orsay

Résumé

1. TITRE DU PROJET : Calcul parallèle pour la propagation et la quantification d'incertitudes dans les modèles

2. Nature du projet : Post Doc (18 mois)

3. Coordonnateur responsable du projet :

Nom : Jean-Marc MARTINEZ

Fonction : Expert Sénior CEA

Institution / Laboratoire : CEA DM2S/SFME

4. Partenaires Digiteo du projet

Laboratoire partenaire n° 1 :

Centre d'Etudes de Saclay

Département de Modélisation des Systèmes et Structures

Service des Fluides numériques, Modélisation et Etudes

91191 Gif sur Yvette Cedex

Laboratoire partenaire n° 2 :

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

CNRS, UPR 325

Université Paris Sud

91403 Orsay

5. Objectifs du projet (résumé) :

La simulation numérique est aujourd'hui de plus en plus employée pour l'étude et l'analyse de phénomènes physiques et de systèmes complexes, tant dans le cadre académique qu'industriel. La généralisation de l'approche numérique repose d'une part sur les développements constants des capacités de calcul grâce à l'émergence des plateformes de parallèles (clusters), et d'autre part sur les progrès de la modélisation (couplages multi-physique, approches multi-échelle,...) et des algorithmes de calcul.

La prise en compte des incertitudes de nature épistémique (modèles) ou stochastique (phénomènes naturels) est devenue un besoin très important compte tenu de l'exploitation des résultats de simulation (analyse de risque, sûreté, conception, ...). L'état de l'art montre l'intérêt croissant des méthodes récemment proposées pour la modélisation des incertitudes, connues sous le nom générique de Polynômes de Chaos. Deux méthodes d'utilisation sont possibles. La première dite non-intrusive, basée sur une exploitation du modèle numérique comme boîte noire est similaire aux méthodes de simulation par tirages Monte Carlo ; suffisamment mature, elle a été proposée dans des projets ANR de réalisation d'outils logiciels (OPUS). La seconde approche est intrusive car elle nécessite la réécriture de la résolution numérique du modèle. C'est cette approche que nous nous proposons d'étudier, plus complexe à mettre en œuvre mais beaucoup plus précise. Elle peut aujourd'hui être envisagée grâce notamment aux progrès constants sur les moyens de calculs haute performance aux architectures massivement parallèle comme ceux disponibles au CEA.

Ce travail permettra de montrer l'adéquation entre ces architectures et les stratégies de parallélisation proposées pour la résolution de problèmes stochastiques. Un autre résultat attendu est la constitution d'une bibliothèque générique d'outils et d'opérateurs numériques sur les polynômes de chaos. Cette bibliothèque pourra servir de base dans les futurs développements et applications des polynômes de chaos.

Fiche

1. TITRE DU PROJET : Calcul parallèle pour la propagation et la quantification d'incertitudes dans les modèles

2. Nature du projet : Post Doc

3. Thématique du projet : Incertitudes et Approximation

4. Coordonnateur responsable du projet :

Nom : Jean-Marc MARTINEZ

Fonction : Expert Sénior CEA

Institution / Laboratoire : CEA DM2S/SFME

Adresse postale : Centre d'Etudes de Saclay 91190 Gif sur Yvette

Téléphone : (33) 01 69 08 58 21

Fax : (33) 01 69 08 96 96

Email : jean-marc.martinez@cea.fr

MARTINEZ Jean-Marc (CEA-DM2S)

Jean-Marc Martinez a soutenu sa thèse d'Etat en 1983 (Paris 11) dans le domaine du contrôle/commande appliqué au pilotage des Réacteurs à Eau Pressurisée. Son travail s'est concrétisé par le dépôt d'un brevet et un recrutement au CEA Saclay en 1983.

Au CEA, ses principales fonctions portent sur le développement de méthodes et outils dédiés à la simulation. Dans un premier temps, il les a développés en robotique (intelligence artificielle) puis en supervision, planification et actuellement dans le développement de méthodes d'aide à la simulation numérique, notamment les outils d'aide à la décision basés sur des représentations simplifiées des modèles numériques et ceux dédiés aux traitements des incertitudes.

Ses travaux de recherche portent sur l'apprentissage statistique, la planification d'expériences numériques et la modélisation des incertitudes. Il a été responsable/directeur de plusieurs thèses menées au CEA et a codirigé une thèse avec G. Dreyfus de l'ESPCI en planification d'expériences numériques pour méta-modèles neuronaux. Dans le domaine du traitement des incertitudes, ses travaux actuels, en collaboration O. Le Maître (LIMSI), portent sur les méthodes spectrales à base de polynômes de chaos.

