
Dossier de renouvellement du GDR Calcul

Pour la période 2022-2026

GDR CALCUL

Janvier 2021

Table des matières

1	Contexte du GDR Calcul	3
2	Bilan des actions menées depuis 2017	4
2.1	Actions initiées par le bureau	4
2.1.1	Mise en place d'outils de communication interne	4
2.1.2	Évolution du site https://calcul.math.cnrs.fr	5
2.1.3	Ressources de calcul pour les mathématiques	6
2.2	Organisation d'évènements scientifiques	6
2.2.1	Action Nationale de Formation "Boîtes à outils éléments finis open source"	6
2.2.2	École thématique PRECIS	7
2.2.3	Journée "Python et Data Science"	7
2.2.4	Mini-symposium au congrès SMAI 2017	9
2.2.5	École Thématique GEOMDATA	9
2.2.6	Mini-symposium sur Jupyter au CANUM 2018	10
2.2.7	JCAD, Journées Calcul & Données	10
2.2.8	JoSy IA 2018	11
2.2.9	Journée différentiation algorithmique	11
2.2.10	Journées "Calcul et Apprentissage"	12
2.2.11	Congrès SMAI 2019 : mini-symposium "S'affranchir du maillage : résoudre les EDP par le deep learning"	13
2.2.12	Journée "Julia et Optimisation"	14
2.2.13	Journée histoire du calcul	15
2.2.14	Autres contributions	16
2.2.15	Bilan des objectifs du projet 2017-2021	16
3	Projet 2022-2026	17
3.1	Précision numérique	18
3.2	Calcul quantique	19
3.3	Modèles stochastiques	21
3.4	Apprentissage machine et interactions	23
3.5	Science ouverte	24
3.6	Mission parité	25
3.6.1	Parité oratrices/orateurs	25
3.6.2	Actions à destination des filles et des jeunes femmes pour susciter des vocations	25
3.7	Médiation scientifique	26

3.8	Jeunes chercheurs	27
3.8.1	Offres d'emploi	27
3.8.2	Participation aux interventions	27
3.8.3	Collection de notebooks Jupyter sur la plateforme PLMSHift-GPU de Mathrice	27
3.9	Organisation interne	28
3.9.1	Porteur du projet	28
3.9.2	Bureau	28
3.9.3	Conseil scientifique	28
3.10	Partenariats	29

1 Contexte du GDR Calcul

Le groupe Calcul a été créé en 2003 au sein du département SPM du CNRS à l'intérieur du groupe Mathrice (GdS 2754 du CNRS regroupant les administrateurs système et réseau des laboratoires de mathématiques). Depuis 2009, le groupe Calcul est structuré en deux entités reconnues par le CNRS

- le GDR Calcul (INSMI, CNRS),
- le Réseau métier Calcul (MITI, CNRS).

Ce découpage a été mûrement réfléchi et correspond à deux aspects bien distincts des domaines du calcul scientifique que sont les aspects recherches et les aspects technologiques. Ils ne sont néanmoins pas disjoints et se recoupent pour un certain nombre de thématiques. Le GDR Calcul s'intéresse aux nouvelles méthodes numériques et aux nouveaux algorithmes qui sont à la pointe de la recherche en calcul scientifique et intensif. Ces thématiques ont une forte composante en mathématiques appliquées et en informatique mais sont également pluridisciplinaires (physique, chimie, biologie, etc.). En règle générale, la validation de ces nouvelles méthodes passe par l'élaboration de logiciels et bibliothèques de calcul qui sont développés au sein des laboratoires et sont mis à disposition de la communauté nationale et internationale. Par ses actions de communication et de formation, le GDR Calcul souhaite participer à la diffusion de ces outils et à leur adoption par la communauté. Il peut également proposer son expertise sur la parallélisation et l'optimisation des algorithmes en s'appuyant sur le Réseau Calcul. Enfin, par son action de veille, le GDR suit les mouvements récents, notamment la percée de l'apprentissage machine et du calcul quantique. Le paysage du GDR et du Réseau Calcul est résumé sur la figure 1.

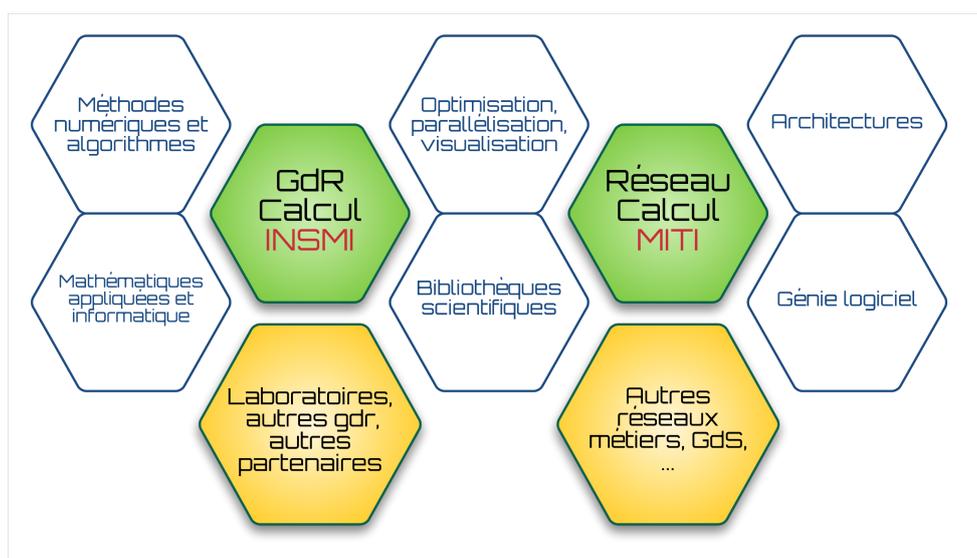


Figure 1: Structuration du groupe Calcul

Depuis la création du GdR Calcul en 2009, de nombreuses actions ont vu le jour permettant de constituer puis de renforcer la communauté autour du calcul en France. Le GdR Calcul offre des formations sous forme d'écoles thématiques et des journées scientifiques. Il anime également le site <https://calcul.math.cnrs.fr> permettant à la fois d'annoncer ses actions et d'archiver les supports numériques associés. Le site <https://calcul.math.cnrs.fr> permet également de publier des offres d'emploi (ingénieurs, doctorats, postdoctorats, etc.) dans le domaine du calcul. Enfin, le groupe Calcul anime une liste de discussions de plus de 2200 abonnés.

La section suivante présente le bilan des actions menées par le GDR sur la période courante (2017-2021).

2 Bilan des actions menées depuis 2017

2.1 Actions initiées par le bureau

2.1.1 Mise en place d'outils de communication interne

Afin de structurer et fluidifier la communication interne, le groupe calcul a adopté le principe d'un groupe Gitlab pour héberger l'ensemble de ses projets et des discussions associées. Initié fin 2016, ce [groupe Gitlab](#) héberge un wiki pour la communication interne et plus d'une vingtaine de projets (privés) contenant nos supports de présentation, nos rapports, le code de nos outils partagés et les sources de notre site web. Ce mode de fonctionnement nous permet de structurer nos échanges distants et de programmer nos actions tout en préservant l'historique des discussions

et des versions dans un espace commun, de sorte que chaque membre du bureau a accès à la même information.

2.1.2 Évolution du site <https://calcul.math.cnrs.fr>

En 2018, nous avons entrepris la refonte du site hébergeant les activités du GDR Calcul et du réseau Calcul, alors basé sur le système de gestion de contenu Spip.

Nous avons fait le choix d'un site statique, généré par un outil libre (Pelican) et un ensemble de scripts et programmes dans lesquels l'utilisation de bases de données est proscrite. Le contenu des pages est entré sous la forme de fichiers texte au format RST qui sont stockés et versionnés au sein d'un projet Gitlab . Dès qu'une modification est effectuée, le site est reconstruit et redéployé automatiquement sur la plateforme PLMShift de Mathrice. Le site est interfacé avec le portail Indico de Mathrice que nous utilisons pour gérer nos manifestations. Le site relaie également les informations du compte Twitter de Calcul et propose un formulaire permettant de déposer une annonce d'emploi. Après validation, une nouvelle version du site incluant cette annonce est construite et une notification est envoyée aux abonnés de la liste Calcul.

L'ensemble des scripts et configurations est publié en *open source* sur <https://gitlab.math.unistra.fr/groupe-calcul/website>. Les sources permettant de migrer les données du kit SPIP CNRS vers un système de scripts sous Pelican sont à la disposition de tous sur <https://gitlab.math.unistra.fr/groupe-calcul/spip2pelican>.

