Initiation à la Recherche

François Boulier

1er septembre 2021

Table des matières

Introduction générale

Ce document constitue les notes de trois cours magistraux de deux heures, d'initiation à la recherche, à Polytech'Lille, pour les étudiants en quatrième année de la filière « Informatique et Statistique ».

Le premier cours est une introduction à l'épistémologie, ou philosophie des sciences. Il aborde des questions telles que « qu'est-ce que la science ? », « comment la connaissance progresse-t-elle dans les disciplines scientifiques ? ». Il présente plusieurs théories classiques, en particulier, la théorie des « révolutions scientifiques » de Kuhn. Au delà de son intérêt intrinsèque, la théorie de Kuhn nous sert de grille de lecture pour les cours suivants.

Les théories présentées au premier cours ont été pensées pour les sciences de la nature. Le deuxième cours se concentre davantage sur les mathématiques et l'informatique. Il présente un programme de recherche et une crise qui ont eu lieu durant la première moitié du vingtième siècle, et qui ont affecté les mathématiques et l'informatique naissante. Nous aborderons plusieurs questions, en lien avec le premier cours : « cette crise a-t-elle correspondu à une révolution scientifique, au sens de Kuhn ? », « les mathématiques et l'informatique sont-elles en fait des sciences, au sens de Kuhn ? » Ce ne sont pas les réponses à ces questions qui sont importantes mais les réflexions qu'elles suscitent.

Le troisième cours est destiné à donner des outils et des méthodes pour aider les futurs chercheurs dans leur pratique de la « science normale » (au sens de Kuhn, bien sûr) en informatique. Il aborde des problèmes aussi divers que : l'écriture d'articles, le positionnement bibliographique, la structure des introductions, l'évaluation de la recherche et des chercheurs : mécanisme de sélection des articles, h-index, facteur d'impact.

Ce cours d'initiation à la recherche n'est pas destiné à fournir des idées toutes faites à ceux qui le suivent. J'espère qu'il donnera aux étudiants l'envie de lire un certain nombre de livres clefs et les amènera à développer leurs propres paradigmes.

Et bien sûr, j'espère qu'il leur donnera envie de faire de la recherche !

Avertissement : ce cours tente une mise en perspective historique de différents événements, ce qui est nécessaire à la présentation de la théorie de Kuhn. Toutefois, je n'ai pas mené un travail d'historien, mais un travail de compilateur, puisque j'ai travaillé le plus souvent à partir de sources secondaires. À plusieurs occasions, j'ai pu me rendre compte que les textes se contredisent et je crains que des affirmations fort discutables, voire franchement fausses, se soient glissées dans ce document. J'incite donc le lecteur à faire preuve de sens critique et à signaler les erreurs à `Francois.Boulier@polytech-lille.fr`.

Pour terminer, je voudrais remercier, par ordre alphabétique, Maxence Delorme, François Lemaire et Alexandre Sedoglavic, pour leurs suggestions et leur travail de relecture. Plus spécialement, je voudrais remercier Simon Boulier pour d'anciennes discussions, ainsi que Jean-Paul Delahaye pour ses suggestions, pour m'avoir autorisé à utiliser certains des articles qu'il a écrits dans *Pour La Science*, et pour avoir organisé, au début des années 90, les séminaires *Histoire et Philosophie de la Logique et du Calcul*, qui ont inspiré certaines parties de ce document.

Chapitre 1

Une introduction à l'épistémologie

Dans ce premier cours, on introduit plusieurs théories de la progression des sciences : l'inductivisme, le falsificationisme et la théorie des révolutions scientifiques de Kuhn. L'accent est mis sur cette dernière théorie. On la met en perspective en décrivant les grandes lignes de la révolution copernicienne (l'exemple majeur de Kuhn) et de la théorie de l'apprentissage chez l'enfant, de Piaget. Les livres importants liés à ce cours sont ceux d'Alan Chalmers [3] et de Thomas Kuhn [10, 9]. La progression est fortement inspirée de celle d'Alan Chalmers. Elle présente l'avantage d'amener par touches successives, un certain nombre d'exemples historiques, nécessaires à l'exposé de la théorie de Kuhn.

1.1 Introduction à l'inductivisme

Il y a de nombreuses variantes de l'inductivisme. Celle qui est présentée ci-dessous, fortement inspirée de [3, chapitre 1, § 2], est caricaturale bien qu'on puisse trouver des textes qui s'en rapprochent [3, citation, page 34].

1.1.1 La théorie

Selon l'inductiviste naïf, la science commence par l'observation. Un scientifique pourrait procéder ainsi :

1. procéder à de nombreuses expériences, en chauffant (par exemple) de nombreux types de barres métalliques, et en variant les conditions expérimentales pertinentes (métaux différents, barres longues ou courtes ...),
2. observer qu'elles se sont dilatées, à chaque expérience,
3. en déduire une loi générale : « toutes les barres métalliques se dilatent lorsqu'elles sont chauffées ».

Les lois générales, déduites des observations, permettent d'expliquer et de prédire (les lois générales qui régissent les mouvements des planètes, permettent de prédire la prochaine éclipse).

Il suffit qu'une observation contredise une loi pour réfuter cette loi (et donc la refuser en tant que loi).

Condition pour qu'une discipline soit une science. Une discipline est scientifique si elle produit des lois générales à partir d'observations, en appliquant la méthode inductive.

Progression continue des sciences. La connaissance scientifique progresse continûment, puisqu'elle s'appuie sur un ensemble d'observations qui augmente avec le temps.

1.1.2 Critiques

La principale critique qu'on peut faire à l'inductivisme est une critique qui s'appuie sur l'histoire des sciences : historiquement, les grands progrès scientifiques se sont appuyés sur des raisonnements complètement incompatibles avec l'inductivisme. On développe ce point dans la section consacrée à la théorie de Kuhn.

On poursuit ci-dessous avec quelques autres critiques, de nature plus logique, où l'inductivisme est combattu sur son propre terrain, si j'ose dire. Bien que ces critiques soient très fortes, il faut, comme le demande justement Chalmers [3, page 71], relativiser toutes celles qui s'appliquent aussi aux théories concurrentes.

Critique logique. Du point de vue purement logique, on ne peut pas démontrer de loi générale à partir d'un ensemble fini d'énoncés d'observation [1]. La seule justification logique de l'inductivisme repose sur un raisonnement inductiviste, ce qui est inadmissible d'un point de vue logique [2]. Ce ne sont toutefois que des critiques un peu secondaires parce qu'elles ne touchent qu'une forme particulièrement caricaturale de l'inductivisme.

Critique de la fiabilité des énoncés d'observation. On critique ici l'idée que les énoncés d'observation puissent être considérés comme plus fiables que les théories. En fait, ils présupposent des théories.

Une phrase comme « le faisceau d'électrons est repoussé par le pôle de l'aimant » suppose une théorie considérable. Les énoncés d'observation sont formulés dans le langage d'une théorie et sont aussi précis que la théorie qu'ils utilisent [3, page 61].

Comme ils s'appuient sur des théories, qui sont faillibles, ils peuvent être faux. Du temps de Copernic (avant la lunette), on considérait comme une observation établie de tous temps que la taille apparente de Vénus ne change pas au cours de l'année, ce qui contredisait la théorie de Copernic. Cet énoncé d'observation s'appuie sur la théorie, fausse, que l'œil humain évalue correctement la dimension de petites sources lumineuses [3, page 65].

1. Le mauvais exemple (mauvais parce que non historique) de la « dinde inductiviste » de Bertrand Russell [3, page 40].
2. Critique de Hume, au XVIIIème siècle [3, page 41].

Critique au sujet de la variation des paramètres pertinents. C'est une variante de la critique précédente : pour déterminer quels sont les paramètres pertinents à faire varier lors des expériences, il faut aussi disposer d'une théorie.

Alan Chalmers donne à ce sujet un exemple historique frappant [3, page 67] : en 1888, Hertz effectue l'expérience qui lui permit d'être le premier à produire et à détecter les ondes radio. S'il avait noté tous les paramètres, il aurait dû noter les couleurs de ses chaussures, les dimensions du laboratoire, le jour de la semaine ... Il se trouve que, parmi tous ces détails apparemment sans intérêt, la dimension du laboratoire est un paramètre pertinent : Hertz essayait en fait de vérifier la théorie de Maxwell qui impliquait que la vitesse des ondes radio était la même que celle de la lumière. À plusieurs reprises, les mesures ont contredit le résultat attendu. La raison n'a été découverte qu'après la mort de Hertz : les ondes radio, réfléchies par les murs de son laboratoire, interféraient dans les mesures [3].

1.1.3 Apprentissage automatique

Le principe de l'induction (si un grand nombre de A ont été observés dans des conditions variées et si tous ces A ont la propriété B, alors tous les A ont la propriété B), qui est à la base de l'inductivisme, est fortement critiqué.

Pourtant, il connaît, en informatique, un succès spectaculaire dans le cadre de l'apprentissage automatique [4]. Une différence importante avec l'inductivisme tient au fait que l'apprentissage se fait dans un contexte formel bien défini. Les critiques faites au principe de l'induction ont poussé les chercheurs du domaine à mieux préciser les hypothèses sous lesquelles fonctionnent leurs méthodes.

On peut, peut-être, faire le parallèle avec la théorie des langages développée par Noam Chomsky, initialement conçue pour les langues naturelles, qui a connu d'importantes applications aux langages formels (dont font partie les langages de programmation), en informatique.

1.2 Le falsificationisme

Le falsificationisme est la théorie défendue par Karl Popper [13]. L'expression « théorie falsifiable » est couramment employée et mérite d'être connue. Pour le falsificationiste, la théorie précède l'observation. Le falsificationisme donne non seulement une condition pour qualifier une théorie de scientifique mais il donne aussi un critère pour affirmer qu'une théorie est meilleure qu'une autre. Dans ses versions plus sophistiquées, il accorde de l'importance aux considérations historiques, ce qui est important puisque la critique principale de Kuhn porte sur cette question.

3. Je ne suis pas parvenu à vérifier cette anecdote.

4. Les algorithmes de reconnaissance de la parole sont fondés sur des techniques d'apprentissage automatique, pour la plupart. Ce sont des algorithmes qui s'améliorent au fur et à mesure qu'on leur parle.

1.2.1 La base de la théorie

Pour le falsificationiste, une théorie scientifique n'est rien d'autre qu'une conjecture qui, une fois énoncée, doit être confrontée à l'observation et à l'expérience. Il faut éliminer les théories contredites par ces observations ou ces expériences [3, chapitre 4].

On peut montrer qu'une théorie est fausse. On ne peut jamais montrer qu'elle est vraie. Au mieux, on peut dire qu'elle est meilleure que toutes celles qui l'ont précédée.

Condition pour qu'une discipline soit une science. Le falsificationisme fournit un critère pour distinguer les théories scientifiques de celles qui ne le sont pas : pour qu'une théorie soit considérée comme scientifique, il faut qu'elle soit « falsifiable ».

Définition 1 *Une théorie est dite* falsifiable *s'il est possible de mettre au point une expérience qui puisse éventuellement aboutir à la conclusion que la théorie est fausse.*

Une théorie contredite par une expérience est dite *falsifiée* par cette expérience. Dans le cas contraire, elle est dite *confirmée* par l'expérience.

Condition pour qu'une théorie soit meilleure qu'une autre. Plus une théorie est falsifiable, meilleure elle est. Par exemple, la théorie (A) « les planètes suivent des trajectoires elliptiques autour du soleil » est meilleure que la théorie (B) « les planètes décrivent des courbes fermées autour du soleil ». En effet, toute observation qui falsifierait (B), falsifierait (A) mais on peut imaginer des observations qui falsifieraient (A) sans falsifier (B).

Progression continue des sciences. Le falsificationiste a une vision continue de la progression des sciences. Idéalement, il aimerait pouvoir dire qu'un ensemble de théories qui constituent l'évolution historique d'une science est fait de théories falsifiables, chacune étant plus falsifiable que celle qui la précède [3, page 92].

1.2.2 Une version plus sophistiquée

Les théories évoluent. En particulier, on peut être amené à les modifier afin de les protéger d'une falsification menaçante.

Définition 2 *Une modification faite à une théorie est dite* ad hoc *s'il est impossible de tester la modification indépendamment de la théorie.*

Des falsificationistes plus sophistiqués accordent de l'importance aux points suivants :

1. ils rejettent les modifications *ad hoc* des théories,
2. ils accordent de l'importance aux falsifications de théories prudentes (difficilement falsifiables) et aux confirmations de théories audacieuses (hautement falsifiables).

L'exemple de la découverte de la planète Neptune permet d'illustrer ces deux points [3, page 96]. Des observations faites, au XIXème siècle des mouvements de la planète Uranus indiquaient que son orbite s'éloignait considérablement de celle prédite par la théorie de la gravitation de Newton. Une réaction possible aurait consisté à protéger la théorie de la gravitation en précisant que la théorie s'appliquait à toutes les planètes, *sauf Uranus*. Cette modification aurait été *ad hoc*. Elle aurait été rejetée par les falsificationistes sophistiqués. En fait, les choses se sont passées différemment : Leverrier en France et Adams au Royaume-Uni ont suggeré l'existence d'une nouvelle planète dans le voisinage d'Uranus. Cette modification-là n'est pas *ad hoc* puisqu'il est possible de tester l'existence de cette planète sans, en même temps, tester la théorie de la gravitation. Le test a eu lieu et Johann Galle a découvert Neptune, en 1846.