Bibliographie:

- Polynomial chaos expansions for uncertainties quantification and sensitivity analysis, Th. Crestaux, J.-M. Martinez, O. Le Maître, 5th International Conference on Sensitivity Analysis of Model Output, SAMO 2007, Budapest, Hungary, 18-22 June 2007
- Neural Networks - Methodology and Applications, G. Dreyfus, co-auteur, Springer, 2005, ISBN 3-540-22980-9
- Réseaux de neurones - Méthodologie et applications, G. Dreyfus, J.M. Martinez, M. Samuelides, M.B. Gordon, F. Badran, S. Thiria, L. Hérault, Eyrolles, 2004, ISBN 2-212-11019-7

- Plans d'expériences itératifs pour la construction de modèles non linéaires, S. Gazut, J.-M. Martinez, S. Issanchou, Société Française de Statistique, 38ème Journées de Statistique, Clamart, juin 2006

5. Avis motivé du Chef de Service auquel appartient le coordonnateur du projet :

Daniel CARUGE

CEA Saclay

Chef du Service Fluides numériques, Modélisation et Etudes

Centre d'Etudes de Saclay

91190 Gif sur Yvette

Tél. : 01 69 08 91 12

Email : Daniel.Caruge@cea.fr

Ce projet s'inscrit dans la thématique du CEA DEN sur la quantification et la propagation des incertitudes en simulation numérique et dans celle d'*Incerteo* de Digiteo, thème 7 de l'appel à projets. Ce travail permettra de débiter une collaboration entre le LIMSI et le CEA-DEN pour développer les méthodes de modélisation d'incertitudes basées sur les polynômes de chaos. Ce projet en s'appuyant sur les moyens de calcul parallèle disponibles au CEA, permettra de valider des stratégies de résolution parallèle de problèmes stochastiques et de constituer une bibliothèque de composants génériques pour les futurs développements et applications des polynômes de chaos dans le traitement des incertitudes en simulation numérique.

6. Partenaire Digiteo du projet

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

CNRS, UPR 325

Université Paris Sud

91403 Orsay

LE MAITRE Olivier (LIMSI)

Maître de Conférences

LIMSI CNRS

B.P. 133

91403 ORSAY CEDEX

olm@iup.univ-evry.fr

Activités professionnelles :

1998-Présent : Enseignant-chercheur, Maître de Conférences, Université d'Evry Val d'Essonne.

2002-Présent : Chercheur associé au LIMSI.

2004-2005 : Consultant auprès de Sandia National Labs (USA).

Formation Universitaire :

2006 : Habilitation à Diriger les Recherches, Université d'Evry Val d'Essonne.

1998 : Doctorat en Mécanique, Université du Havre.

Principaux thèmes de recherche :

Propagation et quantification d'incertitudes dans les modèles numériques.

Mécanique des fluides numérique (écoulements incompressibles, réactifs, convection naturelle).

Sélection de publications récentes :

- L.Mathelin and O. Le Maître, Dual Based a posteriori error estimation for stochastic finite element methods, soumitted to CAMCS, (2006).

- O. Le Maître and O. Knio, A stochastic particle mesh method for vortical flow, *J. of Comput. Physics*, accepted (2007).
- O. Le Maître, H. Najm, P. Pébay, R. Ghanem and O. Knio, multi-resolution analysis for uncertainty quantification in chemical systems, *SIAM Journal of Scientific Computing*, in press (2007).
- O. Knio and O. Le Maître, Uncertainty propagation in CFD using Polynomial Chaos decompositions, *Fluid Dynamics Research*, 38(9), pp.616-640, (2006).

7. Description du projet :

7.1 Contexte

La simulation numérique est aujourd'hui de plus en plus employée pour l'étude et l'analyse de phénomènes physiques et de systèmes complexes, tant dans le cadre académique qu'industriel. La généralisation de l'approche numérique pour l'investigation de problèmes de très grande dimension repose d'une part sur les développements constants des capacités de calcul, notamment ces dernières années grâce à l'émergence des plateformes parallèles (clusters), et d'autre part sur les progrès de la modélisation (couplages multi-physique, approches multi-échelle, ...) et des algorithmes de calcul.