Depuis sa mise en production en juin 2019, ce site est particulièrement actif : plus de 200 modifications pour l'année 2020. La rubrique de publications des offres d'emploi est particulièrement sollicitée : sur cette même année 2020, nous avons publié 145 offres réparties de la façon suivante.

Type de poste	Nombre d'offres
CDD	37
CDI	23
Concours	11
Postdoctorat	20
Stage	41
Thèse	17
Total	149

L'ensemble des manifestations organisées par le GDR sont répertoriées sur <https://calcul.math>.

cnrs.fr/tag/evt_sci.html.

2.1.3 Ressources de calcul pour les mathématiques

En 2019, nous avons fait l'acquisition d'un nœud généraliste puis en 2020 de deux nœuds de calcul équipés d'accélérateurs GPU de dernière génération en collaboration avec la PLM : le GDR Calcul contribue à un projet piloté et hébergé par Mathrice, financé par l'INSMI à partir des remontées de crédits 2020 et destiné à l'ensemble de la communauté mathématique. Les membres du bureau du groupe Calcul travaillent actuellement avec les équipes de la PLM pour le déploiement des nœuds sous la forme d'un portail JupyterHub au sein de la plateforme <https://plmshift.math.cnrs.fr>. Nous préparons également la rédaction d'exemples d'utilisation et de démonstrations permettant de communiquer et mettre en valeur cet équipement.

L'objectif est d'offrir à la communauté une plateforme pour expérimenter le calcul sur GPU et permettre l'exécution de codes produits dans un contexte de science reproductible. Un cas typique est l'exécution de notebooks Jupyter dans un contexte logiciel rigoureusement identique (au bit près) à celui des cas-tests de la publication.

2.2 Organisation d'évènements scientifiques

Sous l'impulsion de son conseil scientifique et grâce à ses ressources financières, le GDR a organisé ou coorganisé une série de formations, écoles et rencontres thématiques qui sont détaillées ci-dessous.

2.2.1 Action Nationale de Formation "Boîtes à outils éléments finis open source"

Nombre de participants : 31

La méthode des éléments finis est utilisée dans de nombreux domaines de la physique et de l'ingénierie (mécanique des solides, des fluides, bio-physique, électronique, ...).

L'objectif de cette ANF était de présenter plusieurs outils récents, qui sont à la fois innovants dans les méthodes, et largement utilisés (FreeFem++, Feel++, Firedrake) en explicitant chaque fois leurs avantages et leurs inconvénients. Ils peuvent effectivement être complémentaires l'un de l'autre suivant le problème étudié. Chacun de ces outils a son propre langage de programmation. Les participants ont appris à utiliser correctement chacun d'eux en s'appropriant leur langage.

Ces outils proposent tous des solutions différentes pour s'affranchir des difficultés de programmation de la méthode des éléments finis, et permettent aux chercheurs, aux doctorants et aux ingénieurs de résoudre des problèmes physiques de plus en plus complexes.

La formation a eu lieu du 24 au 26 janvier 2017 à Jussieu (Paris). Le lien sur le site groupe Calcul est <https://calcul.math.cnrs.fr/2017-01-ecole-elements-finis.html>.

Programme

- Simulations numériques parallèles et multiphysiques en FreeFem++, F. Hecht et F. Nataf (LJLL, UPMC, Paris)
- Solving PDEs with Feel++, C. Prud'homme (IRMA, Strasbourg)
- Solving PDEs with Firedrake, L. Mitchell (Imperial College, Londres)

Les statuts des 31 participants étaient également répartis entre enseignant-chercheur, ingénieur et doctorant. L'organisation de cette formation a permis de promouvoir des logiciels développés dans les laboratoires de l'INSMI : FreeFem++ est une production du LJLL et Feel++ de l'IRMA.

2.2.2 École thématique PRECIS

Nombre de participants : 30

La reproductibilité des résultats est un des fondements de la recherche scientifique. Au cours des dernières années, ce sujet a pris une place centrale dans des débats, parce qu'un nombre important d'études publiées récemment s'est avéré non-reproductible. Les raisons sont multiples et varient fortement entre les disciplines, mais l'une d'elles est transdisciplinaire : il s'agit de la difficulté de reproduire les résultats d'un calcul numérique. Au vu de l'importance croissante des simulations numériques en physique, chimie, et biologie, mais aussi dans des applications industrielles (aéronautique, automobile, nucléaire, etc.) et médicales, il est impératif de travailler sur l'amélioration de la fiabilité des logiciels scientifiques.

Quel crédit numérique peut-on accorder à un logiciel ? Quelle est la précision des résultats obtenus ? Ces résultats sont-ils reproductibles (dans le temps et dans d'autres configurations) ?

Lorsque la reproductibilité n'est pas garantie, la validation, la vérification des logiciels, le processus de développement doivent être abordés différemment. On doit être en mesure d'estimer la précision des résultats numériques d'un logiciel et mettre en place des solutions pour contenir les sources d'imprécision.

Cette école a eu lieu au CAES de Fréjus du 15 au 19 mai 2017 et a réuni une trentaine de personnes : <https://precis.sciencesconf.org>.

2.2.3 Journée "Python et Data Science"

Nombre de participants : 35

Nous avons remarqué que la majorité des sujets abordés dans les grandes conférences Python (Scipy, Pycon) concernaient les sciences des données en réduisant la part consacrée au calcul scientifique. L'objectif de cet évènement était de faire un état des lieux des bibliothèques disponibles en Python pour les statistiques en général sachant que le langage R est plus couramment utilisé dans ce domaine.

Le groupe Calcul, avec le soutien de l'institut de recherche mathématique de Rennes, a organisé cette journée consacrée au langage Python, à l'écosystème disponible pour tous les "Data scientists" et à des exemples concrets d'utilisation.

La journée a eu lieu le 19 décembre 2017 dans les locaux de l'IRMAR à Rennes.

Programme

- Maude Pupin (Université Lille 1) : *L codent, L créent ou comment démystifier la programmation auprès des jeunes filles ?*
- Francis Wolinski (Société Yotta Conseil) : *Présentation et illustration de l'écosystème Python pour la data science*
- Vincent Vigon (Université de Strasbourg) : *TensorFlow et les réseaux de neurones*
- Pierre Poulain (Université Paris 7) : *Lutter contre le paludisme avec Python*
- Lynn Cherny (EMLYON Business School) : *Job Offer Skills Analysis*
- Gaël Pegliasco (Makina Corpus) : *Présentation des différents concepts du parallélisme et du paysage des librairies de calcul distribué/parallèle en Python.*

Maude Pupin a fait une présentation centrée sur l'apprentissage du langage. Elle a montré les actions entreprises à l'université de Lille pour encourager les vocations des jeunes filles vers la programmation. Francis Wolinski a fait un état des lieux des bibliothèques Python disponibles pour traiter et analyser des données avec des exemples concrets et ludiques comme l'évolution des prénoms dans la société française. Gaël Pegliasco a présenté les outils pour optimiser et paralléliser les codes écrits en Python. Pierre Poulain a présenté son travail sur la conception d'un outil d'analyse des images médicales pour aider à combattre le paludisme. Dans cette présentation on a pu voir la polyvalence du langage pour développer le traitement statistique et l'interface web qui permet aux techniciens médicaux d'utiliser facilement l'outil dans les centres d'analyse en Afrique. Vincent Vigon a présenté les réseaux de neurones avec les outils Python liés qui sont les plus performants actuellement. Enfin, Lynn Cherny a fait une démonstration de son outil d'analyse d'offres d'emploi qui a beaucoup intéressé les participants issus du secteur privé.

Les supports des présentations ont été mis en ligne et sont accessibles sur le site du Groupe Calcul : <https://calcul.math.cnrs.fr/2017-12-journee-python-data.html>.

2.2.4 Mini-symposium au congrès SMAI 2017

Nombre de participants : 20

L'évolution actuelle et prévisible des architectures de machines est en train d'imposer une révision de nombreuses méthodes numériques. En particulier les notions de parallélisation, de localité des données et de vectorisation deviennent essentielles pour tirer le maximum de performance des calculateurs. Les laboratoires d'informatique et de mathématiques développent des techniques pour le calcul hautes performances qui interviennent à des niveaux différents, sans forcément beaucoup d'interaction mutuelle. L'objectif de ce mini-symposium était d'aborder des thématiques à l'interface des deux disciplines et de présenter des points de vue complémentaires sur l'efficacité calculatoire des méthodes numériques et les optimisations automatiques proposées par la compilation de *stencils* et le *tiling*.

