L'informatique mathématique en France

Conseil Scientifique du GdR IM

12 juillet 2023

Membres du Conseil Scientifique du GdR IM — Pierrick Gaudry, Bruno Gaujal, Francis Lazarus, Frédéric Magniez, Assia Mahboubi, Jean-Yves Marion, Claire Mathieu, Myriam Preissmann, Bruno Salvy, Sophie Tison, Pascal Weil.

Les rédacteurs remercient celles et ceux qui leur ont apporté leurs lumières, et en premier lieu les responsables de GT qu'ils ont consulté.es et les anciennes membres du Conseil (Sophie Laplante, Stéphanie Delaune).

Introduction

Le périmètre des communautés scientifiques qui se réclament de l'informatique mathématique est le produit d'une évolution graduelle, qui n'a rien d'achevé.

Les ancêtres du GdR (le PRC puis le GdR Maths-Info par exemple) ont d'abord voulu fédérer des groupes d'informaticiens et de mathématiciens qui se reconnaissaient dans une certaine pratique de la recherche, à l'un des points d'articulation entre les deux disciplines. Dans cette pratique partagée, les problématiques ou les motivations sont issues de l'informatique, que ce soit par les objets étudiés ou par l'adoption d'un point de vue résolument calculatoire. Et les méthodes sont de nature profondément mathématiques, même si les mathématiques dites classiques ne fournissent que rarement les solutions recherchées, du fait de la nature des objets manipulés ou des questions posées. Pour citer les premières directrices du GdR, ceci mène *au développement de techniques mathématiques spécifiques, imprégnées du point de vue informatique. L'informatique mathématique est donc un domaine de l'informatique qui utilise non seulement des mathématiques, mais qui se révèle aussi créateur de nouvelles mathématiques.* [1]

Ce point de vue précurseur s'est révélé très fructueux et les évolutions profondes de la science informatique et des disciplines connexes ces dernières décennies font que de nombreuses communautés pourraient se retrouver aujourd'hui dans les objectifs initiaux du GdR. On pense par exemple, en informatique, aux domaines des réseaux, des bases de données, de la sécurité ; en mathématiques, au traitement du signal, à la théorie des nombres, à la géométrie, aux probabilités ; ou dans une liste toujours plus riches de domaines à l'interface de ces deux grandes disciplines, à l'optimisation, au traitement des données massives, à l'apprentissage, à l'intelligence artificielle, etc.

Ces évolutions, extrêmement positives, expliquent que les frontières entre l'informatique mathématique et les domaines couverts par d'autres GdR soient poreuses, et susceptibles d'évoluer.

Cependant, chacun de ces domaines se développe selon ses méthodes propres, et il en va de même de ceux qui rentrent dans le périmètre du GdR Informatique Mathématique. Le présent rapport, qui n'est pas un rapport d'activité du GdR, cherche à présenter sans souci d'exhaustivité l'évolution de nos thématiques de recherche, à l'interface entre informatique et mathématiques, afin d'aider les autres communautés des sciences informatiques et mathématiques à mieux situer nos domaines, nos contributions et leur pertinence, et les pistes qui se dessinent pour les années à venir.

Nous avons organisé ce texte en 8 sections qui chacune rassemblent plusieurs thématiques incarnées dans des GT (groupes de travail) du GdR :

- Logique informatique
- Calcul algébrique, arithmétique et cryptographie
- Géométrie(s) et image
- Calculabilité et complexité

1. Christiane Frougny et Brigitte Vallée, Dossier de création du GdR Informatique Mathématique, 2006.

- Calcul quantique
- Combinatoire(s), systèmes dynamiques, aléatoire
- Graphes et algorithmes
- Algorithmique

Ces sections sont complétées par de courtes discussions d'intérêt transverse sur les liens de l'informatique mathématique avec l'apprentissage et l'intelligence artificielle d'une part, avec l'industrie d'autre part. Une conclusion, encore plus courte, revient sur le périmètre scientifique de l'informatique mathématique, et évoque le rôle structurant du GdR dans les communautés qu'il fédère et quelques perspectives sur les contributions qu'il pourrait apporter au développement d'une informatique durable.

Quelques enseignements...

La lecture de ce rapport confirmera le large spectre scientifique couvert par le GdR. Chacune de ses sections a été rédigée en pensant à un lectorat qui dépasse les frontières du GdR. Elles comportent chacune une description synthétique des domaines scientifiques qui y sont évoqués, que nous espérons largement lisible ; une présentation rapide, nécessairement un peu plus technique, de certaines des avancées récentes les plus marquantes ; et une discussion des perspectives qui s'ouvrent à la communauté correspondante.

On comprend alors qu'il serait difficile, et peut-être pas très utile, de produire une synthèse de ces synthèses sans totalement diluer le propos. Nous nous risquons cependant à une réflexion sur les perspectives tracées dans chacune des sections qui suivent.

Ces perspectives peuvent être regroupées selon deux grandes tendances : la *convergence*, de méthodes, de concepts, d'outils, voire de disciplines qui échangent les unes avec les autres ; et l'*approfondissement* tant des concepts élaborés dans chaque champ que de leurs applications, lesquelles requièrent d'habitude aussi une bonne dose de convergence.

Ainsi, on repère en logique une convergence entre des branches de la logique et de la théorie des modèles auparavant plus distantes (logique linéaire, réalisabilité) et avec le vaste champ de l'apprentissage. Cela permet l'approfondissement des applications au développement des assistants de preuve (et à travers eux des applications dans de nombreux champs des mathématiques), au développement de langages et outils de programmation, et à la vérification qui peut maintenant s'attaquer à des problèmes de taille réaliste ainsi qu'à de nouveaux champs comme la vérification des algorithmes quantiques.

Le calcul formel s'appuie de plus en plus sur une convergence des méthodes numériques et des méthodes symboliques (les unes plus exactes, les autres plus rapides), et cherche aussi à progresser en approfondissant la compréhension de l'impact des structures spécifiques de certains domaines d'inputs. En cryptographie, la prise en compte de l'émergence du calcul quantique était déjà avancée mais elle continue à orienter les recherches, en même temps que la discipline prend en charge d'autres évolutions de son environnement, par exemple avec le développement d'outils pensés pour des systèmes sous contraintes de ressources (énergétiques, calculatoires) comme l'internet des objets.

L'évolution de l'arithmétique des ordinateurs s'inscrit dans le courant plus large de la recherche de reproductibilité, et approfondit ses apports au développement d'applications cryptographiques et à la recherche de garanties.

En géométrie, la convergence est frappante avec la topologie, dans une approche algorithmique et d'évaluation de la complexité qui va bien au-delà des notions plus classiques de constructivité ou d'effectivité ; avec l'algorithmique des données, notamment avec les données de grande dimension dont le volume s'accroît très rapidement avec la massification de la numérisation d'images ou d'objets 3D ; et avec la géométrie des maillages et l'analyse numérique. Dans le même temps, les communautés concernées approfondissent leurs méthodes, tirées par des domaines d'application comme l'impression 3D, l'imagerie médicale, la conception et le design d'objets, etc.

L'impact des notions fondamentales de calculabilité et de complexité se fait sentir toujours davantage dans un grand nombre de disciplines scientifiques qui vont bien au-delà des mathématiques (physique, biologie, sociologie, philosophie, etc). Cela amène les spécialistes à approfondir les défis posés par l'accès à de grandes masses de données ou par les nouveaux modèles de calcul comme le calcul quantique, sans négliger la dimension énergétique de la notion même de calcul.

L'informatique quantique est, par construction, le fruit d'une convergence entre physique fondamentale, photonique, informatique, etc. L'arrivée probable d'ordinateurs quantiques, même petits, va constituer un tournant qui forcera les quantistes, et sans doute beaucoup d'autres, à approfondir les interactions de ce mode de calcul avec l'algorithmique, la cryptographie, les codes correcteurs d'erreur, la programmation, la vérification, etc.

En combinatoire et dans l'étude des systèmes discrets, la convergence déjà connue de la combinatoire avec la biologie (bio-informatique) et avec la physique statistique continue à porter ses fruits, de même que celle des systèmes discrets avec la théorie des groupes et avec les probabilités. Les applications à l'algorithmique soulèvent de très nombreuses et difficiles questions à la combinatoire, de même que des questions fondamentales de complexité passent par un approfondissement des propriétés de certains systèmes dynamiques.

La théorie des graphes est, par nature, en relation avec de très nombreux champs disciplinaires, mais la convergence récente avec les domaines mathématiques de la topologie et de la géométrie est remarquable. Les différentes notions de complexité des graphes récemment introduites (les différentes largeurs arborescentes, twin, etc) mènent à un approfondissement des applications de la théorie des graphes à l'algorithmique et à la théorie de la complexité.

L'algorithmique enfin, transverse par essence, est le lieu de nombre des convergences évoquées ci-dessus. Elle est aussi au cœur de convergences avec des sujets contemporains brûlants, tant en matière environnementale (consommation énergétique du calcul, optimisation de l'usage des ressources) qu'en ce qui concerne des questions aussi fondamentales que celles d'éthique ou d'équité.

Ce rapide tour d'horizon se veut un encouragement à la lecture du reste du rapport, où sont explicitées plus concrètement les idées brossées ici à grands traits.

Logique informatique

Contours

La logique informatique recouvre l'étude des fondements mathématiques des concepts de calcul et de preuve, ainsi que l'étude et la conception de langages formels. Ces derniers sont ultimement utilisés pour exprimer des connaissances, des propriétés de structures ou de systèmes. Le développement d'algorithmes pour assurer que ces propriétés sont vérifiées, ou encore pour interroger des données, sont partie intégrante de cette thématique et y occupent une place importante.

Une facette importante de la logique informatique traite des questions mathématiques liées à la représentation, la mise en œuvre et l'étude des propriétés du calcul. La plupart de ces questions relèvent en fait de champs mathématiques plus vastes, comme la théorie de la démonstration, la théorie des jeux, celle des catégories, des co-algèbres, etc. ... À leur tour, ces champs motivent des connexions naturelles avec des thématiques moins intrinsèquement liées au calcul comme la topologie algébrique, les probabilités, ... La perspective offerte par la logique sur le calcul est intimement liée à la théorie des langages de programmation, à laquelle elle apporte des formalismes comme les systèmes de types, la sémantique, la réécriture ou encore les machines abstraites, mais aussi des outils comme les assistants de preuve ou les logiciels de preuve automatique. Comme dans d'autres champs de l'informatique théorique, les catégories offrent un langage commun qui permet de construire des liens féconds avec d'autres disciplines, en particulier la topologie. Le groupe de travail (GT) « Logique, Homotopie, Catégories » (LHC), s'intéresse précisément à ces ponts, et inclut des membres qui ne se définissent pas comme informaticiens et viennent plutôt des communautés de la topologie algébrique, des catégories d'ordre supérieur, de la physique mathématique, etc. Le groupe de travail « Structures formelles pour le CALcul et les Preuves » (SCALP) quant à lui s'intéresse plus spécifiquement à la logique informatique, dans une large acception.

La conception et l'étude de langages formels constituent une autre facette importante de la logique informatique. Une problématique centrale de cette thématique est celle de la mise au point de langages pour la modélisation de systèmes informatiques et de bases de connaissances. Cette modélisation permet en retour d'analyser et de certifier des systèmes informatiques, en particulier lorsque ceux-ci sont considérés comme « critiques », c'est-à-dire que leur bon fonctionnement est crucial d'un point de vue économique, médical ou sociétal : des exemples notoires relèvent des systèmes embarqués (avions, satellites, fusées, automobiles), des systèmes répartis ou mobiles, des protocoles de sécurité (paiement sécurisé, vote électronique, cryptomonnaies), etc. Au delà de la vérification, la modélisation formelle des systèmes permet également de s'intéresser au problème de leur synthèse : l'idée est alors de construire de manière automatique un système (un contrôleur, un programme informatique) à partir de sa spécification. Les techniques d'analyse et de modélisation formelles des systèmes sont diverses, et permettent l'analyse de ces systèmes à différents niveaux, qui vont de la spécification abstraite au code. Elles reposent sur des fondements variés comme les automates, la logique ou les jeux, ainsi que sur des techniques d'abstraction, d'interprétation abstraite, de réécriture, et sur l'utilisa-

tion d'outils de preuve automatique, typiquement de satisfiabilité (SAT) et satisfiabilité modulo théorie (SMT). Ces thèmes de recherche se retrouvent principalement dans les groupes de travail « Data, Automata, Algebra, & Logic » (DAAL), et « Vérification ». L'étude des automates et de leurs applications occupe une place centrale dans le groupe de travail DAAL, dont le périmètre inclut aussi l'étude des fondements des bases de données. Les thématiques motrices du groupe de travail « Vérification » sont celles de la modélisation, de l'analyse et de la certification.

Par ailleurs, la logique informatique a des connexions naturelles avec des thématiques structurées par d'autres GdR. On observe ainsi des liens assez forts avec les GdR Top (en particulier pour le GT LHC), GPL (en particulier son GT LTP), Jeux, et le GdR Sécurité Informatique à travers en particulier son GT « Méthodes formelles pour la sécurité » créé en 2017. Ce dernier a vocation à rassembler plus largement autour de la sécurité, alors que la communauté Informatique Mathématique est plus en pointe sur les approches formelles.

Avancées récentes

Ces dernières années, les techniques de vérification se sont beaucoup diversifiées avec la prise en compte d'aspects temporels et probabilistes, et plus généralement le passage d'une réponse booléenne (est-ce que le système peut atteindre un cas d'échec ?) à une réponse quantitative (avec quelle probabilité le système peut-il atteindre un cas d'échec ?). La vérification paramétrée a également connu un regain d'intérêt. L'idée est de vérifier une famille de systèmes, comme par exemple un système de vote dont le nombre de votants serait un paramètre. La vérification paramétrée permet alors d'établir des propriétés indépendamment de la valeur du paramètre, et donc dans cet exemple du nombre de votants. Les problématiques actuelles en sémantique des langages de programmation suivent une évolution similaire, et cherchent à modéliser de façon pertinente l'*équivalence probabiliste* de deux programmes, c'est-à-dire, les situations où ils se comportent de la même façon, non plus *toujours* mais *avec une probabilité* suffisamment élevée. L'étude des langages de bas niveau, avec prise en compte des ressources et de la concurrence, sous l'angle de la théorie mathématique des langages de programmation, est également un sujet très actif aujourd'hui.

Parmi les avancées récentes obtenues par la communauté, on peut citer la série de travaux sur la complexité du problème de l'accessibilité dans les réseaux de Petri. Les réseaux de Petri constituent un modèle très utilisé en vérification ainsi que dans d'autres domaines, en fait l'un des plus simples que l'on puisse considérer. Par exemple, la détection d'états indésirables dans un système se formule naturellement en terme d'accessibilité dans le réseau associé. La décidabilité du problème d'accessibilité est connue depuis le début des années 80, mais sa complexité était restée inconnue depuis. Les avancées récentes obtenues en théorie de la complexité pour des classes très élevées, ont permis d'obtenir début 2020 des bornes inférieures et supérieures pour ce problème. Il s'agit d'une première brèche très significative, mais d'importantes questions reliées restent encore ouvertes à ce jour.