Cette confirmation spectaculaire de la théorie de la gravitation universelle ne présente pas d'intérêt pour un falsificationiste naïf. Elle est importante pour un falsificationiste sophistiqué parce que la conjecture de Leverrier et Adams était audacieuse.

Rôle de l'histoire des sciences. La version sophistiquée du falsificationisme présentée ici, tient compte de l'histoire des sciences : une théorie est prudente ou audacieuse en fonction du savoir de l'époque où elle est formulée. La théorie de l'existence de Neptune était audacieuse à l'époque de Leverrier, elle ne l'est plus aujourd'hui [3, chapitre 2, § 4].

1.2.3 Critiques

La principale critique qu'on peut faire au falsificationisme est la même que celle faite à l'inductivisme : historiquement, des grands progrès scientifiques se sont appuyés sur des raisonnements complètement incompatibles avec le falsificationisme.

On en a déjà mentionné deux exemples. Du temps de Copernic, on avait observé (et on s'était trompé) que la taille apparente de Vénus ne change pas au cours de l'année. Les partisans de la théorie de Copernic ont continué à accepter cette théorie malgré cette falsification. De même, on a continué à accepter la théorie de Maxwell malgré la falsification apparente de Hertz. D'autres exemples[5] sont donnés dans [3, chapitre 2, § 4]. Ce type de situation semble être la norme [7, chapitre 5].

Critique de la fiabilité des observations. Rappelons un principe clef du falsificationisme : le rejet d'une théorie, en raison d'une falsification, est un acte définitif, à la différence de l'acceptation, qui est toujours incertaine.

Les critiques faites au sujet de l'inductivisme s'appliquent alors aussi au falsificationisme : les expériences et les observations présupposent une théorie, et sont donc faillibles. Aucune falsification ne peut donc être concluante.

5. En septembre 2011, une expérience aurait montré que des neutrinos peuvent se déplacer à une vitesse supérieure à celle de la lumière [5]. Les physiciens continuent à accepter la théorie de la relativité malgré cette falsification apparente.

1.3 Les paradigmes de Kuhn

La théorie de Thomas Kuhn est exposée dans [10]. Son exemple majeur est exposé dans [9]. La motivation première de Kuhn a été de proposer une théorie de la progression des sciences qui soit compatible avec les études historiques [10, Préface, page 7]. La théorie de Kuhn s'oppose à l'inductivisme et au falsificationisme sur plusieurs points essentiels : la progression des connaissances n'est pas considérée comme continue ; les changements de théories sont davantages provoqués par des phénomènes psychologiques ou sociologiques que par des arguments de logique.

La théorie de Kuhn a eu un retentissement considérable. Des révolutions scientifiques au sens de Kuhn ont été observées dans différentes disciplines [8], et même très récemment [6]. Elle est liée à certains mécanismes d'apprentissage à l'échelle de l'individu (théorie de Piaget). Elle donne une intéressante définition de l'activité d'un chercheur en sciences. C'est une théorie qui « fait réfléchir ».

Toutefois, elle a été pensée pour les sciences de la nature (physique, chimie, biologie, avec un regard insistant sur la physique). Son rapport avec les mathématiques et l'informatique n'est pas immédiat.

Elle est moins ambitieuse, en un sens, que le falsificationisme : elle ne donne aucun critère indiquant si une théorie est meilleure qu'une autre. Mais peut-être ces objectifs sont-ils inaccessibles. Voir [7] pour une charge qui a fait polémique sur ce sujet ... et au delà.

1.3.1 La théorie

Selon Kuhn, une discipline scientifique progresse en accord avec le processus suivant (les trois dernières étapes se répétant indéfiniment) :

1. pré-science,
2. émergence d'un premier paradigme,
3. science normale (avec apparition intermittente d'anomalies),
4. crise,
5. révolution scientifique et changement de paradigme.

En physique par exemple, la théorie d'Aristote a constitué un paradigme. Elle a été remplacée par la physique de Newton (suite à la révolution copernicienne), qui a constitué un nouveau paradigme. Cette dernière a cédé la place à la théorie de la relativité.

1.3.2 Notion de paradigme

Il est dans la nature des paradigmes de résister à une définition précise [3, chapitre 8, page 152].

Un paradigme, dans une discipline donnée, est un ensemble de de lois, de méthodes, voire de principes métaphysiques [10, chapitre III, pages 66-70]. qui constituent, ce qu'on pourrait appeler « le noyau dur » de la discipline : si une expérience contredit le paradigme, on cherche l'erreur ailleurs que dans le paradigme.

Exemple. Dans le cas des problèmes qui ont conduit à la découverte de Neptune, les scientifiques ont cherché l'erreur ailleurs que dans la loi de la gravitation universelle : ils n'ont pas remis en cause un élément du paradigme que constituaient les lois de Newton.

Le paradigme d'une discipline a deux rôles importants pour les membres de la discipline :

1. il joue le rôle de programme de recherche [6] structurant,
2. il façonne la vision du monde des chercheurs.

Rôle de programme de recherche. Les lois qui constituent un paradigme sont générales. Ce sont des promesses de succès. C'est le rôle de la science normale que de réaliser ce programme. Kuhn donne une classification des types de problèmes qu'on rencontre habituellement dans les programmes de recherche issus des paradigmes [10, chapitre II].

Exemples. Mesurer les masses des planètes a fait partie du programme de recherche newtonien. Les lois du mouvement de Newton contredisent les lois de Kepler, dès que le système planétaire comporte plus d'un Soleil et d'une planète. Résoudre le problème des n corps est une question qui est apparue dans le cadre du programme de recherche du paradigme newtonien et qui a conduit à d'importants développements en mathématiques.

Façonnage de la vision du monde. Exemples. Dans le paradigme aristotélicien, l'univers est fini et l'espace vide est inconcevable. Dans le paradigme newtonien, l'univers est infini et la masse des corps est conservée. Dans le paradigme einsteinien, l'univers est (peut-être ?) fini mais en expansion et la masse peut se convertir en énergie.

Les tenants de paradigmes différents ne font pas les mêmes observations, ou donnent des interprétations de ces observations radicalement différentes. Exemple. Après la révolution copernicienne, les astronomes virent, pour la première fois, dans les novae et dans les comètes, des phénomènes qui se produisent au-delà de l'orbite de la lune. Ils utilisèrent pour cela des arguments et des outils qui existaient depuis deux mille ans mais le paradigme aristotélicien interdisait une telle interprétation [9, chapitre 6, *Tycho Brahé*, page 247].

Pour réussir à « voir » de l'oxygène, là où les autres chimistes ne voyaient que de l'air « déphlogistiqué », Lavoisier a dû changer sa manière de voir nombre d'autres substances plus familières [10, chapitre IX, page 166].

Problèmes qui changent de statut. Des problèmes considérés comme scientifiques dans le cadre d'un ancien paradigme deviennent subitement vides de sens dans le nouveau. Exemple. Mesurer la masse des planètes est un problème vide de sens dans le paradigme aristotélicien, où les planètes, faites d'ether, sont impondérables.

L'exemple suivant ne vient pas de Kuhn. Jusqu'au XVIIIème siècle, les axiomes de la géométrie euclidienne sont considérés comme « vrais ». Vers la fin du XIXème, ce ne sont plus que des axiomes possibles parmi d'autres. Bourbaki écrit [2, pages 26-27] : Gauss et

6. Après Kuhn, Imre Lakatos a défini le concept de « programme de recherche » [11] en un sens très proche de celui des paradigmes de Kuhn [3, chapitre 7, et chapitre 8, page 150]. J'avoue ne pas avoir étudié le texte de Lakatos.

Lobatchevsky croient que le débat entre les différentes géométries peut être tranché par l'expérience [...] mais c'est là un problème qui n'a plus rien à voir avec la mathématique.

1.3.3 Science normale, crise et révolution.

Science normale. On a déjà présenté l'activité de science normale un peu plus haut.

Au fil de cette activité, des anomalies apparaissent. Quand elles sont trop nombreuses pour être ignorées (mais ce n'est pas une raison suffisante), une crise apparaît, qui permet à de nouvelles théories de concurrencer le paradigme puis, à l'une d'elles, de le remplacer en tant que paradigme. Une nouvelle activité de science normale peut commencer.

Crise. Qu'est-ce qui différencie la science normale de la science en état de crise [10, chapitre VII, page 117] ? Certainement pas le fait que la première ne rencontre pas de contre-exemples. S'il n'y avait pas de contre-exemples, il n'y aurait pas d'énigmes à résoudre et donc pas d'activité scientifique. La différence tient à la perception de la situation. En période de science normale, le contre-exemple est vu comme une nouvelle énigme à résoudre. En période de crise, il est perçu comme un signe de l'inadéquation du paradigme.

Révolution et changement de paradigme. Un paradigme n'est abandonné qu'en présence d'un nouveau paradigme, prêt à prendre la relève. Rejeter un paradigme sans lui en substituer immédiatement un autre, c'est rejeter la science elle-même [10, chapitre VII, page 117].

1.3.4 Fonctions des différents stades

Pour Kuhn, la science normale et les révolutions ont toutes deux un rôle nécessaire pour qu'il puisse y avoir progrès. Sa théorie ne se limite donc pas à une pure description du passé. Elle est donc susceptible de nous éclairer pour le futur, ce qui est un point important.

Pour qu'il puisse y avoir progrès, il est nécessaire que la science normale soit dans une large mesure non critique. L'acquisition d'un paradigme permet d'attaquer des problèmes de recherche « ésotériques », impensables autrement [10, chapitre I, page 31]. Penser au dernier accélérateur de particules construit à Genève, qui a demandé un investissement gigantesque aussi bien financier et scientifique ... qu'humain. Il permet de tester des raffinements d'une théorie déjà fort sophistiquée. Aurait-il été construit si cette théorie sophistiquée n'avait pas la force d'un paradigme ?

Toutefois, si tous les scientifiques ne pratiquaient que la science normale, toutes les disciplines scientifiques resteraient confinées à un unique paradigme et ne progresseraient jamais au-delà. Qu'est-ce qu'une révolution en politique ? C'est un changement des institutions par des moyens que les institutions interdisent [10, chapitre VIII, page 134]. Seul un phénomène comparable (une révolution scientifique) peut conduire à un changement de paradigme.

1.3.5 Irrationnalité de la progression des connaissances

Kuhn insiste sur le fait que la science normale, qui constitue presque l'essentiel de l'activité scientifique, ne consiste jamais à vérifier le paradigme [10, chapitre XI, page 200]. Il n'est pas question de falsifier le paradigme. La théorie de Kuhn s'oppose au falsificationisme.

Kuhn insiste beaucoup sur le fait que le paradigme façonne la vision du monde des chercheurs. C'est une idée importante, parce qu'elle attaque l'idée qu'il soit possible, pour une communauté scientifique, de décider par des arguments rationnels, qu'un nouveau paradigme est meilleur que l'ancien. C'est une deuxième opposition au falsificationisme. Pour un exemple assez proche, voir [9, chapitre 5, *L'harmonie du système copernicien*, page 215].

La différence entre science normale et science en crise tient à la perception qu'on a des anomalies, pas aux anomalies elles-mêmes. Pour un exemple, lire [9, chapitre 5, *Les raisons pour une innovation — La préface de Copernic*, page 162].

1.3.6 Science et recherche scientifique

Activité pré-scientifique. Qu'est-ce qui différencie l'activité « pré-scientifique » ou « non scientifique » de l'activité de « science normale » ? L'existence d'un paradigme, bien sûr. En l'absence de paradigme, les chercheurs éprouvent le besoin de reconstruire toute la théorie de zéro pour justifier chaque avancée, il y a quasiment autant de théories que de chercheurs.

Condition pour qu'une discipline soit une science. Une discipline est une science si elle comporte au moins un paradigme [10, chapitre I, page 44],

Les scientifiques écrivent pour leurs collègues plutôt que pour le public cultivé. Kuhn écrit : il est difficile de trouver un autre critère, applicable sans l'avantage de la rétrospection, qui proclame qu'un domaine de recherche est devenu une science [10, page 44].

Progression non continue des sciences. La non continuité est due à la nature même des révolutions. Elle est due au fait qu'une partie essentielle du changement est conceptuelle : une vision du monde remplace l'autre [10, chapitre VIII, pages 141-156]. Quiconque observe sérieusement la réalité historique en arrive obligatoirement à penser que la science n'approche pas l'idéal suggéré par un processus cumulatif [10, chapitre VIII, page 139].

Une objection est parfois faite, au sujet de la physique de Newton et de celle d'Einstein. Ne peut-on pas considérer que la physique de Newton est une sorte de cas limite de la physique d'Einstein, et qu'il y a donc quand même un processus cumulatif ? La réponse de Kuhn à cette question (c'est : non) est développée sur plusieurs pages [10, pages 141-147]. Je dois dire qu'elle n'est pas totalement claire, pour moi. Il suggère [10, page 141] qu'on peut, d'un point de vue logique, voir une inclusion entre deux paradigmes consécutifs mais que c'est très improbable historiquement. Il écrit aussi que, même si on peut voir la théorie de Newton comme celle d'Einstein avec la contrainte « vitesse très inférieure à celle de la lumière », les mots « masse », « temps », « position » n'ont en fait plus le même sens. La masse newtonienne est conservée, pas celle d'Einstein [10, page 146].

Qu'est-ce qu'un chercheur ? Un chercheur, en période de science normale, est quelqu'un qui résout des énigmes, mais des énigmes d'un type particulier : celles qui lui permettent de prouver son habileté en utilisant les méthodes admises par le paradigme et qui ont nécessairement une solution, le paradigme étant admis [10, chapitre III].