Programme

- Alain Darte (LIP, ENS Lyon), Thierry Dumont (ICJ, Université Lyon 1), Repenser les schémas numériques à l'aune de la localité et de la vectorisation
- Florian De Vuyst (CMLA, ENS Paris-Saclay), Schémas numériques et modèles de performances associés. Concepts et exemples.
- Emmanuel Agullo (INRIA, équipe HiePACS/INRIA Bordeaux-Sud-Ouest), Programmation et passage à l'échelle de bibliothèques mathématiques numériques à base de tâches sur moteur d'exécution.
- Philippe Helluy (IRMA, Université de Strasbourg), Palindromic discontinuous Galerkin method.

2.2.5 École Thématique GEOMDATA

Nombre de participants : 36

De nombreux domaines appliqués reposent sur l'analyse de données géométriques : médecine, neurosciences, sismique, météorologie, vision par ordinateur, apprentissage statistique. Cette variété d'applications se retrouve dans la forme, la qualité et la sémantique des données ainsi que dans la nature des problèmes mathématiques qu'elles posent. Les données géométriques peuvent être de plusieurs formes : images formées de pixels ou voxels (en 2D/3D), nuages de points, ou complexes simpliciaux. Elles peuvent être bruitées ou incomplètes. L'information sémantique contenue dans ce type de données dépend également du domaine : la donnée peut représenter une distribution de masse ou d'énergie (distribution des nuages en météorologie ou d'énergie cinétique en imagerie sismique), contenir une structure topologique intéressante (protéine), ou une information géométrique fine locale ou globale (vaisseaux sanguins). Enfin les problèmes à résoudre sont de natures variées : il peut s'agir de reconstruire des structures, de seg-

menter l'image, de mesurer des différences, d'apparier des données, de débruiter, déconvoluer ou reconstruire des données manquantes, de découper le jeu de données en classes (clustering), etc. L'analyse et le traitement des données géométriques offrent donc un sujet d'étude très riche, faisant l'interface entre des applications et de nombreux domaines des mathématiques appliquées (optimisation, statistiques, EDP, calcul des variations, géométrie, topologie, etc.).

L'objectif de GEOMDATA était de présenter cette variété de questions et d'approches à des collègues et des étudiants en mathématiques appliquées ou informatique qui voudraient en faire usage pour leur propre recherche. Le parti pris était d'insister sur les aspects numériques et la mise en œuvre pratique (en Python), via des bibliothèques comme GUHDI, PyTorch, etc.

Cette école a eu lieu au CAES de Fréjus du 10 au 14 septembre 2018 et a regroupé une trentaine de participants : <https://geomdata.sciencesconf.org>.

2.2.6 Mini-symposium sur Jupyter au CANUM 2018

Nombre de participants : 40

Jupyter et son écosystème offrent un ensemble d'outils permettant d'avoir un environnement de travail convivial aussi bien pour la recherche que pour l'enseignement. Destiné à la base pour développer du code Python, Julia ou R, c'est aujourd'hui bien plus que ça. Les notebooks (page web où l'on peut écrire et exécuter du code numérique, du texte, des équations mathématiques, insérer des vidéos, des graphiques interactifs, ...) sont de plus en plus utilisés. Leurs usages sont variés : feuilles d'exercices pour les étudiants, cahiers de recherche, tutoriels...

Les outils développés par le projet Jupyter sont en constante évolution. On peut citer quelques-uns de ces outils largement utilisés par différentes communautés allant du calcul scientifique à la science des données.

Au cours de cette soirée thématique, nous avons fait un tour d'horizon des outils proposés par le projet Jupyter afin de montrer leur intérêt pour la productivité de la recherche et la diffusion des travaux scientifiques. Nous avons également présenté les initiatives en cours de réalisation dans la communauté des mathématiques au niveau national pour faciliter leurs utilisations aussi bien pour la recherche que pour l'enseignement.

2.2.7 JCAD, Journées Calcul & Données

Les GIS FRANCE GRILLES et GRID'5000, le Groupe Calcul, le GDR RSD, GENCI et les partenaires de l'Equipex EQUIP@MESO ont organisé ensemble les JCAD, Journées Calcul Données : Rencontres scientifiques et techniques du calcul et des données. Ces journées rassemblent à la fois les objectifs, les thématiques ainsi que les publics cibles des journées Mésocentre (Groupe Calcul),

des journées Méso-challenge (Equipex EQUIP@MESO) et des [journées SUCCES](#) (France Grilles, Grid'5000, GDR RSD, Groupe Calcul) et en prennent le relais.

Les JCAD sont donc dédiées à la fois aux utilisateurs et aux experts techniques des infrastructures et des services associés.

Ces journées associent exposés pléniers, sessions dédiées aux posters, démonstrations et tables rondes sur des sujets d'actualité.

Les JCAD ont eu lieu en [2018](#), [2019](#) et [2020](#) et ont regroupé une centaine de participants.

2.2.8 JoSy IA 2018

Nombre de participants : 60

En partenariat avec Resinfo (réseau des administrateurs systèmes et réseaux) et le GDR MaDICS, le GDR Calcul a participé à l'organisation d'une [JoSy](#) sur le thème de l'intelligence artificielle (IA).

Les thèmes suivants ont été abordés :

- une introduction à l'algorithmie de l'IA
- la théorie et application de l'apprentissage machine
- les infrastructures techniques pour l'IA
- l'utilisation de l'IA dans la science des données

Cette journée thématique a lieu à Strasbourg en octobre 2018 et sur deux demi-journées.

2.2.9 Journée différentiation algorithmique

Nombre de participants : 35

Calculer une dérivée est parfois délicat, surtout lorsque la fonction à dériver est elle-même le résultat d'un calcul numérique. En pratique, cela pose des problèmes numériques et algorithmiques : d'une part, il est difficile de contrôler l'erreur de troncature d'un calcul aux différences finies et, d'autre part, le calcul symbolique peut conduire à une explosion de la complexité des expressions.

Cette difficulté se rencontre dans les problèmes où l'on cherche à optimiser une quantité résultant elle-même d'un calcul approximatif (optimisation d'un critère sous contrainte), notamment : le contrôle optimal, l'assimilation de données, les problèmes inverses, l'analyse de sensibilité et l'apprentissage profond.

La différentiation algorithmique (AD) permet de s'affranchir d'une partie des calculs. L'AD prend un code de calcul, ainsi qu'une description des variables d'entrée et de sortie, et produit un nouveau code qui calcule les dérivées des sorties par rapport aux entrées. Ceci utilise des

concepts de compilation et d'analyse de programme (sans recours au calcul formel). Il existe deux modes de DA qui diffèrent par la manière d'appliquer la règle de dérivation d'une fonction composée :

- le mode direct, qui est utilisé lorsque le nombre d'entrées est inférieur au nombre de sorties,
- le mode inverse, qui est utilisé lorsque le nombre d'entrées est supérieur au nombre de sorties, notamment pour calculer le gradient d'une fonctionnelle par rapport à un grand vecteur.

Le mode inverse équivaut au calcul de l'état adjoint (utilisé dans le contrôle optimal) et est proche de la rétropropagation dans les réseaux de neurones. Ces résultats, bien connus des spécialistes de chaque discipline, ne sont pas nouveaux, mais ils méritent d'être présentés dans une perspective plus large.

Le but de cet atelier était de réunir des spécialistes de ces différentes technologies afin de mettre en évidence les besoins communs, les différents cas d'utilisation et les limites des différentes méthodes. Un autre objectif était de rencontrer différentes communautés qui partagent ces techniques.

L'atelier comprenait des présentations éducatives présentant les principes des méthodes, des commentaires sur différentes applications et des présentations d'extensions récentes ou plus originales. Il s'est tenu le 24 janvier 2019 à l'IPGP et a regroupé une trentaine de participants : <https://calcul.math.cnrs.fr/2019-01-journee-algorithmic-differentiation.html>.