La théorie des automates quant à elle, entretient des liens toujours plus forts avec

la théorie des jeux. Par exemple, le problème de la synthèse (construire un contrôleur à partir de sa spécification) a pu être formulé dans la théorie des jeux.

Les liens de la logique informatique avec la théorie des bases de données sont toujours plus nombreux. Historiquement présente en France depuis de nombreuses années, la communauté des bases de données est toujours en expansion, et ce domaine est très actif au niveau international. Alors qu'à l'origine, la théorie des bases de données se fondait uniquement sur la formalisation des bases de données relationnelles en tant que structures logiques, les données semi-structurées, représentées sous forme d'arbre ou de graphe (à l'origine des standards XML et RDF pour le Web sémantique), posent de nouveaux défis fondamentaux, où les automates sont des outils de premier plan. L'importance et l'évolution des systèmes de gestion de données ont aussi renforcé l'intérêt de notions comme la gestion de l'incertitude et des informations incomplètes, la prise en compte des connaissances, l'explication des résultats et leur certification, l'évaluation en streaming, l'analyse fine de la complexité de l'évaluation d'une requête et de l'énumération des résultats... Par exemple, de nombreux travaux récents traitent de la provenance qui permet d'expliquer les résultats d'une requête, que ce soit pour en fournir une certification grâce à une formalisation en Coq, pour la représenter de façon compacte en utilisant les circuits, pour l'étendre à Datalog et aux bases de données graphe, pour analyser la complexité de son calcul en utilisant des notions comme la largeur d'arbre. Cette tendance a d'ailleurs motivé le renommage du GT ALGA en GT DAAL afin de permettre une meilleure représentation de cette communauté. On peut noter l'implication de chercheurs de la communauté du GDR dans la standardisation de GQL, 'Graph Query Language'.

En 2011, des travaux originaux autour de l'algorithme de décision de Hopcroft et Karp, apportaient une belle illustration du succès de l'informatique mathématique. Dans ces travaux, le point de vue coalgébrique sur un problème d'automates très étudié depuis une quarantaine d'année avait permis de proposer un algorithme complètement nouveau, significativement plus efficace que l'état de l'art. Ce succès a stimulé la confluence des centres d'intérêts des GT DAAL et SCALP. Cette convergence s'est depuis significativement renforcée, trouvant des applications majeures par exemple dans l'étude des sémantiques de programmes, et fertilisant d'autres domaines fondamentaux comme la théorie de la démonstration. On ne peut qu'espérer voir se resserrer toujours plus les liens fructueux entre la théorie de la programmation (avec ses outils catégoriques, algébriques, topologiques, etc.) et celles de la vérification, des automates et de la complexité.

Dans le périmètre du GT SCALP, des résultats récents ont fait avancer les problématiques intrinsèques au domaine. On peut mentionner par exemple la complétion du programme d'étude de la complexité des machines abstraites (machine de Krivine, machine à jetons de la géométrie de l'interaction, etc.), en temps et en espace, achevé en 2022. On trouve aussi dans ce cadre des interactions fécondes avec d'autres thèmes de l'informatique fondamentale. L'étude du concept de différenciation dans les langages offre un bel exemple de confluence entre théorie des langages de programmation et théorie de l'apprentissage (descente de gradient, propagation arrière, etc.).

La théorie homotopique des types (HoTT) et les fondations univalentes ont été in-

troduites par Voevodsky vers 2011. Ce point de vue extrêmement novateur a jeté un pont entre topologie algébrique, théorie des types et fondements des mathématiques et a stimulé les interactions entre sémantique, réécriture de dimension supérieure et ∞-catégories. Depuis, l'étude du contenu effectif de ces fondations univalentes a permis de donner une version calculatoire de HoTT, avec la mise au point de théories des types dites cubiques. Le développement d'assistants de preuve, qui permettent de mener à bien en pratique la formalisation de preuves, est en plein essor. Nourri par une recherche fondamentale très active en théorie des types, cette activité de développement sous-tend des travaux de vérification formelle, en particulier la vérification de programmes et de propriétés de langages de programmation.

Prospective

On assiste à une *convergence des concepts, des méthodes et des systèmes* (vérification, sémantique, preuve,...). En conséquence, l'analyse de langages réalistes devient accessible et *la certification de logiciel à grande échelle* va devenir possible, ce qui constitue une évolution considérable.

La logique linéaire et la réalisabilité sont par exemple tout à la fois des domaines très actifs de recherche en eux-mêmes, et la source d'outils qui jouent un rôle majeur dans la conception de langages de programmation modernes, par exemple pour l'analyse fine des effets et la gestion des ressources. L'étude fine de la notion de ressource en logique (logique linéaire) a également amené à formaliser la notion de différentiation dans ce cadre (logique linéaire différentielle). Au vu des avancées récentes, on peut penser que les interactions avec le domaine de l'apprentissage se consolideront, avec peut-être l'émergence de nouveaux langages programmation.

Les automates sont un objet fondamental qui permet souvent de formaliser et résoudre des problèmes issus de nouveaux contextes. Par exemple, les automates ont joué et continuent à jouer un rôle important dans des domaines comme la théorie des jeux, et la théorie des bases de données. Les prochaines années pourraient voir émerger des liens forts entre théorie des automates et apprentissage, et peut-être permettre de mieux comprendre des techniques déjà largement développées comme les réseaux de neurones. Plus généralement, l'application des techniques de vérification aux algorithmes utilisés en IA pourrait permettre de mieux les expliquer et par suite de garantir leur bon fonctionnement. En retour, l'apprentissage a également un rôle à jouer pour améliorer les performances des techniques utilisées en vérification formelle. L'articulation d'outils d'IA générative et d'outils de vérification pourrait également permettre la conception d'outils de génération de programmes tout à la fois performants et fiables.

Plus généralement, le domaine des méthodes formelles, et en particulier celui de la vérification, est aujourd'hui confronté à un panel varié de nouvelles applications aux enjeux sociétaux importants, qui le nourrissent de nouvel. En particulier, l'essor des cryptomonnaies et de la blockchain a suscité un regain d'activité autour de l'étude formelle des algorithmes distribués et de leur implantation. La vérification de programmes quantiques, ainsi que la sécurité post-quantique fournissent un autre exemple de telles applications, et induisent la conception de nouveaux outils de vérification formelle.

Pour finir, on mentionnera les nouvelles perspectives d'utilisation des assistants de preuves, outils qui permettent de concevoir des bibliothèques de résultats mathématiques formalisés et vérifiés. À quelques notables mais rares exceptions près, les applications significatives de ces outils concernaient jusqu'à maintenant essentiellement des problèmes de vérification de programmes (compilateur, micro-noyau, protocoles cryptographiques, etc.), pour des utilisateurs avec une bonne culture informatique. Une nouvelle communauté d'utilisateurs d'assistants de preuve constituée de chercheurs en mathématiques fondamentales a récemment pris son essor. Leur activité se concentre sur la formalisation de résultats mathématiques contemporains, voire même en cours d'élaboration. Signe des temps, la preuve formelle a fait l'objet de deux exposés invités à l'International Congress of Mathematicians (ICM) 2022, dont un plénier. Le choix de la théorie des types (dépendants) comme fondement logique s'est avéré crucial dans ce succès. La communauté française compte plusieurs experts mondiaux de cette famille de formalismes et a de fait conçu la variante actuellement utilisée par des assistants de preuve comme Lean, Coq ou Agda, et donc en particulier pour ces applications mathématiques. Les assistants de preuve et leurs bibliothèques vont devoir évoluer pour s'adapter aux besoins et questions suscitées par ces nouveaux usages, liées aussi bien à leur implémentation qu'à leurs fondements logiques.

Calcul algébrique, arithmétique et cryptographie

Contours

La manipulation informatique des objets fondamentaux de l'arithmétique et de l'algèbre a une place importante au sein du GdR. Cela concerne les nombres entiers ou flottants, les polynômes à une ou plusieurs variables, les séries, les matrices, les solutions d'équations différentielles, etc. Selon le contexte, les techniques et les communautés peuvent être différentes, mais l'ensemble forme des domaines de recherche aux frontières poreuses. L'arithmétique des ordinateurs est plus centrée sur des objets élémentaires, mais avec des contraintes très fortes (garanties sur les arrondis, vitesse d'exécution, taille des circuits, etc). Le calcul formel prend le relai lorsqu'il s'agit de structures algébriques plus complexes ou de taille plus importante, avec des problématiques de complexité asymptotique, de calculabilité et décidabilité en mathématiques, ou d'effectivisation de résultats jusqu'alors purement existentiels. Beaucoup d'efforts ont porté sur le développement d'applications en direction de la cryptographie, la cryptanalyse multivariée, les codes correcteurs d'erreurs, la combinatoire, les preuves formelles, le calcul fiable, la théorie du contrôle, l'automatique, la modélisation algébrique en lien avec la biologie, et la modélisation géométrique par des primitives algébriques. Le GdR IM a contribué significativement aux développements de ces interactions et transferts entre disciplines.

La cryptographie se positionne à la fois comme utilisatrice et comme demandeuse d'algorithmes plus spécifiques. C'est particulièrement vrai en cryptographie à clef publique où, au-delà des traditionnels calculs modulaires de RSA, on fait désormais appel à des objets plus avancés comme les courbes elliptiques ou les réseaux euclidiens. La théorie des codes correcteurs, et notamment les codes ayant une forte structure algébrique, est traditionnellement proche de la cryptographie. Pour ces deux domaines, le calcul quantique a pris une importance considérable ces dernières années. Il remet en cause la sécurité d'une partie de la cryptographie actuellement déployée, tout en offrant de nouvelles perspectives. De plus, ce nouveau modèle de calcul nécessite, de manière fondamentale, des codes correcteurs adaptés. Les interactions avec le groupe de travail Informatique Quantique sont donc naturellement très fortes, et nous renvoyons à la section correspondante de ce rapport.

Que ce soit en arithmétique des ordinateurs, en calcul formel ou en cryptographie, une préoccupation croissante est de garantir que les algorithmes et leurs implémentations sont corrects. On trouve donc de nombreux liens avec les activités de preuve formelle et de vérification du GdR. Par exemple, des outils comme l'assistant de preuve Coq sont utilisés dans bien des travaux. En cryptographie, on trouve des travaux cherchant à prouver grâce à ce type d'outils qu'une construction de chiffrement, de signature, d'échange de clés a bien les propriétés attendues, et ce de manière vérifiable par une machine. Lorsque les preuves concernent des protocoles plus compliqués, cela relève plutôt du GdR Sécurité. Mais la frontière est floue, notamment lorsqu'il s'agit de prouver du logiciel en plus de la spécification. Les liens du calcul formel avec la partie combinatoire du GdR sont également très forts. Ainsi, le calcul formel fournit des outils pour l'expérimentation, notamment en combinatoire algébrique et en combinatoire énumérative, où ils sont utili-

sés pour classifier des modèles combinatoires (rationnels, algébriques, différentiellement finis, différentiellement algébriques,...). À l'inverse, combinatoire énumérative et analytique suscitent le développement de nouveaux algorithmes de calcul formel, tant pour l'étude des récurrences que pour l'évaluation du comportement asymptotique de suites.

Les développements logiciels constituent une partie importante de l'activité de recherche dans ces domaines. Ils tissent des liens entre les équipes en France et à l'international. Les logiciels sont nécessaires pour valider les résultats de complexité théoriques. Ils sont aussi largement utilisés pour des applications pratiques, en lien avec d'autres disciplines et des industriels.

Avancées récentes

Les principaux axes de recherche en calcul formel continuent à être organisés autour de problèmes théoriques fondamentaux en mathématiques effectives, de problèmes d'efficacité d'algorithmes de base, et d'algorithmes spécialisés pour certaines applications. Encore maintenant, des problèmes fondamentaux continuent de progresser, notamment sur la complexité des opérations de base, comme le produit des entiers, le produit des matrices, le déterminant, la composition modulaire, le résultant dit « bivarié », etc. Plusieurs percées majeures sur ces sujets ont été menées par des équipes françaises durant ces dernières années. Concernant les logiciels, on peut noter une certaine maturité des systèmes de calcul symbolique (SageMath, Pari-GP, Mathemagix, etc.), traduite par un interfaçage accru avec de nombreuses bibliothèques spécialisées. Ce travail de fond sur lequel la communauté du GdR a été très active porte ses fruits, et les utilisateurs en profitent au jour le jour.

Des résultats majeurs ont été obtenus dans la plupart des domaines de la cryptographie. La communauté du GdR a été très réactive lors des appels à standardisation du NIST de ces dernières années (cryptographie à bas coût, cryptographie résistant à l'ordinateur quantique), ainsi que pour analyser les nombreuses soumissions à ces appels. Au-delà de ce contexte, l'étude des fondements de la théorie algorithmique des nombres sur lesquels s'appuient la plupart des algorithmes de la cryptographie asymétrique (factorisation, réseaux euclidiens, logarithme discret, systèmes polynomiaux, isogénies, codes correcteurs), la cryptographie symétrique (conception et cryptanalyse, classique ou quantique) ainsi que la conception de primitives avancées pouvant servir dans les contextes décentralisés (cryptographie multi-parties, cryptographie fonctionnelle) sont autant de points forts de la communauté française.

En arithmétique des ordinateurs, les avancées de ces dernières années concernent à la fois la prise en compte, voire la conception de nouvelles architectures, et les contraintes de nouvelles applications. Par exemple, tirer partie de la puissance des processeurs récents nécessite d'utiliser les instructions vectorielles (comme l'AVX), qui offrent un parallélisme de données au plus bas niveau. De même, ces dernières années ont été marquées par le développement important de travaux sur l'architecture matérielle libre et ouverte RISC V, un modèle de processeurs très modulaires pouvant aller de l'embarqué jusqu'aux supercalculateurs. Un autre virage récent est le retour vers la faible précision. Sous l'impulsion des énormes besoins de calcul liés à l'intelligence artificielle, le besoin d'opérateurs

matériels sur des nombres représentés typiquement sur 16 bits est immense ; les questions de gestion des arrondis ou des dépassements n'ont pas les mêmes réponses dans ce cadre que lorsque l'on travaille en grande précision.