Sur les sujets de recherche. Les scientifiques choisissent des sujets de recherche du type énoncé dans le paragraphe précédent. Ils sont beaucoup moins concernés que les non scientifiques (poètes, médecins) par l'approbation des non spécialistes. Kuhn remarque par exemple, que les sociologues (qu'il considère comme non scientifiques) ont souvent tendance à justifier leur sujet de recherche par l'importance sociale de la solution recherchée [10, chapitre XII, page 225]. Voir aussi une remarque [10, page 139].

Sur la formation. La formation des étudiants en science s'appuie exclusivement sur des manuels récents. On ne demande jamais aux étudiants de lire les articles de recherche dont ces manuels dérivent. Encore moins de lire les textes importants historiquement (de l'Antiquité, par exemple). Cette technique pédagogique n'est possible que par la confiance des scientifiques en leurs paradigmes.

1.4 Deux mises en perspective de la théorie de Kuhn

1.4.1 La révolution copernicienne

Résumé. Avant Copernic, l'univers est décrit par une combinaison de la théorie physique d'Aristote (350 avant J.C), et de l'astronomie de Ptolémée (150 après J.C). La Terre est une sphère immobile au centre du monde. La frontière extérieure du monde est une autre sphère, (la sphère des étoiles), qui est en mouvement. L'espace entre ces deux sphères est partagé en plusieurs sphères concentriques (un peu comme un oignon), qui ont une réalité physique, qui sont en mouvement, mouvement qui leur est communiqué, de proche en proche, par la sphère des étoiles. Ces sphères d'*ether* contiennent la Lune, Mercure, Vénus, le Soleil, Mars, Jupiter et Saturne [7]. Les corps célestes aussi sont faits de cet *ether* pur, inaltérable et impondérable. La rotation des sphères d'*ether* explique le mouvement *moyen* des corps célestes dans le plan de l'écliptique, mais pas leurs irrégularités. En voici quelques unes [9, chapitre 2] :

1. le Soleil met une année à effectuer sa rotation sur l'écliptique mais sa vitesse de parcours n'est pas constante : il met presque six jours de plus pour passer de l'équinoxe de printemps à l'équinoxe d'automne que pour passer de l'équinoxe d'automne à l'équinoxe de printemps ;
2. d'une façon générale, les planètes se déplacent vers l'est parmi les constellations mais, à l'exception du Soleil et de la Lune, elles se déplacent vers l'ouest pendant de courts

7. Le nombre de sphères a varié suivant les auteurs (Aristote en compte 55), l'ordre des planètes aussi. Ptolémée ne croyait pas en la réalité physique des sphères. Les théories ont évolué dans les 500 ans qui ont séparé Aristote de Ptolémée.

intervalles de temps. Pendant ce mouvement *rétrograde*, les planètes Mars [1], Jupiter et Saturne apparaissent plus brillantes ;

3. les planètes Mercure et Vénus ne s'éloignent jamais beaucoup en apparence du Soleil, à la différence des autres planètes.

L'approche suivie par la plupart des astronomes de l'Antiquité, un peu après Aristote et perfectionnée jusqu'à Ptolémée, consiste, dans sa version la plus simple, à décrire les trajectoires des corps célestes en combinant deux trajectoires circulaires : la planète est en rotation uniforme sur un petit cercle, l'épicycle, dont le centre est en rotation uniforme sur un grand cercle, centré sur la Terre : le déférent. En réglant convenablement les vitesses des épicycles des différentes planètes, vis-à-vis de leur déférent, on parvient à expliquer qualitativement les irrégularités 1 et 2 mentionnées ci-dessus. L'irrégularité 3 se résout en imposant une contrainte *ad hoc* : le centre de l'épicycle de Mercure et de Vénus doit toujours appartenir à la droite qui passe par la Terre et le Soleil [9, chapitre 5, *L'harmonie du système copernicien*, page 204]. En fait, le système composé d'un seul épicycle et d'un seul déférent par planète a dû être compliqué pour tenir compte d'autres irrégularités (intervalles de temps variables entre deux rétrogradations, planètes pas toujours exactement dans le plan de l'écliptique ...). Plusieurs modèles ont été imaginés : des épicycles ont été ajoutés sur d'autres épicycles, des déférents ont été centrés ailleurs que sur la Terre (cercles *excentriques*), et Ptolémée lui-même introduisit le mécanisme des *équants* : on centre le déférent sur la Terre, on impose que la vitesse angulaire de la planète soit constante, mais attention : les angles ne sont pas calculés par rapport au centre du déférent mais par rapport à un point excentré, appelé point équant. La planète parcourt alors le déférent avec une vitesse variable. Dans sa forme la plus élaborée, le système des combinaisons de cercles était une réussite étonnante. Mais il n'a jamais bien fonctionné [9, chapitre 2, *L'anatomie de la croyance scientifique*, page 83].

Au cours du Moyen-Âge, les théories d'Aristote et Ptolémée ont été combattues, oubliées, redécouvertes et testées par les scolastiques. Or cette étude approfondie, sur plusieurs siècles, n'est pas parvenue à résoudre les imperfections du système ptoléméen. Kuhn écrit [9, chapitre 5, *Les raisons pour une innovation — La préface de Copernic*, page 163] : après treize siècles de recherches infructueuses, un astronome attentif pouvait se demander ce que Ptolémée n'aurait pu faire, si d'autres tentatives dans le cadre de la tradition pouvaient réussir. En outre, les siècles qui séparaient Ptolémée de Copernic avaient grossi les erreurs.

Copernic (1473-1543) se convainc qu'une approche radicalement nouvelle est nécessaire. Il met en place un nouveau système et le présente, en 1543, dans un livre intitulé *De Revolutionibus Orbium Caelestium*. Voici les principaux aspects du système copernicien : la sphère des étoiles est conservée mais elle n'est pas en mouvement, le Soleil est au centre, la Terre et les autres planètes décrivent des orbites circulaires autour de lui, la Lune décrit une orbite circulaire autour de la Terre. Le système copernicien résout élégamment, mais seulement *qualitativement*, plusieurs difficultés : les rétrogradations des planètes, le fait qu'elles apparaissent plus brillantes lors des rétrogradations et le fait que leur vitesse de rotation sur l'écliptique peut varier d'une rotation à l'autre [9, chapitre 5, *L'astronomie copernicienne — les planètes*, page 198]. Toutes ces observations peuvent, en principe, se résoudre sans utiliser

d'épicycles.

Malheureusement, le système de Copernic ne fonctionne pas *quantitativement* et Copernic y introduisit des épicycles [8] et des cercles excentriques. Par contre, il évita les équants [9].

Au total, le système de Copernic est à peine plus simple que celui de Ptolémée (les deux utilisent plus de trente cycles) et d'une précision comparable. Outre les avantages qualitatifs décrits ci-dessus, il explique pourquoi Mercure et Vénus ne s'éloignent jamais en apparence du Soleil et fixe l'ordre des planètes. Ces avantages sont contrebalancés par plusieurs inconvénients : le diamètre de la sphère des étoiles est beaucoup plus grand (pour expliquer l'absence apparente de parallaxe), ce qui rend la dernière sphère d'*ether* anormalement grande, Vénus devrait présenter des phases, comme la Lune, ce qui est contredit par les observations de l'époque, et surtout ... il faut que la Terre soit en mouvement.

Kuhn écrit [9, chapitre 5, *L'harmonie du système copernicien*, page 215] : les arguments de Copernic [en faveur de son système] ne sont pas pragmatiques. Ils ne font pas appel au sens utilitaire de celui qui pratique l'astronomie, mais à son sens esthétique et à lui seul. Les harmonies nouvelles pouvaient donc attirer ce petit groupe limité et peut-être irrationnel qui s'occupait d'astronomie mathématique et dont l'intérêt néo-platonicien pour les harmonies mathématiques ne pouvait pas être gêné par des pages et des pages de mathématiques complexes qui finalement aboutissaient à des prévisions numériques à peine meilleures que celles qu'ils avaient connues jusqu'alors.

Copernic était considéré comme un des plus grands astronomes européens de son temps. Le *De Revolutionibus* était le premier exposé d'un astronome européen qui pût rivaliser avec l'*Almageste* de Ptolémée. Le livre de Copernic fut utilisé par les astronomes malgré son étrange hypothèse cosmologique plutôt qu'à cause d'elle [9, chapitre 6, *L'accueil fait à l'œuvre de Copernic*, page 222]. En 1551, à partir de ce livre, furent calculées les premières tables de positions astronomiques depuis trois siècles. Ces tables étaient un peu meilleures que les anciennes, pas dans tous les domaines, mais elles furent utilisées.

Si Copernic fut le plus grand astronome européen de la première moitié du XVIème siècle, Tycho Brahé (1546-1601) fut la figure dominante de la seconde. Il s'opposa toute sa vie à la théorie de Copernic. Il fut le plus grand observateur astronomique à l'œil nu. Il fut le premier à faire des observations régulières des planètes dans leur course, au lieu de ne les observer que dans des conditions particulièrement favorables [9, chapitre 6, *Tycho Brahé*, page 238]. Tycho Brahé apporta trois importantes contributions à la révolution copernicienne :

1. ses mesures, exceptionnellement précises, qui ont fourni des données expérimentales fiables à ses successeurs alors que les données plus anciennes étaient truffées d'erreurs,
2. un système astronomique centré sur la Terre [10] (redevenue immobile), mathématiquement équivalent à celui de Copernic, qui eut pour effet de rallier à lui les astronomes

8. Copernic n'introduisit que des épicycles *mineurs*, c'est-à-dire ceux qui servent à faire varier les vitesses de rotation sur l'écliptique [9, chapitre 2, *L'astronomie ptoléméenne*, page 77]. Les épicycles *majeurs*, qui servent à créer des mouvements rétrogrades, sont évités.

9. L'absence d'équants était considérée par Copernic comme un des plus grands avantages de sa nouvelle théorie [9, chapitre 2, *L'astronomie ptoléméenne*, page 80].

10. Le Soleil décrit une orbite circulaire autour de la Terre. Les planètes décrivent des orbites circulaires autour du Soleil. Les orbites de Mars et du Soleil se coupent.

non-copernicien et d'accélérer ainsi la chute du système de Ptolémée,

3. des observations de la nova de 1572 et de plusieurs comètes, qu'il prouva correspondre à des phénomènes supra-lunaires, et qui forcèrent les astronomes à abandonner la théorie aristotélicienne des sphères d'*ether*.

Jean Kepler (1571-1630) fut un collègue de Tycho Brahé. Il fut copernicien toute sa vie. Il commença par un travail de nettoyage de la théorie de Copernic, qu'il accusait d'avoir gardé un point de vue trop ptoléméen, et à en tirer toutes les conséquences [11]. Il réduisit ainsi le nombre de cercles du système copernicien [9, chapitre 6, *Jean Kepler*, page 249]. Ensuite, en étudiant l'orbite de Mars à partir des mesures de Tycho Brahé, il aboutit à ce qu'on appelle les « lois de Kepler », les deux premières étant publiées en 1609, la troisième en 1619 :

1. les planètes se déplacent sur des ellipses dont le Soleil occupe un des foyers,
2. la vitesse orbitale de chaque planète varie de telle sorte qu'une droite, qui relie la planète au Soleil, balaie dans l'ellipse, des aires égales en des temps égaux,
3. si T_1 et T_2 sont les temps nécessaires à deux planètes pour accomplir une révolution sur leur orbite et si R_1 et R_2 sont les distances moyennes des planètes au Soleil, alors $(T_1/T_2)^2 = (R_1/R_2)^3$.

Kuhn fait remarquer que la première loi de Kepler ne se déduit pas uniquement de l'observation et du calcul [9, chapitre 6, *Jean Kepler*, page 255]. À moins de supposer que les trajectoires des planètes se referment sur elles-mêmes (ce qui n'était pas considéré comme acquis, de son temps), il faut disposer d'une loi qui donne la vitesse. Kepler fit un usage constant de son intuition.

Kepler avait résolu le problème des planètes et sa version de la théorie de Copernic aurait fini par convertir tous les astronomes, surtout après 1627, où il publia de nouvelles tables de positions, supérieures à toutes les autres. Mais l'histoire ne s'arrête pas là. En 1609, Galilée eut l'idée de pointer une lunette astronomique vers le ciel et y découvrit un grand nombre de confirmations de la théorie de Copernic. Les phases de Vénus purent être observées. Ces observations apportèrent aussi des réponses à certaines énigmes concernant la taille des étoiles. La lunette de Galilée n'apportait pas une preuve, mais c'était un instrument de propagande [9, chapitre 6, *Galilée*, page 264].

Les lois du mouvement de Newton, combinées à la loi de la gravitation universelle (1687) permirent de démontrer les lois de Kepler, sous plusieurs hypothèses simplificatrices. Certains phénomènes, comme la « précession du périhélie de Mercure » s'expliquent mal avec les lois de Newton et la théorie de la relativité donne des prédictions nettement plus précises. L'histoire ne semble donc toujours pas s'être arrêtée !

Analyse à la lumière de la théorie de Kuhn. Le résumé précédent est extrait d'un livre de Kuhn, publié avant la théorie des révolutions scientifiques. Kuhn n'y fait donc pas explicitement référence à sa théorie mais on la sent sourdre (ou plutôt poindre) par endroits.

11. Copernic faisait, comme Ptolémée, jouer un rôle spécial à la Terre. En particulier, il supposait que les plans des orbites planétaires se coupaient au centre de la Terre. Kepler les fit se couper au centre du Soleil.