2.2.10 Journées “Calcul et Apprentissage”

Nombre de participants : 50

Ces journées ont eu lieu en 2019 les 24 avril (14h-18h) et 25 avril (9h-17h), à Lyon sur le campus de l'université Claude Bernard Lyon 1. Elles ont été financées par le GDR Calcul et le Labex MILyon. L'objectif de ces journées était de présenter les problématiques et les principaux outils de l'apprentissage automatique, en se concentrant sur les aspects algorithmiques et implémentation, à l'intention de la communauté du calcul scientifique. Les interventions ont pris trois formes différentes : exposés introductifs, exposés de recherche et travaux pratiques. Deux exposés introductifs, sous forme de cours d'une heure et demie chacun, ont été donnés par Aurélien Garivier (UMPA, ENS Lyon) :

- Introduction à l'apprentissage automatique : cadre formel, premiers algorithmes ;
- Minimisation du risque structurel versus réseaux de neurones.

Ces cours ont permis d'introduire les bases théoriques et de mettre en évidence les questions ouvertes concernant l'apprentissage, en particulier avec un point de vue statistique. Un exposé

a permis d'aborder avec un point de vue théorique une autre problématique dans les problèmes d'apprentissage :

- Optimisation stochastique pour l'apprentissage.

Cet exposé de 45 minutes a été donné par Aymeric Dieuleveut (CMAP, Polytechnique). Deux exposés de recherche de 45 minutes chacun ont présenté des travaux récents et en cours, concernant des applications :

- Rare event computations and machine learning for climate dynamics, par Freddy Bouchet (Laboratoire de Physique, ENS Lyon) ;
- Deep Learning for inverse problems, par Patrick Pérez (Valeo.ai).

Deux séances de travaux pratiques, de 2 heures chacune, ont été animées par Jean Feydy (DMA, ENS). La première séance a permis de se familiariser avec les algorithmes présentés dans le premier cours et de tester leurs limites. Des outils plus élaborés et plus récents ont été présentés lors de la seconde séance.

Les supports de ces présentations ont été mis en ligne et sont accessibles sur le site du Groupe Calcul : <https://calcul.math.cnrs.fr/2019-04-journees-calcul-apprentissage.html>.

2.2.11 Congrès SMAI 2019 : mini-symposium “S’affranchir du maillage : résoudre les EDP par le deep learning”

Nombre de participants : 30

Les équations aux dérivées partielles (EDP) sont omniprésentes dans presque tous les domaines de la physique appliquée et de l'ingénierie, englobant un large éventail de modèles physiques et phénoménologiques et de lois de conservation. En particulier, de nombreux phénomènes physiques peuvent être décrits comme un système d'EDP non linéaires. Avec l'avènement des schémas de discrétisation spatiale tels que les différences finies, les volumes finis, ou les méthodes spectrales, beaucoup de ces problèmes peuvent maintenant être résolus, en utilisant des algorithmes HPC.

Toutefois, d'importants défis subsistent en ce qui concerne la précision des solutions, la convergence et le calcul du coût associé à ces méthodes pour les problèmes présentant une grande complexité.

Ces dernières années, les techniques d'apprentissage machine ont été proposées comme une alternative prometteuse aux méthodes numériques conventionnelles. Il a été démontré que ces techniques présentent plusieurs avantages par rapport aux méthodes traditionnelles, y compris :

- La solution n'a pas de maillage, est analytique et continuellement différentiable.

- L'utilisation de réseaux de neurones fournit une solution avec de très bonnes propriétés de généralisation.
- Le coût de calcul de l'apprentissage ne dépend que faiblement de la dimension du problème.
- Relativement peu de paramètres peuvent être nécessaires pour modéliser des solutions complexes.
- Les dérivées temporelles et spatiales peuvent être traitées de la même manière.
- L'apprentissage est hautement parallélisable sur les GPU en utilisant des bibliothèques d'apprentissage approfondi *open source*.

Au cours de ce symposium, nous avons donné un aperçu des techniques de pointe utilisées aujourd'hui pour résoudre les EDP grâce au deep learning et nous avons essayé de donner une idée de ce qui sera possible à l'avenir.

Les supports de ces présentations ont été mis en ligne et sont accessibles sur le site du Groupe Calcul : <https://calcul.math.cnrs.fr/2019-05-mini-symposium-smai.html>.

2.2.12 Journée “Julia et Optimisation”

Nombre de participants : 110

Julia est un langage qui est de plus en plus populaire dans la communauté des mathématiciens. C'est une alternative libre à MATLAB et c'est un langage conçu pour faire des sciences, performant, avec une syntaxe qui permet un apprentissage rapide.

La journée “Julia et optimisation” a été organisée le lundi 17 juin 2019 sur le campus de l'Université de Nantes. Elle a accueilli 110 participants. Les intervenants et organisateurs étaient : Mathieu Besançon, François Févotte, Rémi Garcia, Benoît Legat, Pierre Navaro, François Pacaud, Xavier Pillet, Anthony Przybylski, Didier Robbes, Charlie Vanaret, François Vanderbeck et Diego Javier Zea.

Le programme disponible sur le site du groupe calcul <https://calcul.math.cnrs.fr/2019-06-journee-julia-optimisation.html> était le suivant:

- Ouverture : Xavier Gandibleux, Frédéric Hérau (Directeur de la Fédération Recherche Mathématiques des Pays de Loire) et Claude Jard (Directeur du Laboratoire des Sciences du Numérique) ;
- LightGraphs, structure, abstractions and algorithms, Mathieu Besançon (Centrale Lille et Polytechnique Montréal) ;
- Coluna.jl : une implémentation générique d'un algorithme de « branch-and-price-and-cut » accessible via l'interface de modélisation JuMP., François Vanderbeck (Atoptima) ;

- Optimisation combinatoire multi-objectif : méthodes de résolution exacte et solveur vOpt, Anthony Przybylski (Université de Nantes) ;
- Résolution numérique du système de Vlasov-Poisson en Julia, Pierre Navaro (Institut de Recherche Mathématique de Rennes) ;
- L'extension de JuMP à la programmation Sum-of-Squares, Benoît Legat (Université catholique de Louvain), Chris Coey (MIT), Robin Deits (MIT), Joey Huchette (Rice University), Amelia Perry (MIT) et Tillmann Weisser (Los Alamos National Laboratory) ;
- Solving large-scale NLP problems with JuMP and Knitro, Francois Pacaud (Artelys France) ;
- Quelques expériences autour de l'arithmétique flottante en Julia, François Févotte (TriScale Innov), Pascal Benchimol, Laurent Plagne ;
- Solving distillation column equilibrium using interval methods, Charlie Vanaret (Fraunhofer ITWM kaiserslautern) ;
- Ateliers :
 - Atelier 1 : Using Knitro efficiently with JuMP par Francois Pacaud (Artelys France) ;
 - Atelier 2 : Getting started with Julia tools for reproducible science par Mathieu Besançon (Centrale Lille et Polytechnique Montréal) ;
 - Atelier 3 : Premiers pas de l'utilisateur de Coluna dans l'implémentation d'une approche de génération de colonnes par François Vanderbeck (Atoptima).

Sur le site <https://julialang.univ-nantes.fr/journee-julia-et-optimisation/> a été mis en ligne l'ensemble des origines des inscrits, toutes les vidéos des interventions, les liens vers les supports, et une sélection de photos. La journée a été un gros succès avec plus de 110 participants. Le programme était centré sur l'optimisation et la recherche opérationnelle, mais il a attiré beaucoup de personnes curieuses de découvrir ce nouveau langage. Grâce au soutien du groupe Calcul, nous avons pu établir des liens durables avec la communauté francophone Julia. D'autres événements autour du langage ont eu lieu et sont prévus car Julia est particulièrement performant dans de nombreux domaines mathématiques: arithmétique, équations différentielles, optimisation et contrôle et apprentissage statistique.

2.2.13 Journée histoire du calcul

Nombre de participants : 80

Dans le cadre des 80 ans du CNRS, nous avons souhaité comprendre les évolutions qui ont conduit au paysage actuel du calcul. La thématique est large, elle est aussi très transverse et trouve son origine bien avant l'émergence de l'ordinateur. Cette journée qui s'est tenue le 28 novembre 2019 à Grenoble avait pour objectif d'aborder l'histoire du calcul sous différents angles, depuis les mathématiques appliquées jusqu'au lien avec l'informatique et le calcul intensif.

Elle a réuni les intervenants suivants :

- Pierre Mounier-Kühn, CNRS & Sorbonne Université,
- Thierry Dumont, Institut Camille Jordan, Université Lyon 1,
- Bernard Ycart, Laboratoire Jean Kuntzmann, Université Grenoble Alpes,
- Patrick Chenin, Laboratoire Jean Kuntzmann, Université Grenoble Alpes,
- Gerhard Wanner, Martin J. Gander, Université de Genève,
- Olivier Pironneau, Laboratoire Jacques Louis Lions, Université Pierre et Marie Curie, Académie des Sciences,
- Alain Guyot, Association ACONIT,
- Jean-Guillaume Dumas, Françoise Jung, Clément Pernet, Laboratoire Jean Kuntzmann, Université Grenoble Alpes.