Avec la possibilité de réaliser des simulations pour une résolution (spatiale et temporelle) élevée et pour une modélisation physique de plus en plus complète, est apparu de nouvelles problématiques et besoins importants et critiques relatifs à l'analyse et à l'interprétation des résultats de calculs. En particulier, l'utilisation de modèles complexes dans le cadre de simulations de haute précision nécessite une définition précise des données d'entrée du calcul, telles que les propriétés physiques du système étudié, les conditions aux limites et les forçages extérieurs qui lui sont appliqués, les constantes de modélisation (paramètres libres ou de calibration), ... La validité de ces informations est cruciale pour une simulation de haute précision car les erreurs sur les données peuvent devenir dominantes et compromettre la validité des résultats obtenus. Or, pour de nombreux systèmes physiques, cette information ne peut être connue exactement a priori, soit du fait de lacunes de connaissance (par exemple dans le cas de constantes de modélisation non mesurables) soit du fait d'une variabilité inhérente (par exemple quand le système simulé est une machine industrielle). Pour ces situations, une quantification de l'impact des incertitudes des données d'entrée sur le résultat de la simulation est nécessaire en vue notamment d'estimer leur validité.

7.2 Propagation et quantification d'incertitudes

Plusieurs techniques, de complexité et de qualité variable, ont été proposées pour estimer l'impact d'incertitudes sur les paramètres d'une simulation numérique. Nous nous intéressons dans ce projet aux techniques probabilistes de projection orthogonale stochastique qui consistent à considérer les paramètres incertains de la simulation comme des variables aléatoires, puis à approcher la dépendance de la solution numérique vis à vis des paramètres sur une base de dimension finie de fonctionnelles orthogonales (décorrélées) en les variables aléatoires. Connues sous le nom générique de méthodes d'approximation par Polynômes de Chaos, ces techniques de projection stochastique permettent une analyse très précise de l'impact des incertitudes. En fait, la connaissance des coefficients de projection caractérise entièrement la solution incertaine et permet l'estimation de sa densité de probabilité et l'influence respective des différentes sources d'incertitudes. Cependant, le coût numérique de ces techniques augmente très rapidement avec le nombre de paramètres incertains considérés; elles sont ainsi limitées à des problèmes faisant intervenir au plus quelques dizaines de paramètres incertains.

Deux approches alternatives sont possibles pour la détermination des coefficients de projection stochastique. Les techniques de **projection non-intrusive** utilisent un échantillon de simulations, correspondant à différentes réalisations des paramètres incertains, pour estimer les coefficients de projection quadrature numérique ou régression. L'intérêt des

approches non-intrusives repose sur l'emploi d'un code de calcul déterministe (approche de type boîte noire) et sur l'indépendance des simulations pour différentes réalisations des paramètres ce qui autorise une résolution parallèle immédiate.

Les techniques de **projection intrusive** reposent au contraire sur la formulation d'un problème dit spectral et la dérivation d'équations pour les coefficients de projection de la solution, aussi appelés modes spectraux, par des techniques de type Galerkin. Le problème spectral couple en général tous les modes stochastiques de la solution incertaine ce qui conduit à la résolution d'un problème de très grande dimension (P fois celle du problème déterministe, ou P est la dimension de la base stochastique). Sa résolution nécessite en outre l'adaptation des outils de simulation et requière donc un travail préliminaire plus important que pour les techniques non-intrusives, notamment pour les modèles incluant des non-linéarités importantes. Les projections non-intrusives présentent cependant un attrait important, en raison particulièrement de la puissance des théories fonctionnelles et spectrales sous-jacentes et en vue de la dérivation de techniques auto-adaptatives pour la construction de bases stochastiques optimales.

7.3 Objectifs

Dans ce projet, nous nous proposons de mettre en œuvre une technique de projection intrusive pour la propagation d'incertitudes dans un modèle d'écoulement in stationnaire de complexité modérée (équations de Navier-Stokes incompressible) mais nécessitant une résolution (nombre de degrés de liberté) élevée. Les incertitudes considérées porteront sur les propriétés du fluide et les conditions aux limites.

Comme nous l'avons mentionné ci-dessus, cette approche conduit à la résolution d'un problème spectral de très grande dimension, jusqu'à quelques centaines de fois celle du problème déterministe, dont la résolution devra être parallélisée. Différents niveaux de parallélisation devront être considérés.

Tant un premier temps, l'utilisation de schéma d'intégration temporelle explicitant les termes non-linéaires permet le découplage, à chaque pas de temps, de la résolution des différents modes stochastiques de la solution. Cette propriété autorise un parallélisme à gros grain où chaque processeur résout indépendamment un ou plusieurs modes (selon le nombre de processeurs disponibles). L'évaluation des termes sources non-linéaires explicites, constitue alors le goulot d'étranglement de la simulation puisqu'elle couple l'ensemble des modes spectraux. Ici, une stratégie de parallélisation devra être définie pour optimiser le calcul des termes sources non-linéaires en minimisant les échanges de données entre processeurs et pour équilibrer la charge de calcul. Cette stratégie devra reposer sur une analyse fine du type de non-linéarité considérée, avec en premier lieu une compréhension approfondie de la structure de l'opérateur de multiplication spectrale.