Voici une analyse du texte, sous l'angle de la théorie des révolutions scientifiques. Il ne s'agit en aucun cas de « prouver » la théorie de Kuhn mais d'illustrer la théorie et de préciser la pensée de Kuhn, sur cet exemple.

La révolution copernicienne est une révolution scientifique dans un domaine scientifique qui touche à la fois l'astronomie, la cosmologie et même la physique. Le paradigme copernicien succède au paradigme aristotélicien.

Le paradigme aristotélicien n'est pas constitué de l'intégralité de la théorie d'Aristote. Un très grand nombre de changements ont été apportés par les derniers aristotéliciens. Par contre, la position centrale de la Terre en fait certainement partie. Kuhn écrit [9, chapitre 3, *L'univers aristotélicien*, page 96] : on peut construire un univers aristotélicien avec trois ou cinq éléments terrestres aussi bien qu'avec quatre, et avec des épicycles aussi bien qu'avec des cercles homocentriques ; mais l'univers d'Aristote ne peut pas résister, et n'a pas résisté, aux modifications qui faisaient de la Terre une planète.

Les scolastiques du Moyen-âge effectuent un travail de science normale dans le paradigme aristotélicien. Ils ne remettent pas en cause le paradigme mais le précisent. Ce faisant, ils mettent en évidence des anomalies. Ce travail de science normale des scolastiques était indispensable pour que la révolution eût lieu. Kuhn écrit [9, chapitre 4, *La critique scolastique d'Aristote*, page 142] : au cours du XVIIème siècle, au moment où toute son utilité était démontrée pour la première fois [12], la science scolastique fut extrêmement critiquée. Les savants du Moyen-Âge trouvèrent plus souvent leurs problèmes dans les textes que dans la nature. Un grand nombre de ces problèmes, aujourd'hui, ne semblent pas être des problèmes du tout. Mais de quelle autre façon la science aurait-elle pu renaître en Occident ? Les siècles durant lesquels a régné la scolastique sont ceux pendant lesquels la science antique fut reconstituée, assimilée, testée. Au fur et à mesure que les points faibles étaient découverts, ils devenaient les foyers des premières recherches efficaces du monde moderne. Les grandes théories scientifiques nouvelles des XVIème et XVIIème siècle trouvent toutes leur origine dans les accrocs faits par la critique scolastique à la pensée d'Aristote.

Le paradigme copernicien comporte l'idée que le Soleil occupe une position centrale mais il comporte aussi (peut-être) des éléments plus subtils, comme l'idée que le problème des planètes doit avoir une solution à la fois simple et précise. Voir [9, chapitre 5, *Les raisons pour une innovation — La préface de Copernic*, page 164].

Les paradigmes façonnent la vision du monde des chercheurs qui y adhèrent. Et cette vision du monde affecte l'interprétation des observations. Au sujet de la découverte du caractère supra-lunaire des novae et des comètes, Kuhn écrit [9, chapitre 6, *Tycho Brahé*, page 247] : il suffit d'un bout de ficelle au copernicien Maestlin pour affirmer que la nova de 1572 était au delà de la Lune. Les phénomènes et les instruments nécessaires existaient deux millénaires avant la naissance de Tycho Brahé, mais les observations ne furent pas faites, ou bien elles ne furent pas correctement interprétées. Durant la seconde moitié du XVIème siècle, des phénomènes connus depuis très longtemps changèrent rapidement de sens et de signification.

12. Comprendre : pour avoir permis la révolution copernicienne.

Enfin, Kuhn insiste à de nombreuses reprises sur le fait que l'adhésion au nouveau paradigme s'est faite pour des raisons qui n'ont rien de rationnel.

1.4.2 La théorie de Piaget

Jean Piaget a passé plus d'un demi-siècle à étudier l'intelligence de l'enfant. Pourtant, ses travaux dans ce domaine ne devaient primitivement constituer qu'un moyen ou un détour pour aborder un problème relevant de la philosophie des sciences : comment se constituent les connaissances dans l'esprit ? Par quels processus, selon quels rythmes, l'esprit passe d'un moindre niveau à un niveau supérieur de connaissance. Cette question révèle la complémentarité de l'épistémologie et de la psychologie de l'intelligence [4, chapitre 3, C, b) Piaget].

Le développement mental de l'enfant, de la naissance à l'âge adulte, est vu comme une marche vers un équilibre psychique toujours amélioré, qui se fait par stades. Piaget en donne six [12, chapitre 1, page 14]. À chaque nouveau stade, l'enfant est capable d'effectuer de nouvelles opérations (en un sens large du mot), qui lui étaient inaccessibles au stade précédent. En voici un exemple [12, iv. L'adolescence, page 90]. Jusque vers douze ans environ, les opérations de l'intelligence enfantine sont uniquement « concrètes », c'est-à-dire ne portent que sur la réalité elle-même et, en particulier, sur les objets tangibles susceptibles d'être manipulés. Si on demande aux sujets de raisonner sur de simples hypothèses, sur un énoncé purement verbal des problèmes, ils perdent aussitôt pied et retombent dans l'intuition prélogique des petits. Par exemple, tous les enfants de neuf ou dix ans savent sérier les couleurs mais ils échouent complètement [13] à résoudre une question comme celle-ci, même posée par écrit : « Édith a les cheveux plus foncés que Lili. Édith a les cheveux plus clairs que Suzanne. Laquelle des trois a les cheveux les plus foncés ? ». En effet, dans cette question, on pose trois personnages fictifs qui ne sont pour la pensée que de simples hypothèses. Or après onze ou douze ans, la pensée formelle devient possible et l'enfant devient capable de déduire les conclusions à tirer de pures hypothèses et non plus seulement d'une observation réelle.

Kuhn écrit [10, Préface, page 9] : « une note explicative me fit connaître les expériences de Jean Piaget explorant les différents univers de l'enfant qui grandit et le processus de transition qui permet de passer de l'un à l'autre ». L'expérience citée ci-dessus suggère que les changements de stades chez Piaget induisent des changements de vision du monde chez l'enfant. Mais il faut se méfier de l'analogie. J'ai cherché dans [12] si Piaget présentait la progression des connaissances comme un phénomène non continu, ou plutôt, non cumulatif. Je n'ai pas trouvé une telle affirmation.

13. L'âge de neuf ou dix ans semble être indicatif.

Bibliographie

[1] Gabrielle Bonnet and Philippe Saadé. Le mouvement rétrograde de Mars. `http://culturesciencesphysique.ens-lyon.fr/XML/db/csphysique/metadata/LOM_CSP_Mouvement_Mars.xml`, 2003.

[2] Nicolas Bourbaki. *Éléments d'Histoire des Mathématiques*, volume iv of *Collection Histoire de la Pensée*. Hermann, second edition, 1969.

[3] Alan F. Chalmers. *Qu'est-ce que la science ?* Le Livre de Poche, Paris, 1987. Titre original : What is this Thing Called Science ? An Assessment of the Nature and the Status of Science and its Methods (1976).

[4] Alain Danset. *Éléments de psychologie du développement. Introduction et aspects cognitifs*. Armand Colin / Bourrelier, 1983.

[5] T. Adam et al. Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. `http://arxiv.org/abs/1109.4897`, 2011.

[6] Benoit Famaey. Gravitation Modifiée et Matière Noire. `http://www.astro.ulb.ac.be/Publications/bf_Matsombre.pdf`.

[7] Paul Feyerabend. *Contre la méthode. Esquisse d'une théorie anarchiste de la connaissance*, volume 56 of *Points Sciences*. Éditions du Seuil, Paris, 1979. Titre original : Against Method (1975).

[8] Anthony Hallam. *Une révolution dans les sciences de la Terre (de la dérive des continents à la tectonique des plaques)*, volume 5 of *Points Sciences*. Éditions du Seuil, Paris, 1976. Titre original : A Revolution in the Earth Sciences : From Continental Drift to Plate Tectonics (1973).

[9] Thomas Samuel Kuhn. *La révolution copernicienne*, volume 3 of *Le phénomène scientifique*. Librairie Arthème Fayard, Paris, 1973. Titre original : The Copernician Revolution (1957).

[10] Thomas Samuel Kuhn. *La structure des révolutions scientifiques*, volume 791 of *Champs Sciences*. Éditions Flammarion, Paris, 1983. Titre original : The Structure of Scientific Revolutions (1962).

[11] Imre Lakatos. *Falsification and the Methodology of Scientific Research Programmes*, pages 91–196. Cambridge University Press, Cambridge, 1974.

[12] Jean Piaget. *Six études de psychologie*. folio essais. Denoël, 1964.

[13] Karl Raimund Popper. *La logique de la découverte scientifique*. Payot, Paris, 1973. Titre original : The Logic of Scientific Discovery (1968).

Chapitre 2

Crise en mathématiques et naissance de l'informatique

Le but de ce deuxième cours : analyser les progrès effectués en mathématiques et en informatique, sous l'angle de la théorie de Kuhn. Cette analyse n'est pas évidente.

Kuhn suggère que les mathématiques sont une science, en son sens, puisqu'il écrit [11, chapitre I, page 35] qu'elles comportent des paradigmes, qui datent de la préhistoire. Mais il ne donne aucun exemple et, en dehors, de deux autres passages, pages 42 et 214, il ne les mentionne plus, autant que j'aie pu m'en rendre compte. Je ne connais aucun texte liant la théorie de Kuhn et les mathématiques. Les ouvrages qui traitent de l'histoire des sciences couvrent, au moins mentionnent, souvent, plusieurs sciences (physique, chimie, biologie, géologie ...) mais les mathématiques sont toujours traitées dans des ouvrages séparés. Pourquoi la théorie de Kuhn a-t-elle si peu d'écho chez les mathématiciens ? Peut-être ne s'applique-t-elle tout simplement pas aux mathématiques ; peut-être a-t-elle été boudée par les historiens des mathématiques, en raison de son caractère plus sociologique et psychologique que mathématique. Comme il est difficile de trouver des paradigmes dans ces disciplines, on pourrait être tenté de conclure que les mathématiques et l'informatique ne sont pas des sciences, au sens de Kuhn. Pour couper court à une certaine polémique, je précise que, pour moi, ces questions ne sont intéressantes que par les réflexions qu'elles suscitent. Les réponses n'ont que peu d'intérêt. Kuhn serait probablement d'accord, puisqu'il écrit [11, chapitre XII, page 220] : une *définition* du mot « science » a-t-elle une si grande importance ? Une définition peut-elle assurer à quelqu'un qu'il est ou n'est pas un homme de science ? Et si c'est le cas, pourquoi les spécialistes des sciences de la nature ou les artistes ne se soucient-ils pas de la définition donnée à ce terme ? On en arrive inévitablement à supposer que l'enjeu du problème est plus fondamental. Sans doute les questions que l'on se pose réellement sont-elles plutôt : pourquoi ma spécialité ne réussit-elle pas à progresser comme la physique, par exemple ? Ce ne sont pas là pourtant des questions auxquelles on puisse répondre en se mettant d'accord sur une définition.

Même si l'on ne trouve pas facilement de paradigme en mathématiques, on trouve des « programmes de recherche ». Parmi eux, les « vingt-trois problèmes » posés par le mathématicien David Hilbert (1862-1943) en 1900, puis précisés en 1928. Au moins deux de ces

problèmes sont liés à l'informatique : le deuxième et le dixième. Faute de temps, j'ai choisi de me concentrer sur une partie de l'histoire du dixième problème de Hilbert, où apparaît une figure majeure de l'informatique : Alan Turing. Ce deuxième cours est lui-même coupé en deux parties. On commence par la partie la plus technique du cours (en profitant du fait que les lecteurs sont encore frais). On termine par une mise en perspective historique, censée rapprocher ce deuxième cours du premier.

2.1 Le problème de l'arrêt des algorithmes

On explique le problème de l'arrêt des algorithmes, en suivant l'idée originale d'Alan Turing, mais de façon un peu imprécise. Avantage : on comprend facilement. Inconvénient : les imprécisions laissent la porte ouverte à une réfutation, qu'on pourrait appeler, l'argument de Marvin Minsky. Pour concilier les deux raisonnements, en apparence contradictoires, on sera amené à préciser quelques « détails ».

2.1.1 Le problème n'a pas de solution

On s'intéresse à des programmes écrits dans un langage de programmation fixé. On suppose l'existence d'un programme A (pour « arrêt »), qui prend en entrée le code d'un autre programme P ainsi qu'une donnée x pour P. Le programme A retourne *vrai* si P s'arrête lorsqu'on l'exécute en l'appliquant à la donnée x (c'est-à-dire, lorsqu'on exécute $P(x)$). Il retourne *faux* si l'exécution de $P(x)$ ne s'arrête pas.

Quelques remarques avant de continuer. On n'a pas précisé comment les données de A et de P étaient lues. Cette imprécision-là n'est pas bien grave : on pourrait fixer un mécanisme précis quelconque. Le fait qu'un programme puisse lire et analyser le code d'un autre programme, ne devrait pas surprendre un étudiant en informatique d'aujourd'hui [1]. Il serait assez naturel d'imaginer que les données en question sont codées par des chaînes de caractères mais, pour simplifier la suite du raisonnement, on peut aussi supposer que les données (aussi bien le code de P que x) sont des entiers naturels : une suite de caractères peut en effet très bien être vue comme un (très grand) nombre écrit en base 2.