Les détails du programme sont disponibles sur <https://calcul.math.cnrs.fr/2019-11-journee-histoire-du-calcul.html>.

2.2.14 Autres contributions

Le GDR participe ponctuellement au soutien financier d'évènements dont il n'est pas organisateur, notamment :

- le [7th Workshop on Parallel-in-Time methods](#) qui a eu lieu à Roscoff en mai 2018,
- le mois thématique "[Mathématiques pour le signal](#)" qui a lieu à Marseille en février 2021,
- la [SIAM Conference on Computational Science and Engineering \(CSE21\)](#) qui aura lieu aux Etats-Unis en mars 2021.

2.2.15 Bilan des objectifs du projet 2017-2021

Dans le projet 2017, nous souhaitions ouvrir notre champ d'action à de nouvelles thématiques et lancer un grand nombre d'actions en incluant la plupart des idées issues de nos réflexions. De nombreuses actions ont été effectivement menées (si l'on exclut l'année 2020 perturbée par la crise sanitaire) et elles ont rencontré un succès d'audience, ce qui nous a permis de réaliser une partie de nos objectifs, notamment :

- les actions de formation dans le domaine de la précision numérique et celui de l'analyse des données géométriques,
- l'exploration de l'apprentissage machine et de la science des données,
- la promotion de bibliothèques mathématiques *open source*,
- la diffusion des nouveaux langages (Julia) et environnements pour le calcul (Jupyter).

Toutefois, essentiellement par manque de temps, nous sommes loin d'avoir couvert toutes les thématiques et réalisé tout notre programme.

Pour la période 2022-2027, nous souhaitons donc construire un projet un peu moins ambitieux, mais plus réaliste en matière de couverture thématique et de nombre d'actions. Ce choix nous a conduits à identifier principalement quatre axes thématiques qui sont détaillés dans la section suivante.

3 Projet 2022-2026

Dans son nouveau projet, le GDR Calcul souhaite concentrer ses actions autour de quatre axes thématiques :

- la **précision numérique** (section 3.1) : elle joue un rôle essentiel dans la simulation numérique et représente un défi dans le contexte du calcul intensif,
- le **calcul quantique** (section 3.2) : il s'agit d'une discipline en pleine expansion qui a encore peu de liens avec la communauté du calcul *classique*,
- les **modèles stochastiques** (section 3.3) : le lien avec la communauté du calcul intensif promet des avancées intéressantes dans ce domaine,
- l'**apprentissage machine** (section 3.4) : on s'intéressera particulièrement à son application au domaine de la simulation numérique, notamment des équations aux dérivées partielles.

En plus de ces axes thématiques prioritaires, nous continuerons notre activité de veille dans le domaine "historique" du GDR autour de la résolution des EDP.

Pour chaque thème, nous mettrons en œuvre des actions en employant la ou les modalités les plus pertinentes parmi la liste suivante :

- journées thématiques,
- ateliers,
- écoles thématiques,
- actions nationales de formation,
- cafés virtuels.

Les cafés virtuels sont une nouvelle modalité que nous avons récemment expérimentée en réaction à la crise sanitaire. Il s'agit d'un séminaire d'une cinquantaine de minutes sur un thème identifié, diffusé en webconférence et suivi d'une séance de questions et de discussions libres. Ces cafés virtuels présentent l'avantage d'être légers à organiser et de pouvoir être suivis par une large audience nationale, y compris en différé grâce à l'enregistrement vidéo.

Outre les thématiques de recherche mentionnées précédemment, nous avons également identifié les sujets suivants pour nos actions :

- la **science ouverte** dans le domaine du calcul scientifique et intensif (section 3.5),

- la **parité** dans le domaine du calcul scientifique (section 3.6),
- la **médiation scientifique** auprès des jeunes (section 3.7),
- les **jeunes chercheurs** comme bénéficiaires de nos actions (section 3.8).

3.1 Précision numérique

La simulation numérique est largement utilisée, que ce soit pour aider à comprendre, modéliser et prédire des phénomènes fondamentaux dans de nombreuses disciplines scientifiques, comme en géophysique, en chimie ou biologie, ou pour aider au développement de procédés, dans l'industrie ou en médecine.

Les problèmes ainsi adressés par le calcul numérique se complexifient constamment, soit par leur caractère multi-échelles induit par la physique, soit par le nombre d'inconnues traitées, conduisant à des problèmes d'algèbre linéaire ou de données de très grande taille, soit encore par leur mise en œuvre pratique sur les machines de calcul intensif.

Maîtriser la précision numérique des simulations dans ce contexte est complexe, l'accumulation des erreurs impactant potentiellement les fiabilités de ce type de simulations, pour lesquelles il n'est pas toujours possible de constituer une base statistique. Outre les imprécisions liées aux modèles, les erreurs proviennent à la fois de la troncature de l'arithmétique flottante, de la précision de l'approximation discrète, en particulier dans un contexte non linéaire, et aussi de la qualité de l'implantation informatique. Dans le cas de calculs massivement parallèles la probabilité des fautes matérielles augmente avec le nombre d'unités de calcul, la problématique de développer des méthodes résilientes aux fautes devient critique et rejoint par certains aspects les problèmes de précision de calcul (e.g. propagation d'un bit-flip). De plus, les architectures offrent désormais une précision réduite et savoir garantir la précision de calculs effectués avec cette précision réduite nécessite de revisiter les résultats classiques, puisque leurs hypothèses ne sont plus applicables.

Plusieurs défis peuvent être identifiés :

- comment identifier les sources d'erreur et estimer la sensibilité numérique du système à la propagation de perturbations ?
- est-il nécessaire de conserver en tout point et à chaque instant la même représentation des flottants en machines ?
- quels sont les paradigmes de programmation parallèle adaptés, notamment dans la perspective de calculs exaflopiques ?
- comment intégrer dans les simulations numériques des capacités de résilience aux diverses fautes ?

Toutes ces problématiques restent encore à l'heure actuelle l'objet de nombreuses recherches

fondamentales qui nécessitent souvent de revoir la mise en œuvre des simulations numériques, aussi bien dans leurs algorithmes, que leurs méthodes et outils associés et ont à notre sens toute leur place dans le cadre du GDR Calcul, tant dans ses aspects de mathématiques appliquées que trans-disciplinaires.

L'objectif, en soutenant cette thématique, serait de renforcer les actions collaboratives des chercheurs travaillant sur ces sujets connexes mais dans des communautés voisines et de promouvoir les techniques et outils numériques et logiciels permettant leur mise en œuvre dans différents contextes applicatifs.

3.2 Calcul quantique

Les premiers modèles d'ordinateur quantique datent du début des années 1980. Des algorithmes quantiques, pouvant en théorie fonctionner sur des ordinateurs quantiques, sont développés dès les années 1990. Où en est-on aujourd'hui ? Pour réaliser un ordinateur quantique fiable capable de résoudre efficacement des problèmes pratiques, il faut pouvoir utiliser plusieurs millions de qubits ensemble avec un taux d'erreur bas et un temps de cohérence élevé. Un tel ordinateur n'arrivera pas avant des décennies. Toutefois, nous sommes aujourd'hui dans une ère intermédiaire appelée l'ère du "NISQ" (Noisy Intermediate Scale Quantum) : des ordinateurs quantiques existent, comprenant plusieurs dizaines de qubits, mais sans correction d'erreur et avec un temps de cohérence limité.

Malgré ces imperfections, de nombreux algorithmes quantiques sont développés afin de profiter de la puissance de ces ordinateurs. L'objectif est d'obtenir des algorithmes ayant une complexité temporelle bien meilleure que celle des algorithmes classiques tournant sur les ordinateurs actuels. On pourra par exemple citer la transformation de Fourier quantique, les marches quantiques qui sont l'équivalent des marches aléatoires classiques à la base de nombreux autres algorithmes, ou encore les algorithmes quantiques d'optimisation comme le "Quantum Approximate Optimisation Algorithm" (QAOA). On retrouve ces algorithmes quantiques dans de nombreuses disciplines comme l'apprentissage machine (une autre thématique développée dans ce GDR), l'algèbre linéaire, l'optimisation, la cryptographie, etc. Comment s'écrivent les algorithmes quantiques ? Comment fonctionnent-ils ? Quels sont les avantages et inconvénients du quantique ? Quels sont les types de problèmes et les domaines d'application qui bénéficieront du calcul quantique ? Le calcul quantique est-il amené à remplacer le calcul classique ou se couplera-t-il au calcul traditionnel en mode hybride comme le monde de l'intelligence artificielle peut se coupler à celui de la simulation numérique ?