Prospective

En calcul formel, l'idée de combiner des méthodes numériques (pour aller vite) et des méthodes symboliques (pour donner un résultat garanti), qui n'est pas nouvelle, reste un thème fort qui va continuer à se développer. Un autre thème qui a pris de l'ampleur, mais demande encore du travail est la prise en compte des structures spécifiques des données d'entrée (parcimonie, symétrie, etc). Tout cela prend des formes diverses selon les algorithmes concernés et les besoins applicatifs, et se fait donc en interaction avec d'autres disciplines, internes ou externes au GdR (biologie, robotique, optimisation, etc). Par ailleurs, les liens avec le calcul haute performance devraient se développer et donc contribuer à cette ouverture vers plus d'applications.

Pour ce qui concerne la cryptographie, il s'agira de poursuivre le fort positionnement sur les sujets chauds de ces dernières années. L'ordinateur quantique étant une perspective de plus en plus réaliste, on doit considérer que le travail sur la conception d'une boite à outils complète de cryptographie résistante au quantique n'en est qu'à ses débuts. L'autre besoin prégnant est la cryptographie pour des systèmes liés à l'Internet des objets, pour lesquels les contraintes de taille ou de consommation sont telles que des algorithmiques spécifiques se révèlent souvent nécessaires. Une des forces de la communauté française est la diversité de sa culture et la continuité de travaux fondamentaux, qui se déploient parfois loin des projecteurs mais qui peuvent se révéler cruciaux lors d'évolutions futures. Ceci doit être préservé, car l'histoire récente montre que c'est ce qui a permis de s'adapter rapidement à des thématiques en émergence [2].

Sur le thème de l'arithmétique des ordinateurs, les travaux sur des algorithmes et des architectures avec des précisions variées, y compris faibles, ont vocation à se poursuivre, en lien avec des applications spécifiques. La reproductibilité des résultats est un sujet qui a émergé ces dernières années. Il nécessite d'avoir des briques de base arithmétiques au comportement bien spécifié, ce qui ne va pas de soi lorsqu'on ajoute des contraintes fortes d'efficacité. Un domaine traditionnellement gourmand en calculs arithmétiques est la cryptographie ; la montée en puissance de systèmes post-quantiques révèle de nouveaux besoins que les arithméticiens doivent satisfaire. Enfin, le désir de garanties, voire de preuves formelles, sur l'exactitude des résultats ne fait qu'augmenter et la communauté du GdR est armée pour y répondre.

2. On peut citer par exemple les cryptographies à base de codes correcteurs ou de systèmes polynomiaux, redevenues à la mode car résistantes à un éventuel ordinateur quantique.

Géométrie(s) et image

Contours

Le traitement de la géométrie en informatique concerne l'étude des formes discrètes et combinatoires, que celles-ci proviennent d'objets abstraits comme les graphes ou les complexes simpliciaux, ou de données réelles telles qu'un nuage de points échantillonnés à la surface d'un crâne ou d'une carrosserie. Il s'agit de comprendre, analyser, traiter, ou produire des formes géométriques. Sur le plan applicatif, l'imagerie numérique est une cible majeure que l'on retrouve dans les applications industrielles pour le loisir (jeux vidéos, animation, photographie,...), la médecine (chirurgie assistée), ou encore le design (Conception Assistée par Ordinateur), voire l'apprentissage en milieu artificiel (via les simulateurs). Parallèlement, des données discrètes sont produites massivement tous les jours, très souvent liées à des informations géométriques, et il est nécessaire d'avoir des outils pour les traiter, les analyser, les indexer, ou même en fabriquer de nouvelles. Trois thématiques relèvent de la géométrie ou de l'image au sein du GdR IM.

La *géométrie digitale*, ou *discrète*, part du principe que le codage des données géométriques conduit à considérer des objets de nature discrète. Elle a donc pour principal objet d'étude les sous-ensembles *discrets* définis sur des réseaux (sous-ensembles additifs discrets de $\mathbb{R}^n$) et tout particulièrement sur $\mathbb{Z}^n$. Les recherches portent sur la représentation, les transformations (affines par exemple), l'analyse ou la synthèse d'objets discrets. On s'intéresse aussi à l'étude de problèmes inverses, en particulier en tomographie discrète où un objet digital est reconstruit à partir de ses projections. Un autre axe de recherche porte sur la *morphologie mathématique*[3] qui désigne une théorie initialement motivée par des applications industrielles et qui repose principalement sur la notion de treillis, permettant de définir dans un cadre algébrique les opérations duales d'érosion/dilatation et d'ouverture/fermeture.

La *géométrie algorithmique* est motivée par des problèmes de nature géométrique dont la réponse peut être apportée par un algorithme. Typiquement, si l'on considère une subdivision de l'espace induite par un arrangement d'objets géométriques tels que des hyperplans dans l'espace, une requête de localisation consiste à déterminer la cellule de la subdivision contenant un hypothétique point de requête. L'objet de la géométrie algorithmique est de développer des algorithmes pour répondre efficacement à ce type de requête. Plus généralement, la géométrie algorithmique s'intéresse à la compréhension et à la quantification de la géométrie ou topologie des formes discrètes ou combinatoires. La nature discrète de ces formes rend leurs caractéristiques potentiellement calculables et justifie une approche algorithmique (constructive) pour les évaluer, voire pour caractériser leur complexité intrinsèque. Ces formes discrètes peuvent elles-mêmes représenter une approximation de formes hypothétiques continues, de sorte que le champ de la géométrie algorithmique déborde sur les mathématiques appliquées. La théorie de la persistance homologique est ainsi née de l'analyse topologique des données, afin d'extrapoler une topologie à partir de données discrètes supposées échantillonner un espace continu. Lorsque

3. Le GT de géométrie discrète a élargi son périmètre en direction de la morphologie mathématique au printemps 2017 pour devenir GT Géométrie Discrète et Morphologie Mathématique (GDMM).

cet espace est lui-même plongé dans un espace de dimension beaucoup plus grande des difficultés algorithmiques et pratiques apparaissent et une approche spécifique liée aux grandes dimensions doit être développée. À l'autre bout du spectre, les questions propres à la topologie de basse dimension font surgir des questions algorithmiques, aussi bien pour explorer de manière efficace des objets de combinatoires complexes, comme les triangulations de variétés de dimension 3, que pour caractériser ces objets, par exemple en considérant la largeur arborescente (treewidth) de ces mêmes triangulations. Par ailleurs, une attention particulière est portée aux approches probabilistes pour la conception et l'analyse des algorithmes ou pour la modélisation des données. Le GT « Géométrie Algorithmique » entretient ainsi des liens forts avec les communautés de combinatoire, de théorie des graphes, d'algorithmique, et avec le GT « Aléa ».

Enfin, la *modélisation géométrique* et plus généralement, le traitement des données géométriques (*Geometry Processing*), s'intéresse à des méthodes numériques et algorithmiques de modélisation, représentation, analyse et acquisition de formes géométriques 3D. Les activités portent sur le développement de modèles géométriques intégrant des contraintes géométriques, topologiques, ou physiques, et permettant d'ajouter de la sémantique dans les modèles. Des outils de modélisation et reconstruction sont proposés pour rendre la création de modèles 3D intuitive et y intégrer des connaissances provenant du domaine d'application. La communauté développe également des outils pour réparer des maillages, les assembler ou les idéaliser. Un contexte d'application important pour ces modèles est la réalité virtuelle qui nécessite de nouvelles interfaces et de nouvelles manières de concevoir et d'interagir avec les modèles.

Avancées récentes

Dans la communauté de géométrie algorithmique, on note ces dernières années une nette inflexion vers des questions de nature topologique. Ceci est d'ailleurs confirmé au niveau international par le nombre croissant de manifestations (conférences, workshops ou écoles) autour de la topologie algorithmique ou de l'analyse topologique des données [4]. L'analyse topologique des données désigne essentiellement la théorie de la persistance homologique et l'un des défis actuels dans ce domaine est la compréhension des modules de persistance multivariés permettant d'analyser les filtrations à plusieurs paramètres. Les aspects algorithmiques intéressent quant à eux les topologues de basse dimension pour ce qui concerne les courbes sur les surfaces, les nœuds, les variétés de dimension trois, les espaces de modules, etc. En effet, les avancées mathématiques sur ces objets permettent de poser des questions combinatoires plus précises, comme le dénombrement des classes d'homotopie de longueur bornée sur une surface hyperbolique, ou la reconnaissance algorithmique de la trivialité d'un diagramme de nœud. Ces questions combinatoires demandent à être validées ou explorées par les méthodes de l'informatique et on assiste ainsi à une interaction de plus en plus forte, et mutuellement intéressée, entre topologues et informaticiens. Sur les aspects plus géométriques, des outils jusqu'alors cantonnés aux

4. On pourra consulter la page `people.clas.ufl.edu/peterbubenik/conferences` pour une liste non-exhaustive de manifestations.

mathématiques pures pénètrent le champ combinatoire et permettent de distinguer des géométries particulières comme les espaces CAT(0) ou les espaces hyperboliques au sens de Gromov et d'adapter en conséquence les algorithmes et analyses. En parallèle de ces aspects purement topologiques et géométriques, on assiste à la prise en compte de l'aléatoire pour proposer une analyse plus fine des algorithmes face à des données réelles satisfaisant des distributions données a priori. Cet aléatoire se retrouve dans l'analyse des données où l'incertitude et le bruit sont modélisés par des mesures de probabilité et une notion appropriée de distance permet d'étendre les techniques d'inférence géométrique à des données incertaines.

Dans le domaine de la modélisation géométrique les problématiques fondamentales concernent l'analyse de données géométriques intra-modèles (point d'intérêt 3D, structure des modèles) et inter-modèles, la mise en correspondance, l'analyse de familles de formes ou la classification par la sémantique. Du coté applicatif, plusieurs avancées font évoluer le domaine : un grand changement de ces dernières années est la $3D$ *pour tous* (démocratisation de la modélisation $3D$), qui permet la création et l'utilisation de contenus numériques $3D$ avec des surfaces de subdivisions qui rendent la création de formes intuitive et rapide. Ces surfaces de subdivision sont désormais intégrées dans les logiciels de modélisation. Par ailleurs, les outils de numérisation et les imprimantes $3D$ amènent de nouveaux besoins et de nouvelles problématiques. Enfin, la réalité virtuelle ou augmentée, notamment dans le domaine médical ouvre de nouveaux défis.

Il est à noter que, en dehors des liens avec les mathématiques, la recherche en modélisation géométrique ne peut se faire indépendamment des domaines d'application, qu'il s'agisse de la mécanique, de la biomécanique, du multimédia/systèmes (transmission de contenus 3D) ou du traitement du signal (compression de maillages 3D).

En géométrie discrète, les travaux fondamentaux sur les objets de base (segments, plans, arcs de cercle) ont permis la définition d'estimateurs différentiels vérifiant la propriété de convergence multi-grille. Ces avancées ont ouvert la voie à la discrétisation d'opérateurs différentiels d'ordre supérieur, par le biais notamment du calcul extérieur discret. Par ailleurs, le problème de la reconnaissance d'objets s'est déplacé ces dernières années vers des objets plus complexes que ces objets de base, les solutions apportées faisant souvent le lien avec des objets issus de la géométrie algorithmique (triangulation, diagramme de puissance, polytopes, etc). En topologie discrète et en morphologie mathématique, des avancées récentes permettent de généraliser des résultats théoriques/algorithmiques en dimension spatiale aussi bien que spectrale, à des modèles hiérarchiques d'images : passage de modèles sur grilles régulières à des modèles sur graphes généraux, qui permettent de prendre en compte une plus grande variété d'espaces de représentation comme par exemple celui des formes et des lignes de niveau d'une image ; morphologie mathématique «couleur» ou sur des ensembles de valeurs partiellement ordonnées comme c'est par exemple le cas pour les images de tenseurs disponibles en IRM ou pour les images hyperspectrales disponibles en télédetection. Les évolutions récentes montrent de nombreux et fructueux rapprochements entre les communautés à l'interface entre mathématiques et informatique. Parmi ces évolutions on notera des liens avec l'analyse numérique pour traduire des EDP (équations aux dérivées partielles) sur des

structures combinatoires ; des liens avec la topologie et la géométrie des données pour obtenir des garanties théoriques de stabilité (via la persistance homologique, ou la convergence multi-grille d'estimateurs différentiels) ; des liens avec l'optimisation combinatoire via la minimisation de nouvelles formulations convexes pour l'analyse d'images ; ou encore des liens avec la combinatoire des mots, l'arithmétique, les systèmes dynamiques (structures quasi-périodiques : pavages, plans ou quasi-cristaux). D'anciennes conjectures liées aux pavages ont été résolues, le rôle des fractions continues multi-dimensionnelles est de mieux en mieux compris.

La production de logiciels open source est un souci commun aux trois communautés de géométrie. Ainsi le GT « Géométrie Discrète » est le principal instigateur de la bibliothèque libre et collaborative `C++ DGtal` (*Best software award* à la conférence SGP 2016). La communauté française de géométrie algorithmique est le principal contributeur au projet collaboratif de bibliothèque `C++ CGAL`. Enfin, la communauté de traitement géométrique des données propose des plateformes comme `Geogram`, ou `Graphite` (prix spécial *most innovative project* et troisième prix *scientific software category* aux *Trophées du libre*).

Prospective

En géométrie algorithmique, les interactions avec la topologie, discipline traditionnellement réservée aux mathématiques, sont de plus en plus fortes. Les collaborations ou conférences regroupant géomètres algorithmiciens et topologues deviennent communes à mesure que les topologues s'emparent de questions algorithmiques et, inversement, que les algorithmiciens s'attaquent à des problèmes de topologie. Ainsi, on sait désormais (Hass 1999, Lackenby 2016) que le problème de la reconnaissance du nœud trivial est dans $NP \cap coNP$, ce qui le place parmi les rares problèmes identifiés dans cette classe pour lesquels aucun algorithme polynomial n'est connu. Parmi les pistes prometteuses dans ce domaine on peut penser à l'usage du paramètre de largeur arborescente pour quantifier la complexité combinatoire d'objets topologiques, comme les triangulations de 3-variétés ou les diagrammes de noeuds, ou au codage des objets via les "straight-line programs" également utiles en théorie combinatoire des groupes.

Les questions de plongement métrique interviennent par ailleurs pour les applications à l'algorithmique des données de grande dimension, où il est crucial de réduire la dimension tout en préservant au mieux les propriétés métriques. Plus généralement l'existence de plongements (métriques ou non) est une source de problèmes difficiles et importants (un cas particulier étant le problème de l'isomorphisme de sous-graphes).

L'explosion du domaine de l'impression $3D$, avec de nombreuses applications industrielles potentielles, devrait fortement impacter tous les domaines de la géométrie (modélisation géométrique, géométrie discrète, géométrie algorithmique). C'est particulièrement vrai dans le domaine de l'imagerie médicale, dont le développement tous azimuts (réalité augmentée, planning d'opération, simulation...) est un véritable enjeu sanitaire mais aussi commercial.