La suite du raisonnement consiste à montrer que l'hypothèse de l'existence de A conduit à une contradiction. On en conclura que A n'existe pas. À partir du programme A, il est possible d'écrire un autre programme M dont le pseudo-code apparaît figure 2.1, page 24. Notons E le domaine de définition de M, c'est-à-dire l'ensemble des entiers naturels q satisfaisant les conditions énoncées en commentaire dans le pseudo-code de M. Notons H (pour « halt ») le sous-ensemble suivant de E :

$$H \underset{\text{def}}{=} \{q \in E \mid \text{l'exécution de } M(q) \text{ s'arrête}\}\,.$$

Dans le langage que nous nous sommes fixés, le programme M est codé par un entier naturel m. On remarque que $m \in E$.

1. La même idée (évidemment, pas énoncée exactement comme cela), présentée par Alan Turing, en 1936, avant la construction des premiers ordinateurs, a dû paraître nettement moins évidente !

```
function M(q)
  L'entier naturel q code un programme Q.
  Le programme Q est lui-même paramétré par une donnée, qui reste à préciser
begin
  if A(q, q) = vrai then
    while vrai do
    end do
  end if
end
```

FIGURE 2.1 – La fonction M, en pseudo-code.

Question. Est-ce que $m \in H$? Supposons que ce soit le cas. Alors l'exécution de $M(m)$ s'arrête (d'après la définition de H), donc $A(m, m)$ retourne *faux* (d'après le pseudo-code de M), et donc l'exécution $M(m)$ ne s'arrête pas (d'après la spécification de A). Première contradiction. Supposons que $m \notin H$. Alors l'exécution $M(m)$ ne s'arrête pas (d'après la définition de H), donc $A(m, m)$ retourne *vrai* (d'après le pseudo-code de M), et donc l'exécution $M(m)$ s'arrête (d'après la spécification de A). Seconde contradiction. Dans tous les cas, on aboutit à une contradiction. On en conclut que le programme A n'existe pas.

2.1.2 Le problème a une solution

De façon abstraite, un ordinateur qui exécute un programme P, calcule les termes d'une suite de la forme

$$u_{n+1} = f(u_n)\,, \quad u_0 = \text{un entier quelconque}\,. \tag{2.1}$$

Les termes u_n sont des entiers codant l'état de la mémoire de l'ordinateur. La mémoire est finie. Pour fixer les idées, supposons qu'elle comporte 10^9 octets. L'ensemble U des états u_n possibles contient un nombre énorme mais fini d'éléments. Son cardinal vaut 256^{10^9}. Le fait que u_{n+1} ne dépend que de u_n traduit le fait que l'ordinateur est une machine déterministe. Supposons que l'exécution de P ne s'arrête pas. Alors la suite (u_n) est infinie. Or toute suite infinie, de la forme (2.1), dont les termes appartiennent à un ensemble U fini, est nécessairement *ultimement périodique*, ce qui veut dire, qu'à partir d'un certain rang, une même séquence de termes est répétée cycliquement. Et il existe des algorithmes pour détecter de telles suites, comme les algorithmes de Floyd et de Brent (voir section 2.4.1). Par conséquent, en combinant l'un de ces algorithmes avec un autre algorithme qui simule l'exécution de P, on peut déterminer à coup sûr, et au bout d'un temps fini, si l'exécution de P s'arrête ou pas.

2.1.3 La notion d'algorithme

Les raisonnements tenus en sections 2.1.1 et 2.1.2 semblent contradictoires. En fait, ils ne le sont pas, parce qu'un glissement de sens s'est opéré, en raison de certaines imprécisions.

La limitation de la mémoire est la clef de voûte du second raisonnement, qui est en fait, dû à Marvin Minsky [14]. Dans le premier raisonnement, il serait donc plus juste de parler d'algorithmes que de programmes d'ordinateurs écrits dans un langage fixé. En effet, quand on décrit un algorithme comme l'algorithme d'Euclide ou le pivot de Gauss, on ne pose jamais de limite sur la taille des entiers ou sur le nombre de colonnes de la matrice.

Mais alors, pourquoi ne pas avoir formulé le premier raisonnement directement en termes d'algorithmes ? Parce qu'on ne dispose pas, *a priori*, d'un langage précis, dont les instructions ont un sens précis, pour décrire les algorithmes. Or cette hypothèse est essentielle dans le premier raisonnement. L'exemple suivant montre que ce problème d'un langage pour décrire les algorithmes est plus profond qu'on ne le croit.

Un procédé algorithmique ? Voici un procédé pour déterminer un plus court chemin entre deux sommets A et B, dans un graphe. Prendre un anneau métallique pour chaque sommet. Relier les anneaux par des ficelles, à raison d'une ficelle par arête (si les arêtes ont des longueurs, on peut même couper des ficelles de différentes longueurs). Prendre l'anneau A dans une main, l'anneau B dans l'autre et tirer. On obtient un plus court chemin en suivant les ficelles qui sont tendues.

Tout être humain se convainc facilement qu'on a bien là un procédé théorique pour déterminer un plus court chemin entre deux sommets. En supposant qu'on dispose d'un ordinateur muni d'une imprimante $3D$, on pourrait même imaginer de le programmer ! Mais, est-ce que ce procédé est un algorithme ?

Les machines de Turing. En 1936, Alan Turing cherche à donner une définition précise des algorithmes et de la façon de les exécuter. Il invente pour cela, ce qu'on appelle aujourd'hui : les machines de Turing (voir section 2.4.3 pour une description). Elles ont une caractéristique essentielle : elles ont une mémoire infinie à leur disposition.

Ce que démontre Alan Turing en 1936 [17], c'est qu'il n'existe aucune machine de Turing A, prenant en entrée le codage d'une autre machine de Turing P ainsi qu'une donnée x pour P, qui retourne *vrai* si l'exécution de $P(x)$ s'arrête, et *faux* si $P(x)$ ne s'arrête pas.

La thèse de Church. Le texte qui suit est fortement inspiré de [13, section 5.7].

Il y a quelque chose de gênant dans le résultat précédent. La définition des machines de Turing est surchargée de détails qui semblent arbitraires (voir section 2.4.3) : pourquoi la mémoire a-t-elle la forme d'un ruban ? pourquoi le ruban est-il fini d'un côté ? pourquoi une tête de lecture et pas deux ?

Pour répondre à cette question, il est utile d'introduire deux définitions, qui permettent de s'intéresser à ce que les machines de Turing calculent et pas à la façon dont elles le calculent :

— un ensemble E d'entiers naturels est dit *décidable par machine de Turing*, s'il existe une machine de Turing qui prenne entrée un entier x et qui s'arrête soit dans l'état « vrai » si $x \in E$, soit dans l'état « faux » si $x \notin E$;
— un ensemble E d'entiers naturels est dit *semi-décidable par machine de Turing*, s'il existe une machine de Turing qui prenne entrée un entier x et qui s'arrête si, et seulement si, $x \in E$.

Des études ont montré que les notions d'ensembles décidables et semi-décidables par machines de Turing sont insensibles aux variations de détail énoncées plus haut. Elles ont montré aussi que les mêmes ensembles (exactement les mêmes) peuvent être définis par des procédés, en apparence très éloignés des machines de Turing, tels que le lambda-calcul de Church [2]. Tous ces résultats montrent que les ensembles décidables et semi-décidables par machines de Turing ont un caractère beaucoup plus « intrinsèque » que leur définition ne le laisse croire.

Bien des algorithmes avaient été inventés avant l'apparition des machines de Turing (l'algorithme d'Euclide, par exemple). Appelons ces algorithmes « des algorithmes au sens intuitif », non pas qu'on doute que l'algorithme d'Euclide ou le pivot de Gauss soient bien des algorithmes, mais parce qu'on serait bien en peine de donner un critère objectif qui décide si un procédé est un algorithme ou non. À cette notion d'algorithme, au sens intuitif, on peut rattacher les notions d'ensembles d'entiers naturels « décidables au sens intuitif » et « semi-décidables au sens intuitif » [2].

Comment la décidabilité et la semi-décidabilité, au sens des machines de Turing, sont-elles liées à la décidabilité et à la semi-décidabilité, au sens intuitif ? Un sens est évident : tout ensemble décidable (respectivement semi-décidable) par machine de Turing, est décidable (respectivement semi-décidable) au sens intuitif. La réciproque est connue comme la « thèse de Church ».

Thèse de Church : tout ensemble décidable (respectivement semi-décidable) au sens intuitif est décidable (respectivement semi-décidable) par machine de Turing.

Problèmes algorithmiquement indécidables. Un « problème de décision » est une question $P(x)$, dont la réponse, *vrai* ou *faux*, dépend des valeurs des paramètres. Par exemple, la question « x est-il un nombre pair ? » est un problème de décision : la réponse, *vrai* ou *faux*, dépend de la valeur de x.

Un problème de décision $P(x)$ est dit *algorithmiquement indécidable* [3] s'il n'existe aucun algorithme qui réponde à la question, au bout d'un temps fini, pour toutes les valeurs possibles du paramètre x (ou des paramètres, s'il y en a plusieurs).

2. Remarquer qu'entre la décidabilité et la semi-décidabilité intuitives, on trouve les mêmes relations qu'entre la décidabilité et la semi-décidabilité au sens des machines de Turing. Par exemple, un ensemble E est décidable au sens intuitif si, et seulement si, E et son complémentaire sont semi-décidables au sens intuitif.

3. Plutôt que de définir la notion de « problème algorithmiquement indécidable », on aurait pu, comme dans le paragraphe précédent, définir la notion « d'ensemble algorithmiquement indécidable ». Ce sont deux façons différentes de définir la même chose : un ensemble E est algorithimiquement indécidable si le problème de décision « x appartient-il à E » est algorithmiquement indécidable.

Le résultat de Turing et la thèse de Church (qu'on admet), montrent que le problème $A(i, x)$ « la machine de Turing numéro i s'arrête-t-elle lorsqu'on l'applique à la donnée x ? » est algorithmiquement indécidable. C'est l'un des tous premiers problèmes algorithmiquement indécidables connus.

La résolution des équations diophantiennes $P(x_1, \ldots, x_n) = 0$ est un autre, célèbre, problème algorithmiquement indécidable. Une équation diophantienne est une équation polynomiale, à coefficients rationnels, en un nombre quelconque d'inconnues $x_1, \ldots, x_n$. Le problème de décision « déterminer si une équation diophantienne quelconque admet une solution telle que $x_1, \ldots, x_n$ soient des entiers naturels » est algorithmiquement indécidable. La question de la mise au point d'un tel algorithme est le dixième problème de Hilbert, posé en 1900. Le résultat d'indécidabilité a été prouvé par Youri Matiiassevitch, en 1970.

2.2 Mise en perspective

En 1936, Turing et Church avaient une raison précise de s'intéresser à la théorie de la calculabilité. Grâce à l'indécidabilité du problème de l'arrêt, Turing a pu montrer que « l'Entscheidungsproblem », posé par Hilbert, n'a pas de solution. C'est l'histoire de ce problème qu'on essaie de raconter ci-dessous.

2.2.1 La notion de vérité en mathématiques

Sur ce sujet, on suggère la lecture de [1, Fondements des mathématiques ; logique ; théorie des ensembles, pages 9-63]

La notion traditionnelle de vérité mathématique est celle qui remonte à la Renaissance. Dans cette conception, les objets dont traitent les mathématiciens sont très proches de ceux dont traitent les sciences de la nature : « des propositions assez cachées, comme la propriété de l'hypoténuse d'un triangle rectangle furent découvertes par l'expérience » [16]. Ils sont connaissables à la fois par l'intuition et par le raisonnement, qui ne sont faillibles que si on en les emploie pas comme il faut [1, La notion de vérité en mathématique, page 22]. En particulier, les axiomes de la géométrie euclidienne sont considérés comme « vrais ».

Le premier coup porté à cette vision des mathématiques est l'édification d'une géométrie non euclidienne, au début du XIXème siècle : on s'aperçoit que l'axiome des parallèles [4] n'est pas « vrai » puisqu'on peut le supposer faux et construire une géométrie sensée. L'opinion courante que les axiomes de la géométrie (ou de n'importe quelle discipline) peuvent être établis en raison de leur évidence, s'effondre [15, chapitre I, page 23].

En ce début de XXIème siècle, nous sommes habitués à l'idée que les mathématiques peuvent receler ... des bizarreries [5] mais on ne l'était pas à la fin du XIXème siècle : « Com-

4. Par un point, en dehors d'une droite fixée, il passe exactement une parallèle à cette droite.

5. L'existence de fonctions continues non dérivables ; la possible mise en bijection de $\mathbb{R}$ et de $\mathbb{R}^n$ avec $n > 1$, au sujet de laquelle, Georg Cantor a écrit : « je le vois, mais je ne le crois pas » ; des paradoxes issus de définitions trop générales de la notion d'ensemble : notons E l'ensemble des ensembles qui ne se contiennent pas eux-mêmes. Est-ce que $E \in E$? Notons F l'ensemble des entiers dont la définition peut

ment l'intuition peut-elle nous tromper à ce point ? » demande Henri Poincaré en 1905 [1, page 27].

Ces problèmes conduisent un groupe dominant de mathématiciens, dont David Hilbert, à adopter un point de vue beaucoup plus abstrait et formel que leurs prédecesseurs, où les mathématiques sont vues comme l'art de déduire des théorèmes à partir d'axiomes par des raisonnements de pure logique, mais où la question de la véracité des axiomes n'a plus de sens [1, page 29]. De nombreuses théories sont axiomatisées dont l'arithmétique de Peano (1889) [20] et la théorie des ensembles de Zermelo (1908) et Fraenkel (1922) [21].