Les ordinateurs quantiques existent déjà aujourd'hui. Il est possible de les utiliser principalement au travers d'un accès *cloud* du fait des conditions d'opérations très contraintes de ces technologies. Des constructeurs comme IBM Quantum experience, Google Quantum AI ou bien

Microsoft Azure Quantum, mettent à disposition leurs ordinateurs quantiques, gratuitement pour un nombre limité de qubits. Si la plupart de ces technologies reposent sur les matériaux supraconducteurs, le passage à l'échelle n'est pas acquis. L'enjeu pour le calcul quantique rappelle celui du calcul scientifique dans les années 80 : on ne peut pas prédire aujourd'hui l'approche qui sortira vainqueur de la compétition technologique. En France, l'écosystème est riche de startups prometteuses qui sont issues du monde de la recherche académique. La société Pasqal disposera en 2021 d'un simulateur quantique de 100 qubits basés sur des atomes froids. D'autres sociétés comme Quandela développent un système quantique à base de qubits photoniques. Si l'enjeu de la fourniture de plateformes matérielles quantiques est important, l'écosystème logiciel est tout aussi crucial. Pour les constructeurs, proposer des API (Application Programming Interface) permettant de simplifier l'écriture des circuits quantiques est un passage obligé : Qiskit (Python) pour IBM, Cirq (Python) pour Google et le langage Q# pour Microsoft. La société française Atos met à disposition un environnement logiciel complet (compilateur, bibliothèques bas niveaux, etc.) permettant de prendre en considération plusieurs types de technologies matérielles et la propriété des implémentations de celle-ci. Il existe aussi des bibliothèques indépendantes, comme la bibliothèque Yao en Julia.

À ce jour, la communauté du calcul scientifique *classique*, qui utilise des ressources de calcul intensif pour ses simulations numériques, fréquente très timidement le monde du calcul quantique. Alors que la première plateforme hybride HPC/Quantique verra le jour fin 2021 sur le site du TGCC (Très Grand Centre de Calcul du CEA) avec plus de 100 qubits, il semble important qu'un travail pédagogique soit mené, sous la forme de journées thématiques, de workshops et de formations. L'objectif de ces actions sera :

- d'expliquer les aspects théoriques du calcul quantique,
- d'identifier les algorithmes qui bénéficient d'une accélération quantique,
- de montrer comment les implémenter sous forme de circuits quantiques dans les bibliothèques existantes,
- de proposer une approche de la programmation hybride HPC/quantique,
- de comprendre quels sont les environnements nécessaires à l'exécution de travaux quantiques (sur des plateformes le permettant) ou hybrides quantiques/classiques.

La thématique du calcul quantique reçoit de plus en plus d'attention et se trouve dans une ère où le développement des architectures quantiques s'accélère rapidement. En France, le Plan National Quantique qui vient d'être annoncé par le président de la République va lui donner une impulsion supplémentaire. Le calcul quantique est fortement interdisciplinaire puisqu'il touche toutes les disciplines concernées par le calcul numérique. Le GDR Calcul se propose donc d'explorer dès maintenant cette thématique par un travail de prospective.

3.3 Modèles stochastiques

La thématique des modèles stochastiques a été intégrée au projet du GDR Calcul lors du renouvellement précédent, avec l'objectif de créer des interactions entre les communautés probabilistes, statistiques et calcul scientifique. Dans ce cadre, des journées "Calcul et Apprentissage" ont été organisées au printemps 2019 à Lyon.

Pour continuer à développer les actions du GDR Calcul autour des modèles stochastiques, nous avons identifié plusieurs problématiques qui pourraient être étudiées, en partie en lien avec d'autres thématiques du GDR.

Ces problématiques peuvent être divisées en 2 catégories: simulation de modèles stochastiques d'une part, utilisation de méthodes probabilistes et statistiques pour des problèmes déterministes d'autre part.

Dans la première catégorie, le modèle étudié est stochastique, et le résultat d'une simulation est aléatoire. L'inférence des propriétés des systèmes stochastiques par méthode de Monte-Carlo demande des échantillons de grande taille. Pour réduire les coûts de simulation, plusieurs approches complémentaires sont possibles. D'une part, on peut utiliser des techniques de réduction de variance (par exemple méthodes multi-niveaux et estimateurs multi-fidélité [6, 4]) pour diminuer la taille des échantillons requis pour obtenir une précision donnée. D'autre part, on peut se tourner vers des méthodes de calcul scientifique haute performance, notamment quand les systèmes étudiés sont de grande dimension, afin de réduire le temps de calcul par réalisation. L'étude des simulateurs stochastiques, par exemple les schémas numériques pour les EDP Stochastiques ou encore la réduction de modèles en grande dimension, est un sujet de recherche en développement [2]. Enfin, ces modèles sont particulièrement adaptés à des techniques d'optimisation de complexité comme le calcul en précision mixte. De plus, le caractère indépendant des nombreuses réalisations à effectuer rend ces codes hautement parallélisables et adaptés aux architectures de calcul haute performance. Cependant, travailler avec ces architectures de calcul spécifiques demande en général de revisiter les algorithmes et de nouvelles techniques pour leur analyse.

Dans la seconde catégorie, le principe est d'interpréter un problème déterministe avec une perspective probabiliste (par exemple la solution d'une EDP parabolique est associée aux trajectoires solutions d'une équation différentielle stochastique) ou statistique (l'entrée et la sortie de l'algorithme sont des mesures de probabilités au lieu de valeurs fixées des paramètres). Particulièrement, on pourra s'intéresser aux sujets suivants. D'abord, les approches stochastiques pour la décomposition de domaine [3]. Ensuite, les approches probabilistes pour l'algèbre linéaire en grande dimension [5]. Enfin, les méthodes de résolution d'EDP par techniques d'apprentissage [1, 7]. Ce dernier sujet est en plein développement et en interaction avec la thématique *Apprentissage machine et interactions* de ce projet (cf. 3.4). On pourra regarder la construction et

l'implémentation de méthodes hybrides, qui combinent approches déterministes et stochastiques selon les échelles considérées.

Dans les années à venir, on cherchera à renforcer les liens avec le GDR [MASCOT-NUM](#), en particulier sur la thématique des simulateurs stochastiques. On profitera de la présence de membres appartenant aux comités scientifiques des deux GDR pour créer des actions et des interactions. Certaines thématiques pourront aussi être développées en lien avec le GDR [MaNu](#).

Les actions envisagées pourront être de deux types, autour des sujets mentionnés ci-dessus, afin de renforcer les interactions entre communautés probabilistes, statistiques et du calcul scientifique. D'une part, il s'agira de communiquer auprès des communautés probabilistes et statistiques sur les possibilités offertes par les techniques de calcul scientifique haute performance (y compris l'utilisation de méso-centres); par exemple, on pourrait choisir la thématique de la simulation Markov Chain Monte Carlo, en lien avec l'interprétation Bayésienne pour les problèmes inverses, pour introduire ces outils avancés. D'autre part, il s'agira de communiquer auprès de la communauté du calcul scientifique des problématiques spécifiques à la simulation des modèles stochastiques, par rapport aux problèmes déterministes; par exemple, on pourrait choisir la thématique du couplage de modèles déterministes et stochastiques pour introduire cette problématique. En plus de journées thématiques, on pourra chercher à fédérer les spécialistes des méthodes numériques stochastiques, afin de faire émerger les problématiques autour du calcul dans leurs travaux.

- [1] C. Beck, M. Hutzenthaler, A. Jentzen, and B. Kuckuck, *An overview on deep learning-based approximation methods for partial differential equations*, 2020.
- [2] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, SIAM, Philadelphia, PA, 2017.
- [3] F. Bernal, and J. A. Acebrón, *A multigrid-like algorithm for probabilistic domain decomposition*, *Computers & Mathematics with Applications*, Vol .72, No. 7, 2016.
- [4] M. B. Giles, *Multilevel Monte Carlo methods*, *Acta Numerica*, 24:259-328, Cambridge University Press, 2015.
- [5] P.-G. Martinsson, and J. Tropp, *Randomized numerical linear algebra: Foundations & algorithms*, arXiv preprint arXiv:2002.01387, 2020.
- [6] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, *SIAM Review*, Vol. 60, No. 3, pp. 550-591, 2018.
- [7] M. Raissi, P. Perdikaris, G. E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, *Journal of Computational Physics*, 378, 686-707, 2019.