Un axe de développement important est le rapprochement avec les communautés *digital geometry* internationales, où ce terme désigne davantage la géométrie sur les maillages,

nuages de points ou données non-structurées. La géométrie discrète peut jouer un rôle important pour faire se rejoindre analyse numérique standard et calcul discret. D'autre part, le rapprochement avec la communauté Morphologie Mathématique, au niveau du GdR mais aussi au niveau international, devrait resserrer les liens avec les problématiques du traitement d'image. L'implication dans des projets autour de la reproductibilité, comme le journal IPOL [5], va aussi dans ce sens.

En modélisation géométrique, certains thèmes émergent ces dernières années : la création intuitive de contenu numérique 3D, le design virtuel d'objets destinés à être fabriqués, la modélisation par apprentissage statistique, la représentation par des grammaires de formes ou de processus de construction. À ceci s'ajoutent les défis posés par l'impression 3D. La numérisation en masse pose par ailleurs la question de l'apprentissage et de la manipulation de grandes masses de données. Ces données sont par nature multimodales, obtenues sur des systèmes et référentiels différents. De manière générale, l'analyse de nuages de points est en pleine expansion avec la prise en compte de caractéristiques topologiques. Notons que les collaborations avec d'autres domaines sont en augmentation (archéologie, astronomie, médecine).

5. `https://www.ipol.im`

Calculabilité et Complexité

Contours

En 1936, Alan Turing définit un modèle de calcul, nommé *la machine de Turing* par Church, qui correspond exactement à ce qui est communément défini comme calculable par nos ordinateurs, classiques ou quantiques. La machine de Turing calcule sur un espace discret (les entiers ou les mots) en temps fini. D'autres modèles de calculs séquentiels, équivalents, ont été définis, comme les systèmes de réécriture (Thue, Post) et différents modèles de Random Access Machines (RAM). On dispose aussi de modèles de calcul parallèle comme les automates cellulaires de von Neumann et les circuits booléens. Les fonctions récursives définies dans les années 1930 - 1940 par notamment Gödel, Kleene, Post et Turing, modélisent mathématiquement ce qui est calculable. Dans un autre genre, le λ-calcul propose une troisième modélisation du calculable qui continue à irriguer la théorie de la programmation. La calculabilité étudie la frontière entre le décidable et l'indécidable avec des méthodes propres (énumération, point fixe, réductions, degrés, méthodes de priorité, etc.).

L'analyse récursive étend la calculabilité classique, discrète par essence, aux espaces infinis, comme les réels, les sous-ensembles compacts d'un espace euclidien ou encore aux espaces de fonctions dans les entiers, avec un lien profond entre continuité et calculabilité. Plusieurs directions de recherche se sont développées comme les mathématiques constructives, avec différents modèles de calcul comme les machines de Turing avec oracles, les modèles algébriques à la Blum, Shub et Smale, ou encore le GPAC *(General Purpose Analog Computer)* de Shannon. Le temps de calcul peut être infini, voire contrôlé par des ordinaux. Dans son ensemble la calculabilité couvre des questions fondamentales, tout particulièrement en théorie des ensembles, en logique et en mathématiques.

Si la calculabilité classique repose sur l'hypothèse que les phénomènes de la nature sont d'essence discrète, d'autres modèles de calcul ont vu le jour dont l'inspiration provient notamment de la biologie et de la physique. Ces questions permettent d'explorer ce qu'il serait possible de calculer avec des ordinateurs basés sur des paradigmes alternatifs : l'utilisation de l'intrication quantique, de composants analogiques, de modèles où le temps est continu, de modèles biologiques, etc.

Enfin, la calculabilité est relative à la dénotation, c'est-à-dire à l'ensemble des fonctions calculables. Or la notion d'algorithme et l'étude de l'ensemble des algorithmes sont des questions différentes et plus délicates. Gurevich a proposé un modèle, les *abstract state machines*, qui identifie tous les algorithmes séquentiels, formalisant ainsi le début d'une théorie d'une « calculabilité intentionnelle ».

Quant à la théorie de la complexité algorithmique, elle s'articule autour de deux grands axes. Le premier étudie la conception et l'analyse de la complexité des algorithmes. L'autre est le prolongement naturel de la calculabilité, où l'on étudie ce qui est calculable sous des contraintes de ressources comme le temps ou l'espace.

Le premier axe concerne la conception d'algorithmes de résolution de problèmes abstraits (comme la coupe minimum dans un graphe) ou concrets (par exemple le placement de ressources). Ces algorithmes permettent d'identifier la complexité en fonction d'une

ressource comme le temps, l'espace, le nombre de processeurs ou de requêtes, etc. Il s'agit alors de déterminer des bornes supérieures ou inférieures de la complexité de la résolution d'un problème donné, ou encore sa complexité en moyenne. Des méthodes spécifiques ont été développées pour calculer ces bornes de complexité. Ainsi, les bornes inférieures sont évaluées à partir d'un ensemble de méthodes algébriques, combinatoires, d'approximation où les mathématiques discrètes et continues jouent un rôle crucial et les bornes supérieures sont obtenues à partir de méthodes de conception (relaxation linéaire, échantillonnage aléatoire, etc.) et d'analyse d'algorithmes. Une illustration de ce domaine est l'étude des problèmes NP-difficiles et des problèmes d'optimisation combinatoire qui ont conduit à la construction d'algorithmes polynomiaux sous certaines hypothèses. Les domaines d'application sont très variés : on peut citer, sans prétendre à l'exhaustivité, l'optimisation, le streaming, les algorithmes pour les graphes, l'algorithmique distribuée, l'algèbre, l'algorithmique des systèmes « naturels » (réseaux sociaux, recommandations, biologie, etc.) ou encore des méthodes d'apprentissage automatique. La prise en compte de la qualité des solutions produites en fonction de conditions liées à une notion de coût, de probabilité d'erreur ou de succès, ou encore de bruit dans les données, est un sujet important.

Le second axe traite de l'étude structurelle des classes de complexité algorithmique (en temps, en espace, etc.) définie à partir des modèles de calcul. Les approches reposent essentiellement sur des outils de la logique provenant de la théorie des modèles finis ou de la théorie de la démonstration et sur des outils de la calculabilité. Cette double approche dessine deux communautés qui interagissent. Dans les deux cas, l'objectif est d'avoir un modèle mathématique des classes de complexité. Historiquement, la complexité descriptive, qui s'appuie sur les modèles finis, a des applications dans le domaine des bases de données. La complexité implicite, qui s'appuie sur des structures infinies, a des applications dans le domaine de la théorie des langages de programmation, voire dans celui de la sécurité.

Il y a un lien profond entre la théorie de la calculabilité et celle des probabilités et statistiques. En effet, la notion d'objet aléatoire, centrale dans l'intuition probabiliste, n'a pas de définition formelle en théorie des probabilités : ces objets y sont vus dans leur ensemble, globalement donc, jamais de façon individuelle. Il y a là une lacune sur laquelle se sont penchés quelques probabilistes de renom, tout particulièrement Kolmogorov. C'est l'une des motivations qui ont amené à la théorie de la complexité en information et à l'étude algorithmique de l'aléatoire. Plus qu'une simple reformulation de la théorie de l'information de Shannon, cette théorie tente d'étudier la question centrale de l'information.

Avancées récentes

En calculabilité, on notera en particulier des résultats de non-calculabilité, comme le fait que l'accessibilité dans les réseaux de Petri est non élémentaire (voir le paragraphe correspondant dans la section *Logique informatique*, qui s'appliquent à la vérification des systèmes, et des résultats sur des modèles alternatifs de calcul, en particulier sur les classes de complexité dans les modèles de calcul sur des domaines continus. Enfin, il

faut mentionner les avancées dans l'analyse calculatoire de théorèmes mathématiques qui affirment l'existence de certains objets, et peuvent être formulés comme des problèmes dont ces objets sont les solutions. Ces dernières ont récemment montré une richesse insoupçonnée dans l'analyse de théorèmes combinatoires comme le théorème de Ramsey.

Dans l'étude des classes de complexité, la complexité descriptive a creusé son sillon et la complexité implicite des calculs a atteint un certain degré de maturité. À titre d'exemple, des résultats récents ont été obtenus caractérisant la classe des fonctions calculées par un algorithme probabiliste en temps polynomial, ou encore caractérisant la classe des fonctions de type-2 (sur les réels) calculables en temps polynomial. En conséquence, ces travaux construisent des modèles mathématiques de classe de fonctions calculables en ressources bornées en s'appuyant notamment sur des systèmes de type et des outils de la théorie de la démonstration.

L'analyse d'algorithmes et de problèmes s'est rapprochée des mathématiques par le biais de la géométrie, de l'algèbre et de la combinatoire. Les directions de recherche actuelles sont nombreuses et variées : les problèmes d'optimisation, les algorithmes d'approximation, la complexité algébrique, la complexité des systèmes dynamiques, ou encore la complexité des circuits. Il faut signaler en particulier les avancées récentes sur l'approximation de certains problèmes NP-difficiles et sur l'apprentissage.

Prospective

L'étude de la calculabilité et de la théorie de la complexité a un impact épistémique sur de nombreuses disciplines comme les mathématiques, la logique, la physique, la biologie, la sociologie et la philosophie. L'étude de la calculabilité et de la théorie de la complexité des modèles non classiques ouvre un champ d'investigation important, lié à la calculabilité classique et aux fondements des mathématiques (logique, théorie de la preuve, théorie des ensembles) avec notamment des travaux sur les mathématiques constructives, le calcul analogique ou encore les calculs en temps infini.

En ce qui concerne la conception et l'analyse des algorithmes, des progrès continus permettent à la fois de resserrer les bornes et de mesurer à quel point nous sommes encore loin de comprendre les mécanismes qui nous permettront de refermer les conjectures et les écarts entre bornes inférieures et supérieures (on pourra méditer, par exemple, le résultat sur l'inexistence de preuves naturelles pour la conjecture $P \neq NP$). Les techniques développées pour cela sont au croisement de l'algorithmique, des mathématiques et de la physique.

Aujourd'hui, l'accès à des masses de données et à de nouvelles approches du calcul comme l'apprentissage profond (*deep learning*) constitue un nouveau défi pour la calculabilité, l'algorithmique, pour l'analyse des algorithmes, et pour la formulation même de ce qu'on peut appeler un algorithme dans ce cadre.

Pour finir, la dualité entre d'une part le développement de modèles de calcul concrets et d'autre part la calculabilité et l'algorithmique est un enjeu fondamental de l'informatique. La construction de l'ordinateur quantique en est une illustration, et les modèles de calcul bio-inspirés en sont une autre. Cette problématique s'étend aux avancées récentes en Intelligence artificielle, notamment à cause des réseaux de neurones vus comme mo-

dèle de calcul, mais également aux questions d'efficacité algorithmique pour faire face à la transition énergétique.

Calcul quantique

Contours

Préambule. Dans les années 80-90, une nouvelle communauté scientifique s'est forgée afin d'identifier les possibilités que permettrait le traitement de l'information quantique. D'abord une fiction, cette possibilité est devenue peu à peu une réalité. Des avantages inégalés ont été mis en évidence pour de nombreuses applications, parfois avec des perspectives de réalisation concrète à court terme. Ainsi, la transmission d'information quantique sur de grandes distances a été réalisée en situation réelle, y compris par satellite, rendant possible le déploiement d'une nouvelle cryptographie. La manipulation d'information stockée sur des mémoires quantiques, pourtant de petite taille, a déjà permis de réaliser des calculs, certes artificiels mais non reproductibles sur nos machines actuelles, y compris avec nos supercalculateurs.

Les applications plus emblématiques restent sans aucun doute un protocole quantique de distribution de clés secrètes dont la sécurité est inconditionnelle, et un algorithme quantique pour factoriser de grands nombres qui remet potentiellement en cause la plupart des procédés cryptographiques à clé publique. La maturité que cette nouvelle discipline a su acquérir en une si courte période est spectaculaire. Elle aborde un très grand nombre de domaines de l'informatique fondamentale (modèles de calcul, complexité, algorithmes, langages de programmation, méthodes formelles, vérification, théorie de l'information, cryptographie) et en influence beaucoup d'autres, que ce soit en mathématiques ou en physique. Cette nouvelle discipline est maintenant représentée à travers plusieurs conférences spécialisées, dont la conférence phare du domaine (Quantum Information Processing), et dans la plupart des conférences généralistes de l'informatique fondamentale.

Cependant l'informatique quantique est encore loin d'englober toute l'informatique, notamment dans ses aspects les plus appliqués. En effet, l'engouement industriel pour cette discipline est tout récent, et de plus l'ordinateur quantique n'existe tout simplement pas encore. Son existence future est d'ailleurs elle-même toujours questionnée.

L'arrivée récente de prototypes quantiques change partiellement la donne, et met notamment l'accent sur les besoins complémentaires qu'elle engendre : il est indispensable de définir des langages de haut niveau et de disposer d'outils de programmation robustes, tout comme d'étudier les modèles de calculs et leurs relations entre eux. Enfin la disponibilité de ces premières machines très imparfaites et encore rares souligne l'intérêt de disposer d'approches théoriques permettant l'anticipation de leur performance, leur benchmarking, ou encore leur utilisation à distance de façon sécurisée.

Algorithmes. Les algorithmes quantiques de première génération ont surtout permis de résoudre des problèmes mathématiques abstraits, comme celui de la factorisation (Shor, 1994). Une deuxième génération d'algorithmes quantiques a grandement élargi le spectre des applications. L'algorithme de Grover (1996) et ses généralisations fournissent un gain en général quadratique aux heuristiques et algorithmes d'optimisation déterministes ou probabilistes. Plus récemment, un troisième type d'algorithmes a suscité encore plus d'in-

térêt, notamment du côté de l'industrie, en partie en raison du développement récent de nouvelles techniques algorithmiques quantiques liées à l'apprentissage automatique. Il s'agit de l'algorithme Harrow-Hassidim-Lloyd (2009) qui peut, sous certaines conditions, résoudre un système d'équations linéaires exponentiellement plus rapidement que les algorithmes classiques. Cet algorithme permet d'accélérer la résolution d'autres problèmes d'algèbre linéaire, avec des applications importantes en apprentissage machine et pour le traitement de données volumineuses.

Langages de programmation. Pour faire le lien entre les algorithmes quantiques et les propositions d'ordinateurs quantiques, il est indispensable de contribuer à l'ensemble de la pile du logiciel quantique. Des langages de programmation quantique ont été développés ces dernières années (Quipper, Qiskit, QLM, ...) par le milieu académique ou industriel (Atos, IBM, Google, Microsoft, ...) Leur utilisation permet le développement de techniques d'analyse de programmes pour l'estimation de ressources (temps, espace, intrication, ...) et de techniques de certification.