2.2.2 Consistance, complétude et décision

Il est important de préciser que, dans ces théories axiomatiques, les règles d'inférence permettant de déduire un théorème à partir des axiomes sont complètement explicitées. Dans ces théories, il est possible de vérifier mécaniquement qu'une suite de formules est une démonstration correcte.

On peut se poser trois questions importantes au sujet d'une théorie axiomatique :

(a) est-elle *consistante* ? Est-on sûr qu'il est impossible de démontrer une chose (un théorème) et son contraire (la négation de ce théorème) ?

(b) est-elle *complète* ? Étant donnée une formule, est-on sûr qu'il existe soit une démonstration de la formule, soit une démonstration de la négation de la formule ?

(c) est-elle *algorithmiquement décidable* ? Existe-t-il un algorithme, qui prenne en entrée une formule, et qui détermine si la formule admet une démonstration, dans le cadre de cette théorie ?

Le problème de la consistance est important, non seulement parce que la question de la véracité des axiomes n'a plus de sens, mais aussi en raison de l'accumulation de phénomènes contre-intuitifs et même de paradoxes, dont on a donné quelques exemples plus haut.

Au congrès international des mathématiques de 1900, Hilbert posa vingt-trois problèmes. Le deuxième consiste à « établir la consistance » de l'arithmétique de Peano, problème qu'on n'aurait sans doute pas songé à se poser avant la fin du XIXème [1, La métamathématique, page 56]. Le dixième problème consiste à mettre au point un « algorithme », qui prenne en entrée une équation diophantienne et qui détermine si elle a des solutions ou non. J'ai mis entre guillemets des expressions qui doivent être précisées sous peine de contresens.

Le deuxième problème. Sur ce sujet, on conseille la lecture de [15] et de [4].

Hilbert et son école cherchent une démonstration de la consistance qui n'emploie que des « procédés finis » [1, page 59] qui sont, informellement, des raisonnements suffisamment simples pour être acceptés par tous les mathématiciens. Cette exigence est très importante. Elle est liée au souci de ne pas démontrer la consistance d'une théorie dont on doute, au

s'exprimer en moins de seize mots français. L'ensemble F est fini. Mais il est contradictoire de définir un entier comme « le plus petit entier qui n'est pas définissable par moins de seize mots français », car cette définition ne comporte que quinze mots. Au sujet de ce paradoxe (paradoxe de Berry), voir le récent article [7].

moyen d'une théorie plus compliquée et donc encore plus douteuse. C'est cette exigence qui amène Hilbert et son école à préciser une théorie abstraite et formelle de la démonstration, où les règles d'inférence sont totalement explicitées. De tels mécanismes sont intégrés à l'arithmétique de Peano.

En 1931, Kurt Gödel (1906-1978) montra que, sous ces conditions, le deuxième problème de Hilbert n'a pas de solution. Gödel eut l'idée de coder les formules de l'arithmétique de Peano sous la forme de nombres entiers, de telle sorte qu'il obtenait des entiers qui codaient des énoncés sur les entiers. Il put ainsi démontrer que des propriétés comme le fait « d'être une preuve » ou « d'être démontrable » ne sont ni plus ni moins arithmétiques que celle « d'être un carré » ou « d'être premier » [10, chapitre II, page 87]. Gödel utilisa ce codage pour construire une formule, appelons-la G, exprimant « la formule G n'est pas démontrable dans l'arithmétique de Peano ». Le raisonnement est alors celui-ci : si G est démontrable dans l'arithmétique de Peano, alors l'arithmétique de Peano est inconsistante ; si G n'est pas démontrable dans l'arithmétique de Peano, alors G est vraie et l'arithmétique de Peano est incomplète.

Quelques remarques s'imposent.

La construction de Gödel est « irréfragable » [15, page 152] puisqu'elle est générique : elle peut être appliquée à toutes les théories axiomatiques qui contiennent l'arithmétique de Peano, c'est-à-dire, à toutes les théories, à l'exception des plus limitées.

En 1936, le mathématicien Gentzen a démontré la consistance de l'arithmétique de Peano mais dans le cadre d'une théorie plus expressive que l'arithmétique de Peano (il n'y a donc pas de contradiction avec le théorème d'incomplétude de Gödel). La théorie utilisée par Gentzen n'emploie pas que les « procédés finis » demandés par Hilbert. Voir [19].

Si on admet sa consistance, l'arithmétique de Peano est incomplète et donc la question de la décidabilité se pose. En effet, si l'arithmétique de Peano avait été complète, elle aurait aussi été décidable : étant donné un énoncé à démontrer, il aurait suffi d'énumérer toutes les démonstrations par longueur croissante. Au bout d'un nombre fini d'étapes, on aurait été certain d'obtenir soit la démonstration de l'énoncé, soit la démonstration de sa négation.

La formule G de Gödel est une formule combinatoire, construite sur mesure par Gödel. Mais existe-t-il des formules plus proches de celles qu'on rencontre couramment en mathématiques, qui soient indémontrables ? Des réponses étonnantes se trouvent par exemple dans [5, 6] et plus généralement à l'URL `http://cristal.univ-lille.fr/~jdelahay/pls`. Un exemple concret est présenté dans le récent article [8]. Considérons la fonction M de la figure 2.2. L'énoncé « $M(x)$ s'arrête pour tout $x \in \mathbb{N}$ » peut s'exprimer dans le cadre de l'arithmétique de Peano. Cet énoncé n'est pas démontrable dans le cadre de cette même arithmétique. Pourtant, cet énoncé est vrai : il a été prouvé, mais dans le cadre d'une théorie plus puissante que celle de Peano.

La question de la décidabilité algorithmique. Au printemps 1935, Alan Turing (1912-1954) suivit un cours de mathématiques qui se concluait par l'étude des théorèmes de Gödel et décida de s'attaquer au problème **(c)** de la décidabilité des théories axiomatiques [10, pages 85-87] et en particulier à la décidabilité de l'arithmétique de Peano. Ce problème **(c)**,

```
function M(x)
  Le paramètre x est un réel
begin
  if x < 0 then
    return −x
  else
    return M(x − M(x − 1))/2
  end if
end
```

FIGURE 2.2 – Une fonction qui s'arrête dans tous les cas mais dont l'arrêt ne peut pas être prouvé dans le cadre de l'arithmétique de Peano.

souvent appelé « Entscheidungsproblem » était l'un des sujets de préoccupation de Hilbert [6] et de son école.

Le dixième problème de Hilbert est indécidable. L'insolubilité du problème de l'arrêt des machines de Turing, combinée à la thèse de Church (qu'on admet), montre que l'Entscheidungsproblem n'a pas de solution et, en particulier, que l'arithmétique de Peano n'est pas algorithmiquement décidable.

Si elle l'avait été, le dixième problème de Hilbert aurait été résolu puisque l'énoncé « telle équation diophantienne a une solution » se traduit très naturellement en une formule de l'arithmétique de Peano. Comme elle ne l'est pas, le dixième problème reste ouvert, puisque, sauf preuve du contraire, il semble ne concerner qu'un sous-ensemble des formules exprimables dans l'arithmétique de Peano.

En 1970, Youri Matiiassevitch mit la dernière pierre à la démonstration que le dixième problème de Hilbert n'a pas de solution non plus. Il suivit une démarche fortement inspirée de celles de Gödel et de Turing. Il eut (avec d'autres) l'idée de s'intéresser aux ensembles de nombres qui peuvent être définis par une équation diophantienne. Il montra en 1970, que ces ensembles « diophantiens » ne sont rien d'autre que les ensembles semi-décidables par machine de Turing. Ce résultat, combiné à l'insolubilité du problème de l'arrêt des machines de Turing et à la thèse de Church, montra que le dixième problème de Hilbert n'a pas de solution (voir la section 2.4.4 pour des détails).

6. Hilbert participa à un autre congrès international des mathématiques, en 1928. Lors de sa conférence, il aborda explicitement les questions de la consistance et de la complétude des mathématiques. La même année, il publia un livre avec Ackermann, où il mentionna, en plus, la question de la décidabilité. Voir [18, page 104]. Toutefois, il faut faire attention : il y a aujourd'hui plusieurs définitions différentes, pas équivalentes, de la notion de complétude par exemple. Ces définitions étaient inconnues à l'époque de Hilbert et il est difficile, voire impossible, de savoir à laquelle il pensait. Une partie importante du travail des logiciens a probablement consisté à clarifier ces notions, et donc à donner un, ou plusieurs, sens aux questions posées par Hilbert.

2.3 Commentaires

2.3.1 Intelligence artificielle et démonstration automatique

Un système expert est un logiciel tentant de reproduire les raisonnements d'un expert à partir d'une base de faits, d'une base de règles et d'un moteur d'inférence. La production de nouveaux faits à partir des faits connus et des règles, par le moteur d'inférence est très proche de la production de théorèmes dans une théorie axiomatique.

Les théorèmes d'incomplétude de Gödel constituent une limitation théorique incontournable pour ces systèmes. Toutefois, les vraies difficultés rencontrées dans leur mise au point sont davantage posées par la base de faits (modélisation de l'expérience et des connaissances pratiques de l'expert) que par les limitations théoriques des théorèmes de Gödel.

Des branches recherche davantage concernées sont les logiciels d'aide à la démonstration automatique et notamment, à la démonstration automatique d'algorithmes mathématiques (Theorema, Coq).

2.3.2 Calcul formel

Les logiciels de calcul formel sont des logiciels qui permettent de manipuler des expressions mathématiques sous forme symbolique et de leur appliquer divers traitements. Le résultat de Matiiassevitch implique directement qu'il ne peut exister aucun solveur général et infaillible d'équations diophantiennes. Remarquer que cela n'empêche pas de disposer de solveurs pour des cas particuliers. Par exemple, il existe des algorithmes efficaces de résolution d'équations diophantiennes en une variable.

Par effet domino, le résultat de Matiiassevitch a amené d'autres théorèmes d'indécidabilité algorithmique, qui limitent les algorithmes de simplification d'expressions mathématiques. Un exemple frappant est dû à Caviness, Richardson et Matiiassevitch :

Soit $\mathscr{F}$ l'ensemble de toutes les formules qu'on peut construire avec des rationnels, π, un symbole x, les opérations $+$, $\times$, sin et abs. Par exemple, $\sin(x+\pi/3)$ et $|x|+1/2$ sont des éléments de $\mathscr{F}$. Il n'existe aucun algorithme qui prenne en entrée deux formules quelconques de $\mathscr{F}$ et décide si ces formules représentent la même fonction de $\mathbb{R}$ dans $\mathbb{R}$.

Ce résultat implique qu'il n'existe aucun algorithme qui prenne en entrée une formule quelconque de $\mathscr{F}$ et décide si cette fonction est la fonction identiquement nulle. Il est donc impossible d'écrire un algorithme général et complet de simplification d'expressions de $\mathscr{F}$ puisque cet algorithme serait incapable de reconnaître zéro !

2.3.3 Relations avec la théorie de Kuhn

Question d'un étudiant : est-ce que les axiomes de la théorie des ensembles (par exemple) peuvent être vus comme un paradigme, au sens de Kuhn ?

Rappelons rapidement, pour commencer, quelques traits marquants des paradigmes (voir la section 1.3 pour plus de précisions) : (1) en cas d'anomalie, on cherche l'erreur ailleurs que dans le paradigme ; (2) un paradigme joue le rôle de programme de recherche ; (3) un

paradigme façonne la vision du monde des chercheurs qui l'adoptent ; (4) on ne rejette pas un paradigme tant qu'un autre n'est pas prêt à lui succéder ; (5) lors des changements de paradigmes, il arrive que des problèmes considérés comme scientifiques deviennent vides de sens.

Je ne pense pas qu'une théorie axiomatique précise puisse être vue comme un paradigme. Voici quelques arguments.

Il existe au moins deux versions de la théorie des ensembles : avec ou sans, ce qu'on appelle l'axiome « du choix ». Il existe des théorèmes qui ne peuvent pas se démontrer sans cet axiome. Plus généralement, les axiomes qui constituent les théories axiomatiques sont vus comme des ensembles arbitraires et donc susceptibles d'être adaptés si le besoin s'en faisait sentir. Je vois là une inadéquation avec (3).

En un sens, une théorie axiomatique pourrait être vue comme un programme de recherche, visant à démontrer « tous » les théorèmes qu'elle implique. Je ne pense pas que ce soit une bonne façon de voir les choses. Dans leur activité de tous les jours, les mathématiciens, pour la plupart, ne partent pas des axiomes de la théorie des ensembles. Je pense que les axiomes de la théorie des ensembles ont été conçus pour donner un fondement solide à des constructions qui pré-existaient. Je les vois davantage comme une tentative de clarification ou de justification de l'activité des mathématiciens que comme le moteur d'un programme de recherche.

Si je devais voir des paradigmes, je les verrais plutôt « en amont ».

L'idée ancienne que les axiomes de la géométrie euclidienne sont « vrais » et que les axiomes d'une théorie quelconque peuvent être établis et justifiés en fonction de leur évidence [15, chapitre 2, page 23] me semble pouvoir être un élément d'un paradigme.

Lorsque cette idée a été abandonnée, il a manifestement dû y avoir un véritable changement de vision du monde chez les mathématiciens (point 2). Un autre indice (5) de la présence d'un paradigme est donné par Bourbaki, qui écrit [1, pages 26-27] : « Gauss et Lobatchevsky croient que le débat entre les différentes géométries peut être tranché par l'expérience [...] mais c'est là un problème qui n'a plus rien à voir avec la mathématique ».

