Sujet de thèse - CRIStAL - 2021

Titre : Calcul efficace de séries de Puiseux

— Équipe : CFHP `cfhp.univ-lille.fr`. Directeur François Boulier.
— Directeur de thèse : Pierre Fortin, `pierre.fortin@univ-lille.fr`.
— Co-encadrant : Adrien Poteaux, `adrien.poteaux@univ-lille.fr`.
— Laboratoire : CRIStAL, Lille (France). Directeur Olivier Colot, `cristal.univ-lille.fr`.

Mots clés : géométrie algébrique, singularités, calcul formel, calcul haute performance, parallélisme.

1 Résumé.

L'objectif de cette thèse sera de travailler sur l'implémentation efficace d'algorithmes de calcul de séries de Puiseux d'un polynôme bivarié $F(x, y)$ (i.e. les séries $S(x)$ t.q. $F(x, S(x)) = 0$). Les enjeux sont multiples : d'une part une amélioration algorithmique pour le calcul des coefficients des séries quand l'on souhaite calculer un nombre « modéré » de coefficients ; d'autre part un déploiement HPC (calcul haute performance) des algorithmes (actuellement théoriques) développés [5, 6].
Ce sujet de thèse s'adresse prioritairement à des étudiants ayant des connaissances en Mathématiques et en Informatique (le travail effectué pouvant demander d'assimiler un certain nombre de connaissances mathématiques théoriques, et la partie HPC demandant des bases en algorithmique et en programmation). Un étudiant venant de l'un ou l'autre des domaines est tout à fait possible, mais une aversion pour l'un des deux domaines serait problématique : le sujet est clairement à l'intersection des deux domaines.

2 Contexte.

L'étude de singularités (parfois décrites comme des « points infiniment proches ») de variétés algébriques et analytiques est un domaine de recherche ancien (l'étude de singularités de certaines courbes apparaît dans les problèmes étudiés pas les géomètres grecs) et actif. C'est d'ailleurs devenu une discipline à part entière depuis les années 1960, à partir des travaux d'Hironaka, Zariski etc.
Différents algorithmes ont été développés pour étudier des singularités. En particulier, le moyen le plus utilisé pour « résoudre » une singularité est de passer d'une courbe algébrique à une autre par des changements de variables bien choisis, afin d'obtenir au final une courbe non singulière. Ces méthodes peuvent se résumer par le calcul des *séries de Puiseux*, dont les premiers termes contiennent les éléments de cette résolution (que l'on appelle la partie singulière ; le nombre minimum de termes pour « séparer » les séries l'une de l'autre). Les séries de Puiseux sont omniprésentes dans le domaine des courbes algébriques, que ce soit théoriquement ou pour une utilisation pratique : elles permettent de calculer le genre d'une courbe via la formule d'Hurwitz, des bases intégrales d'un corps de fonction [7], une base pour l'espace de Riemann-Roch $\mathcal{L}(D)$ associé à un diviseur D (via l'algorithme de Dedekind-Weber's [2]) ; elles donnent naturellement toute l'information locale de la courbe (permettant d'obtenir une approximation des valeurs de fonctions algébriques quand le corps de base est un sous corps de $\mathbb{C}$), etc.

L'algorithme le plus connu pour ce faire est celui de Newton-Puiseux, dont de nombreuses variantes existent. Le meilleur résultat de complexité pour calculer les parties singulières a été obtenu récemment [5] : elle est de l'ordre de d^3, l'entrée étant de taille d^2 (on peut donc qualifier cet algorithme de rapide, car sous-quadratique).
Néanmoins, il n'existe pas à ce jour d'implémentation efficace d'un tel algorithme : celui développé dans [5] reste théorique (de nombreuses « boîtes noires » sont théoriquement bien connues mais non implémentées à ce jour). De plus, cet algorithme n'a pas une complexité aussi fine si l'on souhaite un peu plus de termes que la partie singulière (par exemple, si l'on souhaite travailler avec une précision n - i.e. que l'on souhaite les n premiers termes de chaque série avec n « modérément grand », l'algorithme va être capable de factoriser le polynôme avec une complexité de l'ordre de $d\,n$, mais le calcul des séries sera lui d'ordre $d^2\,n$). Il y a ainsi encore des progrès pratiques et théoriques à obtenir dans le cadre du calcul de séries de Puiseux.

3 Travail envisagé.

L'objectif de la thèse est double :

1. améliorer les algorithmes existants pour calculer les n premiers termes de chaque série ;
2. mettre en oeuvre une implémentation efficace du calcul de séries de Puiseux.