Codes correcteurs. Une spécificité du traitement de l'information quantique est qu'elle est rapidement corrompue par le bruit qui affecte inévitablement les systèmes physiques. Il est aujourd'hui admis que les ordinateurs quantiques devront implémenter des routines de correction d'erreur pour protéger cette information. La stratégie de correction d'erreur la plus répandue est celle des codes de surface. C'est l'approche que suivent les grands acteurs du calcul quantique comme IBM ou Google pour les qubits supraconducteurs. Cette approche est hélas actuellement beaucoup trop gourmande en coût matériel pour être exploitée sur les premiers prototypes quantiques qui sont donc très bruités.

Technologies NISQ. C'est pourquoi de nombreux algorithmes ont été proposés pour apporter des solutions à plus court terme sur ces prototypes très bruités dits NISQ (Noisy Intermediate Scale Quantum), comme lors de la démonstration de force de Google en 2019 qui annonçait avoir atteint la suprématie quantique. Si l'avantage quantique proposé est souvent difficile à prouver, ces algorithmes offrent néanmoins des heuristiques intéressantes, grâce auxquelles des bénéfices concrets en matière de résolution de problèmes pratiques sont attendus.

Cryptographie. Enfin la cryptographie n'est pas en reste, et heureusement puisque l'algorithme de Shor mettrait à mal la totalité de la cryptographie à clé publique actuellement déployée. Historiquement, c'est en cryptographie que le premier résultat d'informatique quantique est apparu, en 1984. Il s'agit du protocole de distribution quantique de clés secrètes de Bennett et Brassard (BB84) qui a été prouvé inconditionnellement sûr plus tard en 1998. Le remède est donc arrivé avant l'attaque de Shor... Une communauté s'est ainsi structurée autour de la conférence QCrypt (Quantum Cryptography) afin de développer une cryptographie quantique offrant de nouvelles garanties de sécurité. Son déploiement en situation réelle sur de très grandes distances a été validé, y compris par satellite, ce qui en fait une solution tout à faire crédible à court terme.

Suivant une approche orthogonale, des solutions de substitution purement classiques, dites post-quantiques, ont aussi été développées, se structurant quant à elle autour de la conférence PQCrypto (Post-Quantum Cryptography). De nouvelles propositions de

chiffrements classiques ont émergé à travers la compétition de standardisation du NIST (2016-2022).

Structuration, plan national et relations avec l'industrie. L'effort financier et de structuration de la communauté l'a profondément transformée. Initialement focalisée principalement sur les aspects les plus théoriques de la discipline, la communauté se confronte maintenant aux aspects les plus appliqués en collaborant avec le monde industriel afin de rechercher, entre autres, les usages des premiers prototypes quantiques disponibles.

L'acte fondateur de la communauté a été la création en 1998 de la conférence Quantum Information Processing (QIP) dédiée au traitement de l'information quantique, puis en 2002 de l'Institute for Quantum Computing (IQC) à l'Université de Waterloo (Ontario, Canada), le premier de ce type dans le monde dédié au calcul quantique sous tous ses aspects, physiques, informatiques et d'ingénierie, fondé grâce au soutien financier de Mike Lazaridis, créateur du BlackBerry.

Très tôt les industriels ont suivi et accompagné cette discipline. Parallèlement au soutien à la création du laboratoire IQC, D-Wave est créée en 1999, première entreprise dédiée au calcul quantique (machines dédiées au recuit simulé quantique). En 2011 une division dédiée aux algorithmes quantiques est mise en place chez Microsoft, avec le lancement d'un environnement de programmation dédié en 2017 (Q#). Suit en 2013 Google, qui se lance dans la construction d'un calculateur quantique, et en parallèle crée une division dédiée aux algorithmes et à la programmation (dont le langage Cirq en 2018), pour finalement initier en 2019 une série d'expériences exhibant la supériorité quantique des prototypes actuels de processeur quantique. En 2016, IBM crée un calculateur quantique accessible dans le cloud, ainsi qu'un environnement open-source de programmation (Qiskit) en 2017.

Précédé par le lancement en 2014 du UK National Quantum Technologies Programme en 4 hubs (Birmingham, Glasgow, Oxford, York), le Quantum Flagship est porté en 2018 par la Commission Européenne pour 10 années. Suivent en 2019 le lancement du National Agenda on Quantum Technology des Pays-Bas (Quantum Delta NL) pour 7 années en 5 hubs (Amsterdam, Delft, Eindhoven, Leiden, Twente), et finalement en 2021 le plan français de stratégie nationale sur les technologies quantiques qui identifie 3 hubs (Grenoble, Paris, Saclay) et lance le PEPR Quantique, ainsi que la construction d'une plate-forme nationale de calcul quantique.

Dans l'écosystème français, dès 2000 est créé le GdR Quantum Information and Communication, qui devient le GdR Quantum Information : Foundations and Applications en 2008, ainsi que le GT Informatique Quantique du GDR IM en 2006. Entre 2016 et 2018, des programmes de R&D sur le calcul quantique (logiciels et simulateurs) sont mis en place chez Atos, EDF et Total, et plusieurs startups voient le jour, dont pour le calcul Veriqloud (2017), QC Ware France (2019), Qubit Pharmaceuticals (2020) ; pour la cryptographie post-quantique Cryptonext Security (2019) ; pour les technologies Quandela (2017), Pasqal (2019), Alice & Bob (2020), WeLinQ (2022) ; ainsi que des organismes financiers tels que Quantonation (fonds d'investissement, 2018), Le Lab Quantique (association à but non lucratif, 2018).

Avancées récentes

Algorithmes.

Plusieurs cadres généraux ont été développés permettant à chacun d'écrire ses propres algorithmes quantiques sans être un expert en informatique quantique. La France est historiquement bien positionnée sur ce sujet, avec des contributions importantes au fil du temps dont les plus récentes portent sur les analogues quantiques des marches aléatoires, des méthodes de Monte Carlo, de sparsification, et sur des algorithmes pour l'algèbre linéaire au cœur de l'apprentissage automatique, de l'analyse numérique et de la résolution d'équations différentielles.

Cette approche a permis l'extension des algorithmes quantiques à de nombreux modèles plus contraints tels que les algorithmes de streaming ou les algorithmes distribués, ainsi qu'à de nouveaux domaines comme l'algorithmique du texte et l'algorithmique géométrique, ou encore la cryptanalyse post-quantique et l'apprentissage automatique, avec de multiples applications très concrètes en ingénierie, dans le domaine médical et en finance.

La plupart de ces approches apportent des gains significatifs mais non-exponentiels, en raison notamment d'un résultat remarquable sur la déquantisation de la plupart des algorithmes pour l'apprentissage automatique, qui a donné lieu à une nouvelle classe d'algorithmes probabilistes dits *quantum-inspired*.

Complexité.

Sous l'impulsion des travaux menés sur les limitations des algorithmes quantiques, la compréhension de la complexité en requêtes a énormément progressé dans le cadre probabiliste comme dans le cadre quantique. Sans surprise, la preuve de la *sensitive conjecture* (2019) utilise des notions proches du quantique. Plusieurs équipes françaises se distinguent, en contribuant à une classification de la complexité en requêtes des problèmes qui a fait des progrès considérables depuis 2015.

La compréhension de la corrélation des états quantiques (distribués entre deux ou plusieurs parties interagissant ou non entre elles) sous l'angle de la complexité est une des forces nationales, avec des applications qui dépassent le quantique. Cette recherche a culminé en 2020 avec le résultat MIP*=RE, c'est-à-dire que la classe MIP* des langages qui peuvent être décidés par un vérificateur classique interagissant avec plusieurs prouveurs quantiques tout-puissants et partageant de l'intrication est égale à la classe RE des langages récursivement énumérables. Ce résultat a pour conséquence la réfutation de la conjecture de plongement d'Alain Connes.

S'il était déjà possible de certifier inconditionnellement qu'un état est quantique lorsqu'il est partagé entre deux entités non-communicantes, ce qui est une généralisation des inégalités de Bell (dont la validation expérimentale a valu le prix Nobel de physique à Alain Aspect en 2022), il est depuis peu possible de certifier directement qu'un état est quantique de façon calculatoire même s'il n'est localisé qu'en un seul endroit. Les applications sont importantes d'un point de vue théorique, voire métaphysique, mais aussi pratique avec la notion de calcul délégué à distance (variante quantique du chiffrement homomorphe). Sur ces questions, plusieurs laboratoires français sont historiquement bien

positionnés avec plusieurs travaux majeurs et même une startup sur le domaine. Les travaux sur le modèle de calcul par mesure et le langage ZX-Calculus y sont pour beaucoup, car ils sont adaptés au calcul délégué.

La complexité quantique a aussi fait son entrée en gravité quantique, où il est maintenant courant de voir des arguments à base de complexité quantique pour expliquer quantitativement des phénomènes propres aux trous noirs et aux vers. Ces liens inédits ont fait couler beaucoup d'encre et un nouveau domaine est peut-être en train de naître.

Heuristiques et NISQ.

Une théorie de la complexité des calculs réalisés sur les machines NISQ a été développée. Relativement à un oracle, les classes de complexité BPP et BQP encadrent strictement la nouvelle classe de complexité NISQ associée, qui correspond informellement à ce qu'une machine classique peut réaliser avec un accès à une machine NISQ.

En parallèle, des heuristiques dédiées aux machines NISQ sont massivement développées de façon heuristiques. On retiendra des techniques telles que Quantum Variational Algorithm, Quantum Approximate Optimization Algorithm, Quantum Variational Eigensolver. Il est cependant difficile de valider ces approches expérimentalement en l'absence de machine plus conséquente. Ce sont donc des modèles mathématiques qui permettent de les étudier mathématiquement. Alors que certaines heuristiques apparaissent finalement moins bonnes qu'espéré, il reste quelques avantages possibles. Ainsi, des systèmes quantiques de petite taille ont pu être simulés sur les machines NISQ du commerce, dont en particulier une molécule composée de 12 atomes d'hydrogène et une réaction chimique dans une molécule contenant des atomes d'hydrogène et d'azote.

Codes correcteurs. Des progrès fulgurants et simultanés en codes correcteurs quantiques et codes correcteurs classiques ont été réalisés, à travers une construction unique qui résout deux conjectures. Tout d'abord, des codes correcteurs classiques localement testables ont été développés. De tels codes possédant de bonnes propriétés (distance linéaire, taux constant) étaient recherchés depuis une vingtaine d'années. Ces codes, de type LDPC (low-density parity-check), sont adaptés à la construction de nouveaux codes correcteurs quantiques performants dits quantum-LDPC, pour lesquels la France est leader y compris sur les aspects de décodage efficace. Ces codes ouvrent la voie à des calculs auto-stabilisés avec un facteur d'amplification en nombre de bits indépendant de la taille du calcul à effectuer. Une première théorique avec des retombées pratiques potentielles !

Cryptographie post-quantique.

Les premières propositions de chiffrement à clé publique, de signature et de génération de clé secrète qualifiées à la compétition NIST ont été annoncées en juillet 2022. Trois des quatre propositions retenues sont coportées par des laboratoires en France.

La France se distingue aussi en cryptographie quantique, au niveau expérimental comme au niveau théorique, et ce continuellement depuis de nombreuses années.

Prospective

Nous sommes à un moment singulier du développement de l'informatique quantique, alors que des ordinateurs quantiques de petite et moyenne mémoire sont sur le point de

devenir une réalité. Cela soulève de nouveaux défis algorithmiques, qui vont de l'analyse comparative et des tests d'ordinateurs quantiques à la compréhension concrète des types de tâches algorithmiques pour lesquelles ces ordinateurs seront utiles. La théorie des tests de dispositifs quantiques interactifs a déjà conduit à de nouvelles connexions profondes avec la cryptographie et la complexité, et à la résolution de questions ouvertes de longue date en mathématiques. Parmi les tâches algorithmiques intensément étudiées figurent la simulation de systèmes quantiques, la chimie quantique et l'apprentissage automatique quantique. Les progrès des protocoles de correction d'erreurs quantiques et de tolérance aux fautes seront essentiels pour le passage à l'échelle des ordinateurs quantiques. Dans le même temps, des concepts issus du calcul quantique jouent un rôle essentiel dans la compréhension de problèmes fondamentaux ouverts en physique, par exemple en théorie de la gravité quantique.

Les défis directs auxquels le domaine est confronté sont, sur le plan expérimental, la réalisation d'un ordinateur quantique, ou a minima d'un calculateur quantique spécialisé mais capable de passer à l'échelle, et sur le plan algorithmique, la recherche d'applications pour un tel calculateur, même limité. En effet, l'écart est actuellement si gigantesque entre la technologie existante et celle nécessaire pour exploiter industriellement l'avantage des algorithmes connus, qu'il faut un effort symétrique sur ces deux sujets.

Sur le plan cryptographique, les défis sont davantage orientés vers le déploiement de solutions quantiques sur de grandes distances et les réseaux de communication existants, et la mise à jour de l'ensemble des dispositifs classiques existants par des versions classiques dites post-quantiques.

Plus indirectement, la quête d'applications concrètes est récente, avec un réel enthousiasme pour la découverte rapide d'un impact de cette recherche pour la société. Cette dynamique fait courir le risque d'un désintérêt pour les recherches les plus théoriques avec, en ligne de mire, une fragmentation de la communauté. Nous assistons déjà à une absorption rapide non seulement des jeunes docteurs mais aussi de chercheurs confirmés dans l'industrie et en particulier dans des startups. Une autre inquiétude, en miroir, porte sur l'effet négatif que l'échec des applications espérées à court terme aurait sur la discipline, jusque dans ses aspects les plus théoriques.

Combinatoire(s), systèmes dynamiques et aléatoire

Contours

La combinatoire étudie la façon dont des objets discrets s'assemblent (se *combinent*) lorsqu'il sont soumis à des contraintes. Ces contraintes peuvent être locales (les noeuds d'un arbre n'ont qu'un parent, les lettres des mots d'un langage régulier dépendent des précédentes, les faces d'un graphe peuvent être contraintes dans leur nombre de côtés) ou globales (le graphe est coloriable avec 4 couleurs, la permutation ne contient pas un motif donné,...). Dans tous les cas, il s'agit de comprendre comment ces contraintes influent sur le nombre ou la forme des objets construits, et souvent aussi comment exploiter la structure induite par ces contraintes pour en déduire des algorithmes efficaces. Ces questions sont présentes dans toutes les sciences, pour modéliser les liens entre objets (par exemple, entres particules en physique, entre individus ou espèces en biologie, pour concevoir des structures de données en informatique, pour décrire des objets en mathématiques).

Du fait de la nature discrète de ses objets d'étude, la combinatoire occupe une place centrale en informatique mathématique. Ainsi, de nombreuses variantes d'arbres sont utilisées comme structures de données et analysées pour quantifier les performances d'algorithmes (voir notamment la section *Géométrie(s) et image*) ; les graphes permettent d'étudier tant des questions fondamentales de calculabilité et de complexité (voir la section qui leur est dédiée, ainsi que la section *Algorithmique*), que des questions d'application concrète comme l'étude du graphe du web ; la combinatoire des mots et leur algorithmique jouent un rôle très important dans l'étude des séquences génétiques ; les boucles en programmation se modélisent et s'étudient par le biais de systèmes dynamiques discrets ; etc.