3.4 Apprentissage machine et interactions

L'apprentissage automatique, depuis le début des années 2010, grâce à des méthodes de plus en plus performantes et des moyens de calcul importants, a pris une importance considérable dans le paysage scientifique. Après de spectaculaires résultats en reconnaissance d'image, l'apprentissage a été utilisé dans un nombre très important de domaines (chimie et molécule, diagnostics médicaux, langage, traitement du signal audio, vidéo, etc.).

D'un autre côté, la simulation numérique et le calcul scientifique (algèbre linéaire, résolution d'EDP ou d'EDO, géométrie computationnelle, optimisation) ont continué à être au centre de nombreuses recherches pour obtenir des codes robustes et efficaces pour les simulations physiques, médicales ou biologiques. Depuis 3 ou 4 ans, les méthodes d'apprentissage ont fait une apparition remarquable dans les méthodes numériques, les EDP ou l'algèbre linéaire. L'apprentissage a commencé à être utilisé pour résoudre des EDP directement ou des problèmes inverses [1]. Plus récemment, il a été utilisé pour compléter et corriger des méthodes numériques [2] ou pour construire de nouveaux modèles [3]. Pour le moment, une grosse partie des interactions entre calcul scientifique et apprentissage semble se faire autour de l'apprentissage supervisé, basé sur des réseaux de neurones convolutifs ou pleinement connectés.

Plusieurs défis peuvent être identifiés :

- Peut-on intégrer plus profondément l'apprentissage dans les codes numériques ? Jusqu'à envisager des codes qui s'adaptent au fur et à mesure des exécutions ?
- Si les méthodes d'apprentissage font partie intégrante de l'approche numérique, comment peut-on certifier à priori les résultats comme la convergence en maillage ou autre ?
- Un appel fréquent de méthodes d'apprentissage peut être coûteux à évaluer. Peut-on réduire ce coût en allant vers des méthodes plus légères (de type « tiny IA ») ?
- À l'inverse est ce que le calcul scientifique peut permettre l'amélioration ou l'analyse des méthodes d'apprentissage profond ?

Toutes ces problématiques qui commencent à faire l'objet de recherches fondamentales ont à notre sens toute leur place dans le cadre du GDR Calcul, tant du point de vue de l'apprentissage que de celui des applications en EDP, calcul scientifique et algèbre linéaire.

L'objectif, en soutenant cette thématique, serait de renforcer les actions collaboratives des chercheurs travaillant sur ces sujets et de sensibiliser la communauté du calcul scientifique à des aspects variés de l'apprentissage qui sont parfois moins utilisés comme l'apprentissage par renforcement, les méthodes dites explicables ou l'apprentissage géométrique. On ferait également la promotion d'outils numériques permettant la mise en œuvre de méthodes d'apprentissage les plus pointues dans des codes de simulations numériques.

[1] M. Raissi, P. Perdikaris and G.E. Karniadakis, *Physics-informed neural networks: A deep learn-*

ing framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, Vol. 378, 686-707, 2019

[2] Deep Ray, Jan S. Hesthaven, *Detecting troubled-cells on two-dimensional unstructured grids using a neural network*, Journal of Computational Physics, Vol. 397, 108845, 2019.

[3] G. Novati, H.L. de Laroussilhe and P. Koumoutsakos, *Automating turbulence modelling by multi-agent reinforcement learning*, Nat Mach Intell 3, 87–96, 2021.

3.5 Science ouverte

La science ouverte est devenue une stratégie clé ces dernières années en Europe et au niveau international (¹, ² et ³). Elle englobe l'ensemble du processus scientifique, depuis le développement d'une idée jusqu'à sa publication.

Pour la recherche expérimentale, elle comprend la description de l'expérience via les carnets Open Notebooks, et l'accès aux données pour refaire l'expérience et valider ses résultats via les données Open Data. Pour le calcul, les aspects principaux sont l'accès au code via l'Open Source, pour permettre l'inspection et la réutilisation, et le partage des environnements de calcul via l'Open Environment, pour assurer la reproductibilité. Enfin, l'accès aux publications via l'Open Access concerne toutes les études scientifiques.

La mise en œuvre de cette chaîne complète rend la science plus transparente, accessible à tous, fiable et reproductible. Elle favorise également la transversalité et la collaboration entre le monde universitaire et l'industrie. Toutefois, faire en sorte que la science ouverte fasse partie intégrante du processus de recherche demande un effort de pédagogie et de communication auprès de la communauté.

Le GDR Calcul propose de faire la promotion de la science ouverte en animant des ateliers et formations, et en diffusant des documents simples et accessibles montrant comment faire en sorte que sa recherche soit plus transparente et reproductible. Dans le domaine du calcul scientifique, la reproductibilité computationnelle est un sujet particulièrement important, qui a déjà été abordé dans formations organisées par le GdR et qui le sera encore plus dans l'avenir. Au-delà des aspects techniques (comment rendre un calcul reproductible), nous nous intéressons aussi au rôle du calcul dans une étude scientifique. On parle alors de répliquabilité, et pose des questions comme par exemple : comment s'assurer que des changements inévitables, par exemple en passant d'un centre de calcul à un autre, n'influent pas sur l'interprétation scientifique des résultats ?

¹<https://www.science-ouverte.cnrs.fr/>

²<https://en.unesco.org/science-sustainable-future/open-science>

³<https://www.openaire.eu/>

Un autre aspect de la science ouverte qui est au coeur des préoccupations du GDR Calcul est l'ouverture des logiciels. Traditionnellement, l'Open Source se concentre sur les licences, dont le choix judicieux est une condition importante mais pas suffisante pour qu'un logiciel devienne inspectable et réutilisable. Il faut également penser à une bonne documentation, et mettre en place une stratégie de valorisation. Les équipes et laboratoires de recherche sont pour l'instant mal préparés à faire face à ces défis, parce que la publication des logiciels n'a pas été une priorité dans le passé.

Le GDR Calcul fera également l'inventaire des infrastructures existantes facilitant l'Open Notebook, l'Open Data, l'Open Source, l'Open Environment et l'Open Access.

3.6 Mission parité

Lors de son dernier conseil scientifique (novembre 2019), le GDR a discuté des actions possibles dans le domaine de la parité femmes-hommes dans le domaine du calcul scientifique, notamment à partir du constat que les femmes sont particulièrement peu nombreuses dans ce domaine. Le calcul scientifique étant à la croisée des mathématiques et de l'informatique, il est naturel de s'appuyer sur les initiatives de l'INSMI ⁽⁴⁾ et de l'INS2I ⁽⁵⁾.

Suite à ces discussions, nous avons identifié les deux pistes d'actions suivantes.

3.6.1 Parité oratrices/orateurs

Dans nos évènements, porter une attention particulière au fait que la proportion de femmes dans la discipline soit au moins respectée :

- d'abord parmi les oratrices/orateurs,
- ensuite dans les comités d'organisation.

Le bureau du GDR veillera à rappeler ce principe aussi souvent que nécessaire.

3.6.2 Actions à destination des filles et des jeunes femmes pour susciter des vocations

Une réponse évidente au déficit de femmes dans les métiers du calcul est d'agir avant que se décide l'orientation dans les études secondaires et supérieures. Nous avons notamment repéré l'initiative "[Filles et info: une équation lumineuse](#)" qui fait rencontrer directement lycéennes et femmes des filières mathématique et informatique dans le but de lutter contre les mécanismes à l'origine du déficit des vocations féminines. Au niveau du collège, nous avons également

⁴<https://parite.math.cnrs.fr/>

⁵<https://www.ins2i.cnrs.fr/fr/parite-egalite-femmes-hommes>

remarqué l'action "[L codent, L créent](#)" qui propose des ateliers de programmation à destination des collégiennes.

D'autres initiatives comme "Women in Sage" ou encore les [Pyladies](#) peuvent également nous inspirer.

Le GDR pourrait intervenir à plusieurs niveaux :

- en participant au soutien financier de ces projets, ce qui permet également de se faire connaître,
- en intervenant directement dans les actions qui sont menées.

Le lien avec "Filles et info: une équation lumineuse" serait aisé dans la mesure où Nathalie Revol qui est membre de notre comité scientifique fait partie des organisateurs du projet.