3.1 Calculer « un peu plus » que la partie singulière

Que ce soit pour les applications de calcul de bases intégrales ou de bases d'espace de Riemann-Roch, connaître la partie singulière des séries de Puiseux est nécessaire mais pas suffisant. Ces algorithmes nécessitent de connaître un « petit » nombre de termes supplémentaires. De part le phénomène d'éclatement inhérent à l'algorithme de Newton-Puiseux (changement de variable de type $x \leftarrow x^q$ qui augmente la taille des données intermédiaires), le meilleur résultat de complexité connu est de l'ordre de $d^2\,n$ [5, Proposition 7] quand d et n sont d'ordre similaire ; si n est très grand par rapport à d, alors il existe des méthodes en $d\,n$ et un surcoût ne dépendant que de d.
Le point bloquant est clairement le phénomème d'éclatement ([5] donne en particulier une famille d'exemples pour laquelle les bornes de complexité sont atteintes). Pour contourner ce problème d'éclatement, Abhyankar a proposé une méthode pour tester si un polynôme est irréductible utilisant les *racines approchées* [1] (ce sont des polynômes bivariés ψ_k associés à un entier k divisant d tel que $\psi_k^{d/k}$ soit le polynôme « le plus proche » de F). Partant de cette idée, un nouvel algorithme rapide de test d'irréductibilité quasi optimal a été récemment développé [6]. Ce dernier permet également de calculer les *termes essentiels* des séries de Puiseux (i.e. les termes définissant le type mathématique de singularité), mais pas les autres termes (ceux n'ayant pas d'impact d'un point de vue théorie des singularités, mais étant nécessaires aux applications évoquées précdemment).
Néanmoins, ces termes peuvent être lus sur les coefficients des racines approchées, par exemple en résolvant un système linéaire. Une telle méthode reste trop coûteuse (le système est de taille n^2). L'objectif de la partie théorique de cette thèse sera de chercher un moyen de calculer ces termes en temps quasi linéaire en n (l'idée étant de définir une équation univariée à partir des coefficients de la racine approchée dont un polynôme correspondant aux termes cherchés est solution ; il existe ensuite de nombreuses méthodes permettant de trouver les coefficients en temps linéaire en n à partir d'une telle équation).

3.2 Déploiement HPC

En plus de la conception de ces algorithmes efficaces adaptés au calcul des séries de Puiseux, nous souhaitons accélérer leur implémentation sur un serveur de calcul moderne dotés de processeurs multi-cœurs. Plusieurs niveaux de parallélisme pourront être visés via une algorithmique et une programmation adéquates.
Au niveau multi-cœur, on pourra en effet introduire du parallélisme au sein des opérations élémentaires (multiplications de polynômes, itération de Newton, PGCD de polynômes, factorisation de polynômes). Mais on pourra aussi viser un parallélisme présent à un plus haut-niveau dans l'algorithme (traitement en parallèle des développements de Taylor, parallélisation d'une factorisation partielle). Le parallélisme de tâches pourra être particulièrement adapté ici pour imbriquer efficacement ces différents niveaux de parallélisme, et aussi éventuellement pour augmenter le degré de parallélisme disponible (voir par exemple [4]).
Au niveau de chaque cœur de calcul, on pourra par ailleurs chercher à exploiter les instructions vectorielles (AVX et AVX-512) qui peuvent offrir un gain en performance très important, comme montré récemment dans le cadre de l'arithmétique modulaire [3].

Compétences souhaitées : un goût pour les mathématiques (en particulier l'algèbre) et l'informatique (algorithmique et programmation ; des bases en calcul haute performance sont bienvenues mais non nécessaires).

Références

[1] S. S. Abhyankar. Irreducibility criterion for germs of analytic functions of two complex variables. *Advances in Mathematics*, 74(2) :190 – 257, 1989.

[2] D. Duval. Diverses questions relatives au calcul formel avec des nombres algebriques, 1987. Thèse d'État.

[3] P. Fortin, A. Fleury, F. Lemaire, and M. Monagan. High performance SIMD modular arithmetic for polynomial evaluation. *Concurrency and Computation : Practice and Experience (accepted with minor revision)*, 2020. https ://hal.archives-ouvertes.fr/hal-02552673.

[4] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-based augmented contour trees with fibonacci heaps. *IEEE Trans. on Parallel and Distributed Systems*, 30(8) :1889–1905, 2019.

[5] A. Poteaux and M. Weimann. Computing puiseux series : a fast divide and conquer algorithm, 2017.

[6] A. Poteaux and M. Weimann. A quasi-linear irreducibility test in $\mathbb{K}[[x]][y]$, 2019.

[7] M. van Hoeij. An Algorithm for Computing an Integral Basis in an Algebraic Function Field. *Journal of Symbolic Computation*, 18 :353–363, 1994.