Une grande partie des objets étudiés sont classiques et fondamentaux : graphes, arbres, mots, chemins, permutations,... La combinatoire mobilise de nombreuses branches des mathématiques pour leur étude, et c'est souvent autour des outils utilisés que les communautés se sont cristallisées.

La *combinatoire algébrique* étudie ainsi les structures algébriques sous-jacentes aux questions combinatoires ou, réciproquement, les problèmes combinatoires apparaissant dans certaines branches de l'algèbre. Par exemple, la théorie des espèces fait un pont entre combinatoire et théorie des catégories, et fournit un point de vue unifié sur les constructions de la combinatoire énumérative ; les fonctions symétriques jouent un rôle important dans l'étude des représentations du groupe symétrique, et interagissent fortement avec la combinatoire des tableaux de Young ; la preuve de propriétés de positivité ou d'intégralité de coefficients de fonctions spéciales ou polynômes orthogonaux dans diverses bases passe souvent par des interprétations combinatoires, ...

La *combinatoire énumérative*, partant de questions de dénombrement, met en évidence des classes d'universalité de modèles aléatoires en analysant comment les contraintes qui pèsent sur la construction des objets permettent ou non de les classifier comme rationnels, algébriques, différentiellement finis ou non, ce qui impacte la croissance asymptotique des nombres d'objets en fonction de la taille, ou le comportement de paramètres comme les distances moyennes entre sommets dans des arbres ou des graphes.

La combinatoire analytique repose sur les méthodes de séries génératrices qui mènent à des équations fonctionnelles, dont l'étude permet d'approcher le comportement des solutions analytiques. L'asymptotique des suites de dénombrement en découle par l'analyse des singularités de ces solutions. On parvient ainsi à une quantification fine du comportement typique de grands objets aléatoires, avec des analyses probabilistes de convergence vers des distributions discrètes ou continues (limites d'échelle, objets browniens). La physique mathématique étudie aussi, et depuis bien longtemps, les interactions entre phénomènes microscopiques et macroscopiques, et la place de l'aléa dans la structure de l'univers. Ses interactions avec la combinatoire sont de plus en plus fréquentes, et fructueuses.

Les systèmes dynamiques enfin, sont des modèles permettant d'étudier l'évolution à intervalles de temps discrets d'objets ou de systèmes, qui peuvent être finis ou infinis : des cellules, des gènes, des mots, des coloriages de la ligne ou du plan, etc. Parmi ces modèles, on trouve les automates cellulaires, les pavages du plan par des formes géométriques, les pavages du plan discret évitant des motifs interdits, les tas de sables, les réseaux booléens et bien d'autres. Ces systèmes peuvent être étudiés de plusieurs manières : soit sous l'angle de la prédiction (que se passera-t-il dans quelques itérations ?), soit sous l'angle des comportements asymptotiques (que se passera-t-il à long terme ?). On peut également s'intéresser aux cas extrêmes : quels sont les comportements les plus compliqués qui peuvent être obtenus par des systèmes dynamiques vérifiant certaines propriétés. Au sein du GdR, les systèmes dynamiques discrets les plus étudiés sont probablement ceux de la dynamique symbolique : les sous-shifts, les coloriages du plan, de la ligne ou d'un autre espace discret évitant des motifs interdits, et les automates cellulaires.

Les relations entre combinatoire et informatique ne se cantonnent pas à l'aspect fondamental des structures combinatoires en informatique. La nature discrète des objets étudiés mène également au développement de nouveaux champs algorithmiques permettant expérimentation, simulation et analyse. Un cadre d'application évident est fourni par l'analyse de séquences génétiques en *combinatoire des mots*, où la quantité de données à traiter renouvelle sans cesse le besoin d'algorithmes ou de structures de données adaptées et ouvre également de nouvelles perspectives (par exemple avec la compression et l'indexation de tera-octets de données bruitées et redondantes). La recherche d'automatisation en combinatoire énumérative, analytique ou algébrique, mène quant à elle à des mathématiques expérimentales grâce au développement d'algorithmes et de logiciels spécialisés, notamment en calcul formel et en génération aléatoire

Avancées récentes

En 20 ans, le coût du séquençage par paire de bases a été divisé par près d'un million, facilitant l'acquisition des séquences génétiques et apportant chaque année de nouvelles questions pour le stockage et le traitement de ces données. L'utilisation massive d'outils théoriques (arbres de suffixes, transformée de Burrows-Wheeler, filtres de Bloom, graphes de de Bruijn, fonctions de hachage minimales parfaites, ...) a dynamisé la recherche fondamentale en combinatoire des mots. De nouvelles questions ou concepts méthodologiques sont apparus (par exemple les attracteurs de séquences, les graphes de Wheeler,

les Spectrum-Preserving String Sets, etc.) menant à de nouvelles structures de données et algorithmes. Les logiciels développés en bio-informatique à partir de ces algorithmes ou concepts théoriques sont maintenant utilisés de manière routinière dans des laboratoires de biologie, pour la compréhension du vivant, ou dans des hôpitaux, pour le diagnostic ou le suivi de maladies.

La combinatoire analytique multivariée est passée en peu de temps d'une collection d'articles de recherche difficiles d'accès à un domaine avec des ouvrages introductifs, des logiciels et de nouveaux développements. Les outils théoriques et logiciels développés permettent une analyse directe de nombreuses questions inaccessibles auparavant, notamment pour l'analyse simultanée de plusieurs paramètres de structures combinatoires de grande taille. Ces méthodes sont en plein essor actuellement.

L'étude des cartes planaires, modèle de géométrie discrète à l'intersection entre combinatoire, physique statistique et théorie des probabilités, s'est fortement développée ces vingt dernières années. En particulier l'étude des cartes planaires aléatoires est maintenant un sujet de recherche à part entière. Si aujourd'hui le comportement des cartes planaires uniformément aléatoires est bien connu, la situation est très différente lorsque l'on considère des cartes aléatoires échantillonnées selon une loi de probabilités différente. Un sujet brûlant dans ce domaine est l'étude des cartes planaires décorées par un modèle de physique statistique (tel que le modèle d'Ising). Des résultats profonds dans ce domaine ont été obtenus récemment sur ces cartes doublement aléatoires (où la carte et son coloriage sont aléatoires) permettant d'accéder à des limites locales, des exposants critiques et des transitions de phase. Ces résultats ouvrent des perspectives fascinantes à étudier dans les prochaines années, car des liens sont conjecturés entre ces modèles et d'une part des modèles de physique sur $\mathbb{Z}^2$ autour de la formule KPZ, d'autre part la gravité quantique de Liouville.

Les interactions avec la physique statistique sur les questions de récurrence topologique et d'intégrabilité touchent de plus en plus de combinatoriciens. Des résultats importants sur ces objets ont été obtenus : modèle d'Ising (notamment sur les cartes), système de particules, 6 (ou 8)-vertex -models, équations de Yang-Baxter. La gravité quantique bi-dimensionnelle des physiciens a été reformulée en termes de cartes, fournissant en retour de nouvelles limites d'échelle en probabilité. Cette stratégie a aussi donné naissance au domaine de la géométrie énumérative, lorsque Kontsevich a reformulé de manière combinatoire le calcul des nombres d'intersection de Witten. Ce domaine demeure aujourd'hui une thématique très active.

Les liens traditionnels entre combinatoire et analyse d'algorithmes ont été renouvelés récemment par des travaux plus proches des langages de programmation, pour lesquels les modèles classiques de calcul ne permettaient pas de capturer les phénomènes observés. En retour, ces travaux ont un impact technologique (découverte d'un bug dans l'algorithme de tri de Java par exemple). Cette tendance où l'étude et l'optimisation d'implantations efficaces en pratique mène a des modèles combinatoires du calcul plus fins devrait se poursuivre.

Du côté des systèmes dynamiques discrets, la question des pavages du plan a vu plusieurs avancées majeures ces dernières années. La plus récente est la découverte d'une

mono-tuile pavant le plan seulement apériodiquement, mais on peut également mentionner que le problème des formes de pentominos irréguliers permettant de paver le plan a été fermé par Michael Rao et que le plus petit jeu de tuiles de Wang pavant apériodiquement a été trouvé par Jeandel et Rao. Une version un peu plus faible de la conjecture de Nivat a également été démontrée récemment.

La conjecture sur les substitutions de Pisot (qui s'exprime en termes de pavage par fractals de Rauzy ou en termes de spectre purement discret de shifts substitutifs) a été prouvée pour les substitutions définies par la beta-numération mais le cas général reste ouvert. Elle a été étendue aux systèmes S-adiques (définis par des suites de substitutions) par des membres du GT SDA2 et prouvée pour de grandes classes de systèmes par ces auteurs et par le groupe Pytheas Fogg. Une réponse à la "conjecture S-adique" (qui consiste à décrire les shifts S-adiques avec complexité linéaire) a été donnée récemment.

On notera également que le preuve de nombreux résultats sur la dynamique des systèmes discrets s'appuie sur des notions issues de la théorie de la calculabilité et de la complexité, soulignant ainsi l'importance des interactions entre ces domaines.

Prospective

De nombreuses questions liées aux chaînes de caractères restent ouvertes. Un exemple est l'assemblage d'une séquence ADN ou ARN a partir d'un ensemble de lectures partielles, bruitées, avec parfois de nombreuses répétitions, et/ou provenant de plusieurs organismes ou versions de génomes proches (métagénomique, génomes polyploïdes) éventuellement en s'appuyant sur des séquences déjà connues pour accélérer le traitement. Dans ce domaine, un but à long terme serait la mise à disposition d'outils rapides et précis résolvant de manière optimale les problèmes posés par la reconstruction et l'analyse des séquences. Il faudra en particulier prendre le temps de développer des regards neufs sur les questions fondamentales pour accompagner les évolutions du domaine.

Une question au cœur de l'actualité de l'étude des cartes aléatoires est de définir des modèles de cartes en dimension supérieure (en recollant des tetraèdres de dimension 3 ou supérieures). Un des objectifs principaux est de définir un "bon" modèle de cartes en dimension supérieure, qui exhibe des comportements asymptotiques intéressants. Au vu de leurs applications en physique statistique, où les cartes donnent un modèle de gravité quantique, le cas de la dimension 4 (correspondant à 3 dimensions d'espace et une de temps) est particulièrement intéressant. Cependant, l'extension à la dimension 3 soulève déjà de nombreuses questions et devrait faire l'objet de développements fondamentaux dans les prochaines années. Au-delà des cartes, l'exploitation des interactions entre la combinatoire énumérative et la combinatoire intégrale développée dans la littérature physique devrait se développer et permettre d'obtenir des intuitions et des heuristiques sur les modèles que l'on cherche à étudier.

On observe que les choix d'implantation d'algorithmes classiques (tris, tables de hachage, ...) dans les langages de programmation contemporains sont parfois très originaux et éloignés des connaissances théoriques. L'analyse d'algorithmes doit maintenant se donner aussi comme objectif d'enrichir les modèles de calcul en y incorporant des aspects modernes d'architecture (hiérarchie de mémoire, prédiction de branchements,...) pour

proposer des analyses plus pertinentes des algorithmes utilisés en pratique. Un défi pour les combinatoriciens est donc de réconcilier la théorie et la pratique, en prenant en compte dans les algorithmes des éléments d'architecture des ordinateurs et des modèles aléatoires adaptés. La principale difficulté, pour ces deux points, est de parvenir à mieux modéliser la réalité tout en restant sur des modèles analysables mathématiquement.

Les liens récemment établis entre théorie des groupes et pavages donnent lieu à tout un programme de recherche dédié à l'identification de groupes qui admettent des pavages apériodiques. Ce programme est lié à des questions de calculabilité sur les groupes.

Les propriétés calculatoires des systèmes dynamiques discrets restent par ailleurs un objet fondamental d'étude. C'est le cas notamment pour les sous-shifts de basse complexité (ceux qui ont peu de motifs différents), dont l'intérêt a été confirmé récemment, qui voient de nombreux travaux sur leurs groupes d'automorphismes.

Enfin, la percolation sur les systèmes dynamiques discrets commence à être étudiée, par exemple sur les automates cellulaires ou certains pavages.

Graphes et algorithmes

Contours

La notion de graphe est l'une des plus simples et des plus intuitives que l'on puisse imaginer et pourtant les graphes sont à l'origine d'une multitude de problèmes mathématiques et algorithmiques dont certains se sont révélés extrêmement difficiles. Un exemple particulièrement célèbre est le fameux théorème des 4 couleurs : *Quatre couleurs suffisent pour colorier tous les sommets de n'importe quel graphe planaire de telle sorte qu'il n'y ait pas deux sommets adjacents coloriés avec la même couleur*. Ce problème, initialement exprimé comme un problème de coloration de carte de géographie, a suscité le premier engouement important des mathématiciens pour la théorie des graphes, au 19ème siècle. Il a aussi fourni l'un des premiers exemples de démonstration utilisant l'aide d'un ordinateur (en 1976), et il n'existe aucune preuve de ce théorème qui ne recoure pas à la puissance des ordinateurs. De plus, en 2005, une version formulée en Coq par Georges Gonthier et Benjamin Werner a permis à un ordinateur de vérifier une preuve du théorème des quatre couleurs. Ce résultat est considéré comme l'un des plus remarquables obtenus par un assistant de preuves.

Les liens entre théorie des graphes, mathématiques et informatique sont riches et variés et l'exemple ci-dessus est loin d'être unique. Les graphes constituent un outil très efficace pour modéliser des problèmes de toutes sortes : la théorie des graphes peut ainsi aider aussi bien à la conception de circuits intégrés qu'à l'élaboration de réseaux fiables et efficaces. Les applications sont nombreuses aussi en recherche opérationnelle (ordonnancement de tâches, établissement d'emplois du temps, problèmes de transports, etc. . .), en biologie, en chimie, en physique statistique, en géométrie algébrique, en théorie des jeux, dans les télécommunications... La nécessité d'établir l'existence d'une solution ou, selon les cas, de construire une solution explicite ou bien toutes les solutions, a conduit à la découverte de nombreux algorithmes efficaces en pratique et aussi à la conclusion que beaucoup de problèmes concernant les graphes sont NP-complets.