Si les théories axiomatiques, prises isolément, ne me semblent pas pouvoir être considérées comme des paradigmes, peut-être que l'idée que toutes les théories mathématiques devraient être axiomatisées [15, chapitre 1, page 19] a pu faire partie d'un paradigme, qui aurait succedé au précédent.

On peut voir, là aussi, une certaine vision du monde, pour les mathématiques. Cette idée a manifestement joué un rôle de programme de recherche. Lorsque des anomalies sont apparues (les paradoxes survenus lors des premières tentatives d'axiomatisation de la théorie des ensembles), le programme de recherche n'a pas été remis en cause.

Que penser des théorèmes d'incomplétude de Gödel ou des résultats d'indécidabilité algorithmique dus à Turing (et à d'autres) ? On considère parfois qu'ils mettent fin à « la crise des fondements » des mathématiques. Avec le point de vue que nous avons pris, ils pourraient plutôt correspondre au début d'une crise, au sens de Kuhn. Ils pourraient être vus comme une première véritable « anomalie » qui n'invalide qu'une interprétation « extrême » du paradigme. Lire [15, Partie III]. Un grand nombre de mathématiciens a considéré que

les propositions « à la Gödel » qu'on ne peut, ni prouver, ni infirmer, sont des propositions construites sur mesure pour les besoins des théorèmes d'incomplétude mais ne correspondent pas aux préoccupations authentiques des mathématiciens [5, page 196]. Or un certain nombre de résultats récents montrent qu'il devient de plus en plus difficile d'ignorer les anomalies [5, 8]. La crise, alors, ne serait pas terminée.

2.4 Compléments

2.4.1 L'algorithme de Floyd

On considère une suite définie par récurrence et condition initiale

$$u_{n+1} = f(u_n)\,, \quad u_0 = a\,. \tag{2.2}$$

De deux choses l'une : ou bien tous les u_i sont différents ou bien il existe deux entiers naturels k et ℓ, avec $k \neq \ell$, tels que $u_k = u_\ell$. Dans ce cas, comme u_{n+1} est complètement déterminé par u_n, la suite est nécessairement périodique à partir de l'indice k. De telles suites sont dites « ultimement périodiques ».

Il existe un cas particulier où toute suite de la forme (2.2) est ultimement périodique : c'est le cas où les termes u_i appartiennent à un ensemble fini.

L'algorithme de Floyd **ne** démontre **pas** si une suite donnée est ultimement périodique ou pas ! Ce qu'il fait ? Si on sait à l'avance que la suite (2.2) est ultimement périodique (par exemple dans le cas où les termes u_i appartiennent à un ensemble fini) alors il permet de repérer l'entrée dans la période. Un autre algorithme, meilleur mais un peu plus compliqué, est dû à Brent [9, section 1.5.4].

Supposons donc la suite (2.2) ultimement périodique et définissons la suite (v_m) par $v_m = u_{2\,n+1}$. Il existe un indice i tel que $u_i = v_i$. La suite est alors périodique à partir de l'indice i. L'algorithme de la figure 2.3, page 33 applique ce principe.

```
procedure Floyd (f, a)
   La suite (2.2) est supposée ultimement périodique
begin
   u = a
   v = f(a)
   while u ≠ v do
      u = f(u)
      v = f(f(v))
   end do
l'entrée dans la période est détectée
   end
```

FIGURE 2.3 – L'algorithme de Floyd, en pseudo-langage

2.4.2 L'arithmétique de Peano

Voici les sept axiomes « arithmétiques » plus le schéma d'axiome qui correspond au raisonnement par récurrence. Il y a un symbole de constante : 0, un symbole s (qui signifie « successeur » : sx signifie $x+1$ par exemple), deux symboles $+$ et $\cdot$ pour désigner l'addition et la multiplication. Les symboles $\neg$, $\wedge$ et $\rightarrow$ correspondent aux opérateurs logiques « non », « et » et à l'implication.

1. $\forall x\ \neg(sx = 0)$
2. $\forall x\ \exists y\ (\neg x = 0 \rightarrow sy = x)$
3. $\forall x\ \forall y\ (sx = sy \rightarrow x = y)$
4. $\forall x\ (x + 0 = x)$
5. $\forall x\ \forall y\ (x + sy = s(x + y))$
6. $\forall x\ (x \cdot 0 = 0)$
7. $\forall x\ \forall y\ (x \cdot sy = (x \cdot y) + x)$

Voici le schéma d'axiome qui correspond au raisonnement par récurrence. On peut l'appliquer pour tout prédicat [7] du premier ordre [8] $P(x)$:

$$[P(0) \wedge (\forall x\ P(x) \rightarrow P(sx))] \rightarrow \forall x\ P(x)\,.$$

Pour obtenir une théorie axiomatique, il faut ajouter à ces sept axiomes et à ce schéma d'axiome, cinq schémas d'axiomes « logiques », où A, B et C sont des prédicats du premier ordre quelconques [3, chapitre 5, § 3] :

1. $A \rightarrow (B \rightarrow A)$
2. $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$
3. $(\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$
4. $\forall x\ A(x) \rightarrow A(t)$
5. $(A \rightarrow B) \rightarrow (A \rightarrow \forall x\ B)$

Enfin, il reste à fixer les règles d'inférence qui permettent, dans une démonstration, de déduire une formule à partir des formules précédentes. Dans notre cas, il s'agit du « modus ponens » et de la « généralisation » :

modus ponens. Des formules $A \rightarrow B$ et A on peut déduire B ;

généralisation. De la formule A, on peut déduire $\forall x\ A(x)$.

7. Intuitivement, un « prédicat » est une formule, dépendant éventuellement de variables, dont la valeur est « vrai » ou « faux ». Par exemple, « x est un nombre pair » est un prédicat alors que « $x+1$ » n'en est pas un. Cette définition intuitive nous suffira mais elle demanderait, en toute rigueur, des précisions. Voir [3, chapitre 6].

8. Un prédicat du premier ordre est un prédicat dont les variables ne peuvent prendre que des valeurs de type « entier ».

Cet ensemble d'axiomes, de schémas d'axiomes et de règles d'inférence forme un exemple de théorie complètement axiomatisée, qui est appelée « arithmétique de Peano du premier ordre ». C'est une théorie axiomatique de ce type (quoique pas exactement celle-là) que Kurt Gödel a considéré dans son article de 1931.

Remarque : lorsqu'on parle de « formule » dans la question **(b)** de la section 2.2.2 ou dans le paragraphe concernant le théorème de Gödel, on parle de « formules closes », c'est-à-dire de formules dont toutes les variables sont quantifiées.

Pour conclure, voici un exemple de démonstration (qui ne fait intervenir que les schémas d'axiomes logiques). Il s'agit d'une démonstration avec hypothèse : on montre qu'on peut déduire $\forall\ z\ p(z,z)$ à partir de l'hypothèse $\forall x\ \forall y\ p(x,y)$ [3, chapitre 6, page 100]. La démonstration est une suite de six formules f_i :

f_1	$\forall x\ \forall y\ p(x,y)$	(hypothèse)
f_2	$\forall x\ \forall y\ p(x,y) \to \forall y\ p(z,y)$	(quatrième schéma d'axiome logique)
f_3	$\forall y\ p(z,y)$	(modus ponens avec f_1 et f_2)
f_4	$\forall y\ p(z,y) \to p(z,z)$	(quatrième schéma d'axiome logique)
f_5	$p(z,z)$	(modus ponens avec f_3 et f_4)
f_6	$\forall z\ p(z,z)$	(généralisation)

2.4.3 Les machines de Turing

Les machines de Turing ont été inventées par Alan Turing en 1937 en vue de donner une définition précise au concept d'algorithme.

Bien qu'il s'agisse de constructions purement théoriques, destinées à faciliter certaines preuves, on en décrit une version [13, chapitre 5, page 77], comme s'il s'agissait d'une machine physique.

Une machine de Turing a une mémoire sous la forme d'un ruban, divisé en cellules. Ce ruban possède une seule extrémité. On peut supposer qu'elle est à gauche. Le ruban est donc potentiellement infini à droite, ce qui implique que, contrairement à ce qui se passe avec de vrais ordinateurs, une machine de Turing n'est jamais à court de mémoire. Par contre, tout calcul particulier ne met en œuvre qu'un nombre fini de cellules.

Les cellules sont un peu l'analogue des mots (ou des octets) des vrais ordinateurs. Une cellule peut être vide ou contenir un unique élément d'un alphabet fini de symboles $A = \{\alpha_1, \ldots, \alpha_w\}$. Des machines différentes peuvent avoir des alphabets différents. Un symbole, $*$, joue un rôle spécial : il marque la cellule la plus à gauche et n'est utilisé nulle part ailleurs. Un symbole supplémentaire, $\Lambda \notin A$, est utilisé pour noter le vide.

Les symboles sur le ruban sont lus et écrits par une tête de lecture/écriture qui, à tout instant du temps, supposé discret, se positionne sur une seule cellule du ruban. La tête peut se déplacer vers la gauche, vers la droite ou rester sur place.

La machine comporte une variable d'état, qui peut prendre un nombre fini de valeurs $q_1, \ldots, q_v$. Un de ces états, q_1, est réputé être l'état *initial*. Un ou plusieurs états sont réputés être des états *finaux*. Des machines différentes peuvent avoir un nombre différent d'états.

Le comportement de la machine est donné par un ensemble d'instructions, qui se pré-

sentent sous la forme de règles :

$$q_i\,\alpha_j \longrightarrow \alpha_k\,D_\ell\,q_m\,,$$

qu'on peut lire ainsi : si la machine est dans l'état q_i et la cellule courante (sous la tête de lecture) contient le symbole α_j, alors le contenu de la cellule courante devient le symbole α_k puis, la tête se déplace suivant la direction D_ℓ et l'état de la machine devient q_m.

Le symbole D_ℓ peut prendre trois valeurs possibles : L : la tête se déplace d'une cellule à gauche ; R : la tête se déplace d'une cellule à droite ; S : la tête ne se déplace pas.

Il y a quelques contraintes à respecter : le jeu d'instructions doit être complet et déterministe : pour tout état non final q_i et tout symbole $\alpha_j \in A \cup \{\Lambda\}$, il doit exister une et une seule règle de membre gauche $q_i\,\alpha_j$; la tête ne doit pas sortir du ruban : si α_j vaut $*$ alors α_k doit valoir $*$ et le déplacement D_ℓ doit être différent de L.

Au début, un segment initial (collé le plus à gauche possible) du ruban est initialisé par des symboles de A, sans trou, le reste du ruban potentiellement infini est vide, la tête est positionnée au-dessus d'une cellule et la machine est dans l'état q_1.

La machine exécute ses instructions étape par étape. Elle s'arrête dès que l'état courant est l'un des états finaux. Il est possible que la machine n'atteigne jamais un état final et que des étapes s'enchaînent indéfiniment.

Le segment initial est l'information donnée en entrée à la machine. Le résultat du calcul, la sortie, est déterminée par le contenu du ruban, la position de la tête sur le ruban et l'état final, lorsque le calcul s'arrête.

L'interprétation des informations d'entrée et de sortie sont des événements extérieurs à la machine de Turing. Supposons que les états finaux soient q_2 et q_3. On peut convenir qu'un arrêt dans l'état q_2 signifie « retourner *vrai* » et qu'un arrêt dans l'état q_3 signifie « retourner *faux* » mais cette interprétation est extérieure à la machine de Turing.

Par exemple, la machine suivante « retourne *vrai* » si le symbole sous la tête de lecture est 0. Elle « retourne *faux* » sinon. La tête de lecture ne se déplace pas. Le symbole présent sous la tête n'est pas modifié. L'alphabet est $A = \{*, 0, 1\}$. Cette machine représente l'instruction « if 1 then *true* else *false* ».

$$q_1 * \to * \,S\, q_3\,, \quad q_1\,1 \to 1\,S\,q_2\,, \quad q_1\,0 \to 0\,S\,q_3\,, \quad q_1\,\Lambda \to \Lambda\,S\,q_3\,.$$

Autre exemple. La machine suivante ne change jamais son état et ne s'arrête jamais. L'alphabet est le même que précédemment. Elle représente l'instruction « while *true* do done ».

$$q_1 * \to * \,S\, q_1\,, \quad q_1\,0 \to 0\,S\,q_1\,, \quad q_1\,1 \to 1\,S\,q_1\,, \quad q_1\,\Lambda \to \Lambda\,S\,q_1\,.$$

Des machines de Turing compliquées peuvent facilement être réalisées par assemblage de machines plus simples, quitte à renuméroter les états. Cette machine-ci a été obtenue par assemblage des deux machines précédentes. L'état q_1 de la deuxième machine a été renuméroté en q_4. Elle représente l'instruction « if 1 then *true* else while *true* do done ».

$$q_1 * \to * \,S\, q_4\,, \quad q_1\,1 \to 1\,S\,q_2\,, \quad q_1\,0 \to 0\,S\,q_4\,, \quad q_1\,\Lambda \to \Lambda\,S\,q_4\,,$$

$$q_4 * \to * \,S\, q_4\,, \quad q_4\,0 \to 0\,S\,q_4\,, \quad q_4\,1 \to 1\,S\,q_4\,, \quad q_4\,\Lambda \to \Lambda\,S\,q_4\,.$$

2.4.4 Le dixième problème de Hilbert

Un ensemble E de n-uplets $(x_1, \ldots, x_n)$ d'entiers naturels est dit *diophantien* s'il existe une équation diophantienne $P(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0$ telle que : un n-uplet d'entiers naturels $(x_1, \ldots, x_n)$ appartient à E si, et seulement si, il existe un m-uplet d'entiers naturels $(y_1, \ldots, y_m)$ tel que $P(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0$.