Dans un second temps, un niveau de parallélisme plus fin pourra être considéré en incluant les techniques de résolution parallèle déterministes telles que les approches de décomposition de domaine (en espace). Les stratégies de communication entre sous-domaines devront être adaptées pour tenir compte du caractère stochastique de la solution. Enfin, dans le cas de l'emploi de techniques d'adaptation automatique du maillage à la solution (Automatic Mesh Refinement), l'adaptation du maillage physique à la structure de chaque mode spectral devra être considérée. Ce dernier point nécessitera la dérivation de nouveaux opérateurs d'interpolation et de prolongation pour le calcul des termes sources non-linéaires, les modes spectraux ayant alors sur discrétisation spatiale variable.

7.4 Résultats attendus

Les développements proposés ci-dessus auront pour objectif principal la démonstration de la faisabilité et de l'efficacité des techniques de projection intrusive pour la propagation d'incertitudes paramétriques dans les simulations de grande dimension. L'utilisation des plateformes de calcul massivement parallèles, disponibles au CEA, permettra de montrer l'adéquation entre ces architectures et les stratégies de parallélisation proposées pour la résolution du problème incertain. La possibilité d'effectuer des tests numériques et des comparaisons entre différentes alternatives d'implémentation, tout en utilisant un nombre

élevé de processeurs, permettra d'analyser finement les facteurs d'efficacité et le rendement des algorithmes parallèles.

Un autre résultat attendu de ce projet est la constitution d'une bibliothèque générique d'outils numériques parallèles, dédiés aux calculs de propagation d'incertitudes par projection non-intrusive. Cette bibliothèque parallèle devra inclure d'une part les principaux opérateurs spectraux couramment utilisés dans les applications (multiplication, exponentiation, extraction de racine carrée,...) et d'autre part standardiser les protocoles d'échange de données spectrales entre processeurs. Elle pourra servir de base dans le futur à l'application des techniques intrusives à d'autres applications.

L'ensemble de ces résultats donnera lieu à la rédaction de rapports de recherche et d'avancement des travaux et par la publication d'articles dans des revues scientifiques du domaine (calcul scientifique et intensif) et communication lors de congrès.

7.5 Calendrier

Outils de base et parallélisation à gros grain (T0 à T0+6mois).

- Parallélisation de la résolution des modes spectraux. (3 mois)
- Parallélisation du calcul des termes sources non-linéaires et constitution d'une bibliothèque parallèle pour les principales opérations spectrales. (2mois)
- Tests et analyse de performance et rendement de parallélisme. (1 mois)

Parallélisation à grain fin (T0+6mois à T0+18mois).

- Application des résultats de la première année à un calcul utilisant une technique de décomposition de domaine spatial, étude des stratégies d'échanges de données entre domaines. (4 mois)
- Extension de la méthodologie parallèle à un outil de résolution utilisant une technique d'adaptation de maillage automatique (AMR). (3 mois)
- Adaptation de maillage à chacun des modes spectraux: définition et implémentation des opérateurs d'interpolation et de prolongation spectraux. (3mois)
- Démonstration de performance sur un problème d'écoulement complexe de grande dimension. (2 mois)

7.6 Références

- O.P. Le Maître, M.T. Reagan, B.J. Debusschere, H.N. Najm, R.G. Ghanem and O.M. Knio, **Natural Convection in a Closed Cavity under Stochastic, Non-Boussinesq Conditions**, *SIAM Journal of Scientific Computing*, **26**(2), pp. 375-394, (2004).
- Olivier P. Le Maître, Matthew T. Reagan, Habib N. Najm, Roger G. Ghanem and Omar M. Knio, **A Stochastic Projection Method for Fluid Flow: II. Random Process**, *Journal of Computational Physics*, **181**(1),pp. 9-44, (2002).
- Olivier P. Le Maître, Omar M. Knio, Habib N. Najm and Roger G. Ghanem, **A Stochastic Projection Method for Fluid Flow: I. Basic Formulation**, *Journal of Computational Physics*, **173**(2), pp. 481-511, (2001).

9. Coordonnées (postales, téléphoniques et électroniques) de deux référents n'ayant jamais travaillé avec les proposants mais compétents pour évaluer le projet.

Pierre Sagaut IJLRA/UPMC, Boîte 162 4, Place Jussieu F-75252 Paris Cedex 5 Tél: (33) 1 44 27 54 68 pierre.sagaut@upmc.fr	Olivier Pironneau Laboratoire Jacques-Louis Lions Université Pierre et Marie Curie Boîte 187 175, rue du Chevaleret 75013 Paris
--	---