Néanmoins, beaucoup de problèmes sur les graphes, souvent reliés à des concepts mathématiques et informatiques, sont étudiés en amont de toute application directe. On peut d'ailleurs remarquer que la théorie des graphes est citée dans quasiment tous les domaines couverts par ce document. Il s'agit d'une réelle interaction entre ces domaines, dans les deux sens : il existe par exemple des théorèmes de la théorie des graphes qu'on ne sait pas démontrer sans recourir à des résultats d'algèbre linéaire, de théorie des nombres ou de topologie algébrique. Dans le monde, on trouve d'ailleurs des chercheurs en théorie des graphes aussi bien dans les instituts de mathématiques ou d'informatique, que dans les écoles de commerce, les laboratoires d'économie, etc. . .

De nombreux problèmes classiques de la théorie des graphes (coloration de sommets, d'arêtes, par listes, clique maximum, connectivité, décomposition,. . .) restent largement étudiés, soit dans leur forme initiale parce qu'ils ne sont pas encore résolus, soit restreints à des sous-classes de graphes ou bien encore sous la forme de variantes.

Les pistes les plus étudiées pour obtenir des algorithmes de résolution efficaces en pratique pour les nombreux problèmes de la théorie des graphes dont on sait qu'ils sont

NP-complets, sont les algorithmes d'approximation, les algorithmes distribués ou parallèles, les algorithmes exponentiels exacts et la complexité paramétrée. Cette dernière branche de la théorie de la complexité, dont l'introduction remonte aux années 1990, a connu des développements spectaculaires ces dernières années, en particulier en Europe. Beaucoup de ses résultats sont liés à l'étude de la structure de graphes caractérisés par une famille finie de « mineurs » interdits (où le terme « mineur » peut avoir diverses significations) et les paramètres considérés sont souvent ceux de la *largeur arborescente (ou treewidth), clique-width, rankwidth, etc.* Une autre méthode consiste en une « kernelisation », c'est-à-dire la réduction à une instance plus petite, appelée noyau. On doit aussi noter les travaux concernant les classes de graphes « nowheredense » et leurs liens avec la théorie des modèles finis.

Quelques autres sujets ont aussi émergé relativement récemment, en particulier la théorie des limites de graphes (tout début en 2003, livre de Lovàsz en 2010), en lien avec les «grands graphes» tels que celui du web, en lien également avec la combinatoire extrémale et d'autres domaines des mathématiques tels que les probabilités, la théorie de la mesure ou l'optimisation semi-définie. Cette théorie des limites de graphes est en particulier liée à la notion de Property Testing, qui consiste à évaluer une propriété d'un grand graphe (ou de manière plus générale, d'un grand objet combinatoire) en ne testant qu'un petit nombre de sommets (pris au hasard).

Avancées récentes

En complexité paramétrée, les progrès récents les plus significatifs portent certainement sur des résultats méta-algorithmiques de «kernelization» montrant l'existence de noyaux polynomiaux pour de larges familles de problèmes paramétrés ainsi que l'apparition de techniques pour déterminer des bornes inférieures sur la taille des kernels. Dans ce domaine, l'optimisation des complexités existantes et les liens entre méthodes paramétrées et méthodes d'approximation sont des enjeux importants.

Très récemment, un groupe de quatre chercheurs membres du GdR ont déini la « twin-width », un nouveau paramètre qui s'applique aux graphes et aux matrices. Ils ont montré que de nombreuses classes naturelles de graphes (en particulier les graphes de clique-width, rank-width ou tree-width bornée) ont une twin-width bornée. Leurs résultats unifient et étendent de nombreux résultats en complexité paramétrée. Depuis, ce nouveau concept a induit beaucoup de recherches et résultats.

Du côté des algorithmes d'approximation, on relève un résultat important de chercheurs parisiens, qui montre que la recherche locale est très efficace pour différents problèmes d'optimisation importants, par exemple pour certaines variantes des problèmes de *facility location.*

Des progrès notables pour le célèbre TSP (problème du voyageur de commerce) ont entraîné une compétition mondiale accélérée. Pour le voyageur de commerce graphique, ainsi que pour la version «chemin» de ce problème avec des métriques générales et aussi d'autres problèmes proches, des records de garantie d'approximation ont été obtenus à Grenoble, en partie en collaboration avec des chercheurs d'autres pays. Après une trentaine d'années sans progrès notable concernant les grandes conjectures connues, le

déclic est venu de l'utilisation de méthodes génériques nouvelles comme la maximisation d'entropie, issue de la théorie de l'information. Par la suite, les meilleurs résultats ont été obtenus par des arguments combinatoires plus élémentaires mais innovants, qui utilisent la théorie des graphes classique (surtout les couplages et et les propriétés de connectivité), la théorie des matroïdes et la programmation linéaire. D'autres résultats impressionnants ont été obtenus récemment ailleurs, par exemple un algorithme d'approximation à facteur d'approximation constant pour le TSP asymétrique, ce qui était un des grands problèmes ouverts du domaine. Une question d'actualité serait de diminuer ce facteur constant.

Des chercheurs marseillais ont aussi obtenu récemment de nouveaux résultats sur des conjectures d'informatique théorique (en particulier un contre-exemple à une conjecture de Thiagarajan sur les structures d'événements régulières et les dépliages de réseaux de Petri). Un autre résultat remarquable est la preuve en 2017 de la conjecture de Erdős-Sands-Sauer-Woodrow (1982) par des chercheurs lyonnais et grenoblois. Notons aussi la classification complète des pavages du plan par des pentagones convexes, qui a clos une question posée un siècle auparavant.

Enfin, un résultat majeur de ces dernières années est un théorème de Dujmović, Joret, Micek, Morin, Ueckerdt et Wood (JACM 2020) qui décrit les graphes planaires comme des sous-graphes du produit fort d'un chemin et d'un graphe de largeur arborescence bornée. Ce théorème a été utilisé par des chercheurs de la communauté française et internationale pour obtenir des résultats qui répondent à des problèmes ouverts importants du domaine : une conjecture célèbre d'Alon, Grytczuk, Haluszczak et Riordan datant de 2002, sur la *coloration non-répétitive* des graphes planaires (un problème très étudié au carrefour de la combinatoire des mots et de la coloration de graphes) ; un résultat quasi optimal sur les schémas d'adjacence dans les graphes planaires, concluant une ligne de recherche entamée en 1986 ; un graphe à $n^{1+o(1)}$ sommets et arêtes qui contient comme sous-graphes induits tous les graphes planaires à au plus n sommets. Ce résultat améliore de manière substantielle la borne précédente de $O(n^{3/2})$ (Babai, Chung, Erdős, Graham, Spencer, 1982).

Prospective

La théorie structurelle et algorithmique des graphes orientés (au contraire de celle des graphes non orientés) est encore balbutiante en dépit d'applications potentielles. Il en va de même de la théorie des graphes définis par sous-graphes induits exclus, pour lesquels une compréhension globale reste à construire bien que des familles de graphes essentielles soient définies de cette manière (les graphes parfaits, les graphes sans-griffe, etc...). Une piste pour y arriver pourrait passer par la théorie des mineurs de matroïdes.

On peut aussi mentionner un domaine largement représenté dans la communauté française, celui de la « reconfiguration discrète » qui pose des questions du type suivant : étant donné un ensemble de solutions (par exemple les k-colorations propres des sommets d'un graphe), est-il possible de passer d'une solution quelconque à toute autre par une suite d'opérations élémentaires (par exemple une suite d'échanges de couleurs le long d'une chaîne de Kempe) ? Combien de telles opérations élémentaires sont-elles nécessaires ? Combien de composantes connexes dans l'espace des solutions ?... Récemment

trois conjectures de Las Vergnas et Meyniel portant sur la recoloration ont été réfutées, on peut espérer de nombreuses avancées dans ce domaine au cours des prochaines années.

Les algorithmes avec garantie d'approximation sont devenus un champ de recherche qui synthétise les idées de la combinatoire extrémale et par modules. Ils utilisent l'optimisation combinatoire exacte. Les méthodes de différents domaines mathématiques y trouvent des applications à la fois utiles, subtiles et esthétiques. Sur ce sujet, un excellent livre par Williamson et Shmoys a paru en 2011. Il ne contient bien sûr pas les résultats les plus récents mais il a certainement stimulé les progrès considérables de ces dernières années. Plusieurs livres récents [6] montrent l'explosion du sujet, due probablement à l'aboutissement de méthodes qui ont progressé depuis des décennies.

Remarquons enfin qu'il existe en théorie des graphes un grand nombre de problèmes ouverts anciens sur lesquels bien des chercheurs aiment toujours se pencher en parallèle avec les « thèmes émergents », dans l'espoir de faire apparaître de nouvelles connections ou de leur appliquer de nouveaux outils ou de nouvelles techniques...

6. Par exemple : Vijay Vazirani, *Approximation Algorithms*, Springer-Verlag, 2004 ; David Williamson and David Shmoys, *The Design of Approximation Algorithms*, Cambridge University Press, 2011 ; et Marek Cygan, Fedor V. Fomin, Daniel Marx, Saket Saurabh, Lukasz Kowalik, Daniel Lokshtanov, Marcin Pilipczuk, *Parameterized Algorithms*, Springer, 2015.

Algorithmique

Contours

Les paragraphes précédents l'ont montré, et c'est d'ailleurs l'évidence, les algorithmes, l'algorithmique sont partout présents en informatique mathématique et bien au-delà. Nous tentons plutôt ici de donner un point de vue transverse sur le développement de l'algorithmique, sans rechercher l'exhaustivité.

Quels problèmes sont étudiés par l'algorithmique ? Pour l'essentiel, ces problèmes évoluent peu. Ce sont les mêmes problèmes fondamentaux — de théorie des graphes, de recherche opérationnelle, d'optimisation combinatoire, de combinatoire des mots, de géométrie, ou numériques, identifiés pour la plupart dès les années 1950 — qui ne cessent d'être explorés par les algorithmiciens. Constamment revisités, ils continuent à nous fournir de la matière pour mieux comprendre, *in fine*, ce que c'est que calculer. Il suffit, pour s'en convaincre, de regarder les papiers primés dans les éditions de FOCS, STOC, ICALP(A), SODA des dernières années : de nombreux résultats, probablement la majorité, sur des problèmes algorithmiques dans les graphes et réseaux (couplages, ensembles stables, coupes, flux, chemins disjoints, voyageur de commerce, isomorphisme, diamètre, coloriage) ; et pour le reste, des résultats épars sur distance d'édition sur les mots, le comptage des bases d'un matroïde, la résolution de systèmes d'équations creux, l'accessibilité dans les réseaux de Petri ou la couverture d'objets convexes.

Comment peut-on continuer à travailler sur les mêmes problèmes pendant des dizaines d'années ?

Plus on creuse un problème donné, plus le travail devient difficile. De fait, on a vu le domaine mûrir, avec en corollaire l'allongement des études nécessaires pour apporter des contributions substantielles à la recherche algorithmique. Occasionnellement arrivent des avancées sur des problèmes anciens, signes de cette maturité. Un exemple de cette catégorie est donné ci-dessous parmi les avancées récentes. Cela demande des progresser sur les « fondements » : d'où l'importance des mathématiques sous-jacentes à la résolution des problèmes fondamentaux traités par l'algorithmique. Il s'agit en quelque sorte d'ouvrir la boîte noire que peuvent représenter les probabilités et les statistiques, les mathématiques de l'optimisation, ou celles de la théorie des graphes et des matroïdes, de la programmation linéaire ou semi-définie, etc. On veut alors comprendre plus en profondeur ces outils mathématiques afin, par exemple, de tirer parti des spécificités ou des symétries des problèmes à résoudre. Les liens avec la combinatoire entrent aussi dans cette catégorie.

Plus souvent, les problèmes étudiés sont en fait des variantes de nos problèmes anciens : *beyond worst case* pour des cas particuliers de données d'entrée (des sous-classes de graphes par exemple ; voir aussi l'analyse probabiliste des algorithmes mentionnée dans la section sur l'Intelligence Artificielle), avec des changements dans le modèle de calcul, ou en adoptant une autre perspective qui modifie objectifs ou contraintes. De telles évolutions sont généralement inspirées par les évolutions technologiques. Ainsi, le développement des travaux en réseaux a naturellement redonné de l'importance à l'algorithmique distribuée. L'avènement de données massives a inspiré le modèle de calcul en

streaming dans lequel les problèmes anciens doivent être réétudiés, mais aussi un renouveau d'intérêt pour la complexité des algorithmes, en ce sens qu'on souhaite progresser non seulement dans la conception d'algorithmes de complexité polynomiale, mais aussi, vers le bas de l'échelle des complexités en temps, pour obtenir des complexités qui sont des polynômes de degré aussi faible que possible. L'impératif de réactivité face aux inévitables modifications des données ou réseaux a servi de tremplin au développement des algorithmes dynamiques. Le faible nombre attendu de changements, d'erreurs, ou d'anomalies, a pu guider le concept de complexité paramétrée. Les applications sociétales des algorithmes ont soulevé de nouvelles questions d'équité et ont redonné du lustre au concept de vérification comme outil de transparence. Ainsi, ce sont les changements des usages du numérique dans le monde qui induisent des changements dans les questions algorithmiques à étudier.

Enfin, quelquefois arrive un problème entièrement nouveau. De nos jours, on ne peut pas dessiner les contours de l'algorithmique sans évoquer les algorithmes d'intelligence artificielle. D'une certaine manière, au moment de l'explosion soudaine de l'efficacité du paradigme d'apprentissage profond, la communauté algorithmique a été prise au dépourvu devant un algorithme efficace en pratique mais sans garantie d'efficacité démontrée.

Avancées récentes

Il est difficile de faire un choix parmi tous les résultats récents en algorithmique, même en se restreignant à ceux qui ne concernent pas un domaine spécifique décrit ailleurs dans ce document. C'est donc avec un peu d'arbitraire que nous braquons le projecteur sur quelques domaines où l'activité est intense.

Un exemple de problème ancien qui a connu des progrès récents est celui du voyageur de commerce (cas métrique général). Depuis les années 70, l'algorithme de Christofides donnait une approximation à un facteur $3/2$ de l'optimum. Récemment a été développé un algorithme avec une approximation très légèrement meilleure : $3/2 - 10^{-36}$ (best paper, STOC 2021). Cette amélioration est quantitativement symbolique mais elle fait suite à une série d'articles apportant de nouveaux éléments de compréhension à différentes versions du voyageur de commerce. C'est un tour de force technique qui repose d'une part sur la structure combinatoire de coupes quasi-minimum, et d'autre part sur une généralisation de théorèmes d'échantillonnage probabiliste pour faire de la génération aléatoire de certains arbres couvrants presque minimaux.

Dans l'étude des complexités en temps très faibles, on peut citer un résultat récent (best paper, FOCS 2022) sur le calcul de plus courts chemins à partir d'une source lorsque les arêtes peuvent avoir des poids négatifs. La complexité est linéaire à des facteurs logarithmiques près. Ce résultat a émergé de manière indépendante dans deux travaux, l'un utilisant des idées empruntées à l'optimisation continue et l'autre, de façon peut-être moins générale mais plus simple, par le biais d'une nouvelle décomposition de graphes. La coïncidence de ces deux travaux montre l'intérêt porté par la communauté à ce type de questions.