Par exemple, l'ensemble E des nombres pairs est un ensemble diophantien puisque

$$x \in E \quad \text{si, et seulement si} \quad \exists\, y \in \mathbb{N},\ x - 2\,y = 0\,.$$

Il peut être représenté par le polynôme $P(x, y) = x - 2\,y$.

On savait bien avant 1970 que tout ensemble diophantien est algorithmiquement semi-décidable. Pour s'en convaincre, on va définir une machine de Turing M qui semi-décide l'ensemble E des nombres pairs. Le procédé employé n'est pas très subtil dans le cas des nombres pairs, mais il présente l'avantage de pouvoir être appliqué à un ensemble diophantien quelconque.

On cherche une machine de Turing M qui prenne en entrée un entier x et qui s'arrête si, et seulement si, x est un nombre pair. L'idée consiste à énumérer toutes les valeurs possibles pour y, à évaluer le polynôme P et à arrêter la machine dès que le résultat vaut zéro [9].

En 1970, Matiiassevitch démontra l'autre inclusion [12] : tout ensemble algorithmiquement semi-décidable est diophantien.

On conclut cette section en montrant comment ce résultat permet de répondre au dixième problème de Hilbert. Soit H l'ensemble des couples (i, x) tels que la machine de Turing numéro i s'arrête lorsqu'on l'applique à la donnée x. Cet ensemble est algorithmiquement semi-décidable. D'après le théorème de Matiiassevitch, il est diophantien. Il existe donc un polynôme $P(i, x, y_1, \ldots, y_m)$ tel que $(i, x) \in H$ si, et seulement si, il existe un m-uplet $(y_1, \ldots, y_m)$ d'entiers naturels tels que $P(i, x, y_1, \ldots, y_m) = 0$. Quel que soit le couple $(\bar{\imath}, \bar{x})$ d'entiers naturels, on peut évaluer les deux premières variables du polynôme P et obtenir un polynôme $Q(y_1, \ldots, y_m) = P(\bar{\imath}, \bar{x}, y_1, \ldots, y_m)$ qui a au moins une solution si, et seulement si, $(\bar{\imath}, \bar{x}) \in H$. Supposons qu'il existe une solution au dixième problème. On pourrait décider de l'existence de solutions pour Q et donc décider de l'appartenance à H. Cette contradiction avec l'insolubilité du problème de l'arrêt montre que le dixième problème de Hilbert n'a pas de solution.

9. L'exemple est particulièrement simple, puisqu'il n'y a qu'une seule variable y. Dans le cas de polynômes P dépendant de deux variables y_1 et y_2, il suffit d'énumérer les couples de la façon suivante : $(y_1, y_2) = (0,0)$, $(1,0)$, $(0,1)$, $(2,0)$, $(1,1)$, $(0,2)$, ...

Bibliographie

[1] Nicolas Bourbaki. *Éléments d'Histoire des Mathématiques*, volume iv of *Collection Histoire de la Pensée*. Hermann, second edition, 1969.

[2] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58 :345–363, 1936.

[3] Jean-Paul Delahaye. *Outils logiques pour l'intelligence artificielle*. Éditions Eyrolles, Paris, 1988.

[4] Jean-Paul Delahaye. Les propositions indécidables. *Pour la Science*, 265 :104–109, novembre 1999.

[5] Jean-Paul Delahaye. Promenade au pays des indécidables. *Pour la Science*, 266 :196–200, décembre 1999.

[6] Jean-Paul Delahaye. L'incomplétude, le hasard et la physique. *Pour la Science*, 355 :90–94, mai 2007.

[7] Jean-Paul Delahaye. Les délicats paradoxes de Berry et de Skolem. *Pour la Science*, 525 :80–85, juillet 2021.

[8] Jean-Paul Delahaye. Mesurer le temps en allumant des mèches. *Pour la Science*, 527 :80–85, septembre 2021.

[9] Michel Demazure. *Cours d'algèbre. Primalité. Divisibilité. Codes*. nouvelle bibliothèque mathématique. Cassini, Paris, 1997.

[10] Andrew Hodges. *Alan Turing ou l'énigme de l'intelligence*. Bibliothèque scientifique Payot, Paris, 1988. Titre original : Alan Turing : the enigma of intelligence (1983).

[11] Thomas Samuel Kuhn. *La structure des révolutions scientifiques*, volume 791 of *Champs Sciences*. Éditions Flammarion, Paris, 1983. Titre original : The Structure of Scientific Revolutions (1962).

[12] Youri Matiiassevitch. Enumerable sets are diophantine. *Sov. Math. Dokl.*, 11 :354–357, 1970.

[13] Youri Matiiassevitch. *Le dixième problème de Hilbert. Son indécidabilité*. Axiomes. Masson, Paris, 1995. Titre original : Hilbert's Tenth Problem (en Russe, 1993).

[14] Marvin Minsky. *Computation. Finite and Infinite Machines*. Prentice-Hall, 1967.

[15] Ernest Nagel, James R. Newman, Kurt Gödel, and Jean-Yves Girard. *Le théorème de Gödel*. Sources du savoir. Éditions du Seuil, 1989.

[16] Émile Picard. La mathématique dans ses rapports avec la physique. In *Actes du IVème congrès international des mathématiques*, Rome, Italie, 1908. Gauthier-Villars. 10 avril 1908.

[17] Alan Mathison Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42 :230–265, 1936.

[18] Hao Wang. *Gödel's and Some Other Examples of Problem Transmutation*, pages 101–109. Birkhauser, 1991.

[19] Wikipedia. Gerhard Gentzen. `https://fr.wikipedia.org/wiki/Gerhard_Gentzen`.

[20] Wikipedia. Peano axioms. `http://en.wikipedia.org/wiki/Peano_axioms`.

[21] Wikipedia. Zermelo-Fraenkel set theory. `http://en.wikipedia.org/wiki/Zermelo-Fraenkel_set_theory`.

Chapitre 3

Pratiques de la science normale en informatique

Dans ce troisième cours, je décris une partie de mon activité de recherche en analysant *a posteriori* le dernier article que j'aie écrit, au moment où je rédige ces lignes. L'article en question, n'est pas un modèle à suivre (si j'avais à le réécrire, je le réécrirais différemment), mais il donne un matériau concret à analyser et à critiquer. Il s'agit d'un article « technique » et donc assez proche d'un article que pourrait écrire un jeune chercheur.

La progression de ce chapitre est la suivante : la section 3.1 présente les articles de recherche en général, et celui-là en particulier ; la section 3.2 raconte l'histoire de cet article, ce qui donne un bref aperçu d'un travail de recherche ; l'article lui-même est donné en section 3.4 ; la section 3.3 décrit quelques indices utilisés, un peu, pour l'évaluation de la recherche.

3.1 Articles de recherche

La majeure partie des articles de recherche sont des photographies instantanées de l'activité de recherche de leurs auteurs. À la différence des rapports d'activité, qui sont souvent écrits à intervalles de temps réguliers, les articles de recherche sont écrits lorsque des résultats de recherche suffisamment aboutis pour être présentés, sont établis.

Un article présente donc un ou plusieurs résultats de recherche. Il est soumis, soit à une revue (un journal), soit à une conférence. Il est alors évalué par des rapporteurs anonymes. Ces rapporteurs ne sont pas des scientifiques pris au hasard : ce sont normalement des scientifiques qui ont déjà travaillé sur cette question ou sur des questions proches. En fonction des rapports, le comité éditorial de la revue, ou le comité d'organisation de la conférence, décide d'accepter ou de rejeter l'article.

Prenons le cas d'un article soumis à une conférence. En raison des contraintes d'organisation de la conférence, les rapporteurs ont un délai assez strict pour établir leur rapport. En général, les auteurs savent si leur article est accepté ou pas au bout de deux mois. S'il l'est, ils renvoient au comité d'organisation une version définitive de l'article, obtenue en

appliquant les corrections mineures suggerées par les rapporteurs. L'article est imprimé dans les actes de la conférence. L'un des auteurs présente l'article, devant les autres participants, le jour venu. Dans le cas de l'article [1], l'exposé, en Anglais, a duré 25 minutes, suivies de 5 minutes de questions.

Prenons maintenant le cas d'un article soumis à une revue. Bien que les éditeurs fassent de plus en plus pression sur les rapporteurs pour qu'ils envoient leur rapport rapidement, les délais sont plus longs que pour une conférence. Ces délais peuvent varier fortement d'une sous-discipline à une autre. Ils sont dus, entre autres raisons, au fait que les rapporteurs peuvent demander des corrections majeures, exigeant une réécriture de l'article et une réévaluation. En « calcul formel », qui est une sous-discipline de l'informatique, il faut compter plusieurs années. Le plus long délai que je connaisse est de 20 ans (un article soumis en 1981 et publié en 2001). Le plus long délai concernant un de mes articles est de 11 ans (article soumis en 1998 et publié en 2009).

3.1.1 Qu'est-ce qu'un bon article ?

Sur le fond

En théorie, un bon article doit contenir au moins un résultat nouveau d'une importance suffisante, et être correct techniquement (il ne doit pas comporter d'erreurs). Bien que cette affirmation constitue une excellente ligne directrice, la réalité est plus compliquée.

Le critère de la nouveauté n'est pas toujours très clair. Je connais par exemple le cas d'un article qui montre comment appliquer une technique connue de physique statistique à des problèmes de modélisation en biologie. La nouveauté ne réside pas dans la technique elle-même mais dans le fait de l'appliquer à un domaine où elle ne l'est pas encore. Une autre difficulté posée par la nouveauté : en présence de résultats vraiment nouveaux, il peut être très difficile de trouver des rapporteurs. Des articles sont parfois rejetés, parce que le comité éditorial de la revue ne parvient pas à en trouver. Des résultats vraiment nouveaux peuvent aussi être rejetés parce que ... trop nouveaux. Un exemple célèbre est fourni par Boris Belousov, qui tenta sans succès de publier sa découverte de phénomènes oscillatoires en chimie, vers 1950 [3, section I.3]. Ses travaux furent toutefois reconnus en 1980, lorsqu'on lui attribua le prix Lénine, 10 ans après sa mort.

L'importance d'un résultat est un critère encore plus subjectif. Par exemple, on observe parfois des situations extrêmes de « bulles scientifiques » (au même titre qu'il existe des « bulles financières ») où un groupe de scientifiques travaillant sur des questions très proches, valident mutuellement leurs travaux, sans toujours réaliser que, vus de l'extérieur, les résultats qu'ils obtiennent semblent peu importants.

Un article doit-il être techniquement correct ? Bien sûr. Toutefois, dans les domaines que je connais en tous cas, plusieurs articles publiés, considérés après coup comme extrêmement importants, contenaient des erreurs. Je pense en particulier à un article, où l'algorithme principal s'appuie sur un sous-algorithme, censé faire partie de l'article, mais qui n'y apparaît pas ! Pour de bonnes raisons probablement puisque le problème à résoudre par ce sous-algorithme n'était pas élémentaire. Cet article techniquement incorrect est pourtant considéré

comme un article fondateur dans son domaine. Que penser des rapporteurs ? Ont-ils été négligents ? Ce n'est pas sûr. Je connais au moins un cas de rapporteurs, ayant laissé publier un article qu'ils savaient être techniquement incorrect, parce qu'ils jugeaient que son contenu « faisait bouger les lignes ». Les rapporteurs (et les comités) peuvent parfois agir ainsi pour que l'inventeur d'un résultat important ne soit pas dépossédé de son invention, en raison d'une erreur considérée comme « réparable ».

On n'a pas parlé de paradigmes. On remarque toutefois que les considérations ci-dessus, qui sont d'une nature assez sociologique, s'accordent plus facilement avec les théories de Kuhn qu'avec celles de Popper, ou celles des inductivistes.

Sur la forme

Un article de recherche doit comporter un nombre de pages raisonnable. Typiquement, entre 10 et 20 pages pour les articles de conférence, jusqu'à une quarantaine pour les articles de revues. Il y a bien sûr toujours des exceptions, comme l'article qui présente la célèbre « astuce de Rabinowitsch », sur une demi-page, dans les très sérieuses *Mathematische Annalen* [5].

Un article est écrit en Anglais (des revues acceptent encore le Français, en mathématiques notamment, mais cela devient de plus en plus rare). Il commence par un résumé (ou *abstract*), accompagné, éventuellement, de quelques mots-clefs. Le résumé est suivi d'une introduction, d'une ou plusieurs sections contenant le corps de l'article, d'une conclusion et d'une bibliographie. Cette structure est d'ailleurs suivie par le générateur aléatoire d'articles de recherche en informatique [6].

Le résumé. Long d'un paragraphe, il explique au lecteur ce qu'il trouvera dans l'article, s'il décide de le lire. Ce point mérite d'être souligné : parmi l'immense masse d'articles publiés chaque année, un chercheur n'en lit que très peu, en raison du temps et de la concentration que leur lecture demande. Le résumé et l'introduction ont pour rôle d'aider le chercheur à décider s'il va, ou non, lire l'article. Il s'agit donc d'aller droit au but. En termes compréhensibles par les étudiants qui assistent à ce cours, le résumé de l'article [1] se réduit à trois phrases : *Cet article présente un [nouvel] algorithme pour calculer [un certain objet]. Un test [algorithmique] est revisité [dans tel domaine] et généralisé [dans tel autre]. Une nouvelle caractérisation [d'un certain objet] est fournie.* Les éditeurs des actes de la conférence ne demandaient pas de mots-clefs.

L'introduction. Elle comporte