3.7 Médiation scientifique

La médiation scientifique représente un enjeu important pour la communauté de la recherche académique en général et pour nos disciplines en particulier. D'une part, le calcul scientifique ne bénéficie pas de la même exposition que d'autres domaines scientifiques : il bénéficierait largement d'actions de communication vers le grand public. D'autre part, menée auprès des jeunes publics, la médiation permet de susciter l'intérêt de futur·e·s étudiant·e·s et de les attirer vers les filières des mathématiques appliquées et du calcul scientifique. Il nous semble particulièrement important de se tourner vers les réseaux d'éducation prioritaire afin de toucher des publics qui sont peut-être moins familiers des métiers de la recherche scientifique. Les actions de médiation peuvent également s'inscrire dans le cadre de la mission parité en menant des actions spécifiques à destination des jeunes filles.

Les membres de la communauté du calcul, qu'ils soient ingénieurs ou chercheurs, sont régulièrement sollicités pour présenter leurs travaux et leur métier au grand public. Or, on a tous constaté qu'il n'est pas aisé d'expliquer cette discipline à un public de non spécialistes. Le GDR Calcul propose donc de faciliter et d'inciter les actions de médiation en collectant des exemples de présentations ou de projets utilisés pour promouvoir les métiers en lien avec le Calcul auprès des scolaires de tout âges. Ce matériel sera mis à disposition sous forme de kit afin d'être utilisé lors de manifestations telles que la Fête de la Science. Il pourra servir également dans le cadre d'initiatives personnelles de chercheurs ou ingénieurs souhaitant communiquer sur leur activité professionnelle, par exemple dans le cadre de projets tels que "[1 scientifique, 1 classe : Chiche !](#)", lancé par le Ministère de l'Éducation Nationale et de la Jeunesse en 2020 ou encore les [Cordées de la réussite](#) portées par l'ONISEP.

Notre objectif est donc de constituer une bibliothèque numérique de contenus (présentations, projets pédagogiques, exemples de codes, etc.), de l'héberger et de la rendre visible et accessible.

3.8 Jeunes chercheurs

3.8.1 Offres d'emploi

Dans cette nouvelle période, le GDR Calcul va continuer son activité de diffusion des offres d'emploi (cf. section 2.1.2) car nous pensons que ce dispositif est particulièrement utile pour les jeunes en recherche d'emploi. Nous prévoyons d'améliorer le système d'annonce par courriel des nouvelles offres en les regroupant afin d'éviter un envoi électronique à chaque publication (le rythme moyen actuel est d'une offre tous les deux jours et demi).

3.8.2 Participation aux interventions

Par ailleurs, soucieux de faire bénéficier les jeunes chercheurs de son action, nous souhaitons inciter leur participation comme intervenants dans nos actions. Lors des journées "Calcul et apprentissage" (cf. section 2.2.10), nous avons eu une très bonne expérience avec l'intervention de deux doctorants dans les exposés et les TP.

3.8.3 Collection de notebooks Jupyter sur la plateforme PLMShift-GPU de Mathrice

PLMShift de Mathrice est une plateforme matérielle et logicielle permettant à tous les mathématiciens de partager des applications ou d'héberger des sites web. Il est déjà possible de partager une application Shiny (tableau de bord interactif construit en langage R) ou un notebook Jupyter qui est une feuille de calcul interactive permettant de partager du texte, des équations et du code informatique en langage Julia, Python ou R. Le groupe Calcul a participé au financement de cette infrastructure avec notamment l'achat de cartes GPU dédiées au calcul scientifique. Afin d'inciter à l'adoption de cette ressource de calcul, nous proposons de coordonner la réalisation de notebooks ou d'application Shiny en faisant participer particulièrement les jeunes chercheurs.

Ce projet aura trois objectifs :

- initier les jeunes chercheurs à faire de la science ouverte et reproductible permise par l'utilisation de ces outils,
- utiliser et promouvoir la plateforme PLMShift-GPU en collaboration avec Mathrice,
- proposer une démonstration attractive et accessible de l'utilisation du calcul scientifique pour la résolution numérique d'un ensemble de problèmes mathématiques.

Concrètement, le bureau lancera un appel à participer aux jeunes chercheurs et proposera son aide, en plus du support technique de Mathrice. Un site web sera construit avec une documentation et une galerie pointant vers les applications.

3.9 Organisation interne

3.9.1 Porteur du projet

Matthieu Boileau, ingénieur de recherche CNRS

matthieu.boileau@math.unistra.fr
IRMA, CNRS UMR7501, Université de Strasbourg
7 rue René Descartes
7084 Strasbourg Cedex
+33 (0) 3 68 85 01 73

3.9.2 Bureau

Le GDR Calcul se dote d'un bureau permettant de faire le lien avec le réseau Calcul et les autres GDR. Il est l'organe exécutif du GDR. Dans le cadre de ce renouvellement, nous proposons la composition suivante pour le bureau :

- Matthieu Boileau (CNRS), Institut de Recherche Mathématique Avancée, Strasbourg,
- Anne Cadiou (CNRS), Laboratoire de Mécanique des Fluides et d'Acoustique, Lyon,
- Roland Denis (CNRS), Institut Camille Jordan, Lyon,
- Benoît Fabrèges (Université Claude Bernard Lyon 1), Institut Camille Jordan, Lyon,
- Loïc Gouarin (CNRS), Centre de Mathématiques Appliquées, Palaiseau,
- Violaine Louvet (CNRS), UMS GRICAD, Grenoble,
- Pierre Navaro (CNRS), Institut de Recherche Mathématique de Rennes

3.9.3 Conseil scientifique

Le conseil scientifique est réuni au moins une fois par an afin de :

- faire le bilan des actions menées,
- définir de nouvelles actions thématiques à partir de discussions et de travail de brainstorming,
- s'organiser en groupes de travail pour mener ces actions avec l'appui du bureau.

Sollicités pour aborder les thématiques visées dans ce projet de GDR, les experts suivants ont accepté de rejoindre le prochain conseil scientifique (CS) :

- Mireille Bossy (INRIA, équipe CALISTO), INRIA Sophia Antipolis, *calcul stochastique*,
- Charles-Edouard Bréhier (CNRS), Institut Camille Jordan, Lyon, *calcul stochastique*,
- Emmanuel Franck (INRIA, projet Tonus), IRMA, Strasbourg, *apprentissage machine et interactions*,

- Luc Giraud (INRIA, équipe HiePACS), INRIA Bordeaux, *précision numérique*,
- Laura Grigori (INRIA, Alpine Group), INRIA Paris, *algorithmie parallèle*,
- Konrad Hinsen (CNRS), Centre de Biophysique Moléculaire, Orléans, *reproductibilité*,
- Fabienne Jézéquel (Université Panthéon-Assas), LIP6, Paris, *précision numérique*,
- Anthony Nouy (Centrale Nantes), Laboratoire de Mathématiques Jean Leray, Nantes, *apprentissage machine et calcul stochastique*,
- Christelle Piechurski (GENCI), Paris, *calcul quantique et apprentissage machine*,
- Nathalie Revol (INRIA, projet AriC), LIP, Lyon, *précision numérique*.

Note : parmi cette liste, Charles-Edouard Bréhier, Laura Grigori, Konrad Hinsen et Nathalie Revol font partie de l'actuel CS. En prolongeant leur participation, ils nous aideront à poursuivre des actions initiées lors de l'actuel GDR.

3.10 Partenariats

Depuis ses débuts, le GDR Calcul collabore régulièrement avec différents acteurs du calcul scientifique français. En particulier, le GDR a une forte proximité avec l'**INRIA**, comme le montre la représentation de cet organisme parmi les membres du conseil scientifique. **Genci**, la **Maison de la simulation** et l'**Amies** sont également des partenaires naturels dont les liens se sont faits par le biais d'actions communes ou par la participation au conseil scientifique. Enfin, la collaboration avec la **SMAI** est régulière : le GDR participe à la fois au financement et à l'organisation de mini-symposiums à l'occasion des congrés SMAI et des CANUM.

En ce qui concerne nos futures collaborations, notre programme scientifique nous conduit naturellement à nous rapprocher des GDR suivants :

- GDR IA, pour son activité dans l'apprentissage machine,
- GDR MaDICS, pour l'apprentissage machine et la science des données,
- GDR Mascot-Num, pour les aspects de calcul stochastique (Mireille Bossy et Anthony Noui en font partie),
- GDR MaNu, dans le domaine de la résolution des EDP.