En termes de concepts, en France se développe actuellement une théorie de la « twin-width ». Les graphes avec largeur d'arborescence bornée ont nécessairement une « twin-

width » bornée, donc la classe des graphes avec « twin-width » borné englobe davantage de graphes, et on peut ambitionner de construire une théorie des graphes avec « twin-width » bornée, similaire à celle développée autrefois pour les graphes de largeur arborescente bornée. Ce développement inclut naturellement tous les aspects algorithmiques, et peut se positionner dans le cadre de la *complexité paramétrée* (où le paramètre supplémentaire est la « twin-width » du graphe).

L'algorithmique interagit depuis longtemps avec les sciences économiques, des conférences y sont consacrées, par exemple *Economic Design and Algorithms* (St Petersburg 2019). Cette interaction fait maintenant évoluer profondément certaines problématiques de l'algorithmique et amène des interrogations toujours plus profondes, comme dans le best paper FOCS 2016 sur le calcul efficace d'équilibres de Nash approchés. Dans le domaine de l'économie aussi, mais avec des applications qui débordent largement les frontières de ce domaine, on notera aussi les résultats à haute visibilité sur les problèmes algorithmiques posés par les technologies de blockchain, y compris leur coût énergétique et leur devenir à long terme.

D'autres avancées à fort impact concernent les fondations mathématiques de la science des données et de l'apprentissage : il s'agit là le plus souvent de résultats obtenus conjointement par des algorithmiciens, des statististiciens et des mathématiciens, qui explorent par exemple la façon dont les algorithmes d'apprentissage automatique fonctionnent, comment ils peuvent être améliorés du point de vue de leur efficacité, ou comment on peut comprendre et évaluer (un jour garantir) la qualité de leurs résultats. Ainsi un best paper de FOCS 2016 établit que, pour certains problèmes d'apprentissage, on ne peut se passer d'un grand espace mémoire. Des avancées de même nature peuvent être observées dans le domaine de la science des données.

Prospective

De nouveaux domaines de recherche et de réflexion s'ouvrent en ce moment en algorithmique, parallèlement à l'approfondissement des aspects décrits dans les sections précédentes. Certains concernent la cryptographie, le quantique, la science des données ou l'IA et sont décrits ailleurs dans ce document. Dans ce paragraphe, nous allons mettre en avant, de façon non exhaustive, deux grandes directions bien distinctes, qui nous paraissent à la fois prometteuses et représentatives de la diversité de ce que recouvre l'algorithmique aujourd'hui.

« Green computing ». Il faut un changement de société pour faire face au défi du dérèglement climatique ; tous les domaines sont concernés, et l'algorithmique n'est pas exemptée. Cherchant à s'éloigner d'une algorithmique plus abstraite, les communautés concernées (voir par exemple le *SIAM Symposium on Algorithmic Principles of Computer Systems*, APOCS) s'attachent à reconnecter les algorithmes classiques avec les machines réelles, pour en améliorer la rapidité ou la frugalité énergétique en tirant parti des spécificités de l'architecture des systèmes. (Voir aussi l'évaluation de consommation énergétique mentionnée dans la section Intelligence Artificielle.) À ce sujet, on peut regretter la quasi-absence dans les conférences purement algorithmiques de questions en

lien avec les systèmes d'exploitation ou avec les bases de données.

Les algorithmes dans la cité. L'impact social des algorithmes est au cœur des préoccupations de nouvelles conférences. Par exemple, l'*ACM Conference on Fairness, Accountability, and Transparency* (ACM FAT) se donne pour mission de rassembler des spécialistes d'algorithmique (théoriciens et praticiens) et des spécialistes d'autres disciplines, singulièrement des SHS, pour discuter des questions d'équité, de responsabilité et de transparence des algorithmes, aussi bien pour améliorer les algorithmes qui peuvent déjà être déployés que pour poser les bases d'une formalisation de ces notions en termes sociaux ou légaux. On retrouve ces préoccupations dans un nombre croissant de communautés, notamment en science des données et en IA.

Intelligence artificielle et apprentissage

La seule mention de l'intelligence artificielle ou de l'apprentissage suscite souvent des débats enflammés, à la fois scientifiques et politiques. Le but de cette courte section n'est pas de trancher ces débats mais de donner quelques éléments pour préciser ce que les activités du GdR Informatique Mathématique peuvent apporter en la matière.

Tout d'abord, si l'IA s'inscrit historiquement dans le périmètre de l'informatique, et même de sa composante la plus fondamentale (inférence de règles, programmation logique), ses développements plus récents ont d'abord émergé dans d'autres communautés, en particulier celles des statisticiens et du traitement du signal. Un GdR spécifique est consacré à ce domaine, le GdR RADIA, qui lui-même interagit avec le GdR IM mais aussi les GdR ISIS (signal et image), TAL (traitement automatique des langues), Robotique, RO (recherche opérationnelle), BIM (bio-informatique), MADICS (masses de données), et avec les très nombreux domaines d'application de l'IA [7].

Cela étant, l'IA est bien présente dans les activités du GdR IM, à travers notamment le travail sur les modèles qui la sous-tendent (classification, apprentissage profond, apprentissage par renforcement), sur des questions fondamentales comme celles de l'explicabilité ou de la certification des algorithmes d'IA, et sur les techniques qui permettent son implémentation.

Une des premières implications de notre communauté scientifique a été de se pencher sur la formalisation des propriétés des algorithmes d'apprentissage. De nouveaux champs de recherche florissants sont nés autour de la recherche de garanties des performances et de l'« explicabilité » des algorithmes d'IA pour lutter contre une approche empirique de type boîte noire. La *certification* de logiciels d'IA ou d'apprentissage joue un rôle crucial dans de nombreuses applications, en particulier lorsque ces logiciels sont utilisés dans le cadre de systèmes critiques. L'*explicabilité*, c'est-à-dire la capacité d'un logiciel à donner les éléments qui justifient la solution qu'il apporte, est quant à elle essentielle, entre autres choses pour l'acceptabilité sociétale de solutions basées sur l'intelligence artificielle.

Ces deux questions représentent des défis scientifiques considérables, qui impliquent en particulier les communautés spécialistes de la logique, des modèles de calcul et du traitement des données, en collaboration avec les communautés qui conçoivent les modèles de connaissance sous-jacents (mathématiciens, statisticiens, spécialistes d'optimisation combinatoire) et celles qui sont à l'origine des données traitées (santé, ingénierie, etc). La situation à cet égard est différente selon les techniques de l'IA utilisées, avec des résultats plus avancés — c'est-à-dire la possibilité d'une « confiance » plus poussée — pour les logiciels qui s'appuient sur la classification ou l'apprentissage par répartition.

La quête de l'explicabilité passe aussi par un travail sur la conception même de langages de description des données et de programmation pour les algorithmes d'apprentissage, une nouvelle déclinaison de la thématique traditionnelle de la sémantique des langages de programmation dans le contexte de l'IA.

Un autre défi est celui de l'« équité » dans les algorithmes d'apprentissage et dans

7. Les applications de l'IA sont de plus en plus nombreuses, et posent des questions éthiques, sociales et environnementales très importantes, qui sortent cependant du cadre de ce document.

l'implémentation des applications de l'IA : il s'agit de mesurer a priori les biais de discrimination de certains logiciels d'apprentissage à travers le prisme de divers modèles d'« équité », pour pouvoir ensuite minimiser ces biais.

La communauté informatique mathématique est également motrice dans la mise en œuvre ou l'analyse de certains des développements de l'IA moderne. Il en va ainsi, par exemple de l'analyse théorique des algorithmes de l'IA ou de l'apprentissage, notamment l'évaluation a priori de leur rapidité mais aussi de leur consommation énergétique, les deux points étant des goulots d'étranglement très importants. Le domaine de l'analyse probabiliste des algorithmes joue ici un rôle central. De même, en calcul formel, l'amélioration des performances des algorithmes de différentiation automatique a un impact direct sur les performances des algorithmes d'apprentissage. On citera aussi l'importance de l'étude des systèmes dynamiques discrets pour la résolution de problèmes d'optimisation en apprentissage par renforcement.

Plus profondément, bien que l'apprentissage automatique se soit développé indépendamment de l'informatique dite théorique, l'ensemble des méthodes et des savoirs de cette dernière pourrait être mis à contribution pour contourner le mur que représente le coût actuel de l'apprentissage, en termes de la quantité d'exemples qui doivent être traités et de la quantité d'énergie nécessaire à cet effet — par exemple en s'appuyant sur des systèmes hybrides, combinant systèmes discrets et apprentissage automatique.

Il est intéressant enfin de noter que certains domaines de l'informatique mathématique sont impactés en retour par la montée en puissance de l'IA dans l'industrie et les services. Par exemple, les nouveaux processeurs conçus pour être efficaces pour les types de calculs employés en intelligence artificielle, ont des unités de calcul avec peu de bits de précision, et cela pose de nouvelles questions dans le domaine de l'arithmétique des ordinateurs. De même, pour garantir la sécurité des données, la recherche en cryptographie s'attache désormais à fournir des outils qui permettent aux algorithmes d'apprentissage de travailler « en aveugle » sur des bases de données sensibles, sans jamais y accéder en clair.

L'informatique mathématique et l'industrie

Loin des clichés sur l'opposition entre informatique fondamentale et informatique appliquée, les liens entre les différents domaines de l'informatique mathématique et le secteur industriel ou celui des services sont anciens et solides, qu'il s'agisse de projets menés en commun ou du recrutement de docteurs.

La présence ou non d'acteurs industriels forts en France et en Europe a bien sûr un impact sur nos domaines. Ainsi, la modélisation géométrique bénéficie de la présence de Dassault Systèmes, leader mondial industriel pour la CAO, ainsi que de nombreux industriels de la CAO ouverts aux partenariats (EADS/Airbus group, EDF, constructeurs automobiles).

En cryptographie ainsi qu'en sécurité, la situation est relativement similaire et on constate de nombreuses interactions aussi bien avec les grands groupes civils ou militaires comme Orange, Thalès ou Enedis, qu'avec des PME comme Quarkslab, Amossys ou Trust in Soft, ou encore avec des organisations gouvernementales comme l'ANSSI, le CEA ou la DGA.

Les chercheurs en théorie des graphes et optimisation combinatoire orientés vers la Recherche Opérationnelle ont d'intenses collaborations avec de grandes compagnies dans le domaine aéronautique (notamment Amadeus) et dans celui des télécommunications (Nokia, Orange).

La partie « Matériel » de l'architecture des ordinateurs profite de la présence sur le territoire d'acteurs importants comme ST Microelectronics ou Kalray, mais les principaux acteurs mondiaux (Intel, AMD, ARM, etc.) sont à l'étranger.

Les liens avec l'industrie se matérialisent également par l'implication des membres du GdR dans les processus de standardisation internationaux de plusieurs domaines, comme les compétitions cryptographiques du NIST, ou l'élaboration des normes IEEE 754 et 1788 sur les nombres flottants et l'arithmétique d'intervalles.

Il faut souligner enfin que, depuis plus d'une décennie, les GAFAM (Google/Alphabet, Microsoft Research, Facebook/Meta en particulier) ont installé en région parisienne des instruments de collaboration avec la recherche informatique française, en particulier avec la communauté informatique mathématique.

La situation de l'emploi des docteurs dans l'industrie et dans la haute administration — en général aussi bien que dans nos spécialités — fait l'objet de nombreux débats. Il se situe à un niveau très bas par rapport à ce qu'on peut observer dans des pays comparables, et les évolutions que l'on peut constater sur le terrain sont lentes en général.

Mais dans certains domaines d'activité, la situation a connu un bouleversement ces toutes dernières années. En cryptographie, pour tout ce qui concerne la science des données ou le calcul quantique, mais aussi sur certains aspects de la cybersécurité, de la vérification, et des fondements de la programmation, l'industrie recrute des chercheurs en grande quantité. Le phénomène n'est pas entièrement nouveau mais le flux de chercheurs du monde académique vers celui des laboratoires de recherche privés s'est singulièrement intensifié, et concerne aussi bien les jeunes docteurs que des chercheurs confirmés. C'est certainement la marque de la pertinence de nos formations et de l'importance de nos

disciplines, mais la rapidité de cette évolution a parfois pour effet de rendre difficile non pas tant le recrutement de doctorants que le maintien de ressources suffisantes pour les encadrer. Quoi qu'il en soit, le mouvement est rapide et intéressant, à l'initiative de start-ups mais aussi d'un petit nombre de grandes entreprises françaises et, très souvent, d'entreprises internationales.

En guise de conclusion très partielle

Nous l'avons souligné dès l'introduction de ce document et chacune de ses sections l'a confirmé, les frontières du GdR Informatique Mathématique avec les autres branches de l'informatique ou des mathématiques sont souvent poreuses, et parfois mobiles. Des exemples frappants sont donnés par les relations entre l'informatique mathématique et la sécurité, articulées notamment autour d'un groupe de travail commun aux deux GdR concernés ; ou par celles entre l'informatique mathématique et l'IA, là aussi à travers un groupe de travail commun à deux GdR mais aussi du fait de la réalisation qu'un nombre croissant de domaines de l'informatique mathématique sont utiles en IA.

Ces quelques exemples, la sécurité, l'IA, on aurait pu parler aussi du calcul quantique ou de la cryptographie, illustrent la centralité de l'informatique mathématique dans le développement de l'informatique et, parfois, des mathématiques, sur des sujets de société dont l'importance ne fait pas débat. Elles illustrent aussi tout l'intérêt de cette porosité des frontières du GdR.

À côté de ces considérations sur la dynamique qui anime l'informatique mathématique et le GdR IM dans leurs interactions avec les domaines connexes, il faut aussi souligner que la communauté scientifique rassemblée par le GdR a une forte unité, parce que ses différentes composantes s'appuient sur un substrat scientifique commun et, peut-être encore plus important, sur des méthodes communes. Cette communauté se retrouve volontiers dans les structures du GdR et dans les manifestations qu'il organise, et elle s'y reconnaît d'autant mieux qu'il n'existe pas d'autres structures de cette granularité où elle pourrait se développer.

La communauté des chercheurs en informatique mathématique enfin est, comme beaucoup d'autres, traversée par les débats qui intéressent nos sociétés dans leur ensemble. Sur le sujet crucial de la transition énergétique, on observe un intérêt très vif pour les questions liées au coût énergétique des algorithmes déployés dans tous les domaines, et en particulier dans celui de l'apprentissage. L'informatique mathématique devrait pouvoir apporter une contribution significative au développement d'une théorie solide du coût énergétique des algorithmes, voire même à l'objectif plus ambitieux de l'élaboration d'une analyse informatique de la correspondance entre information et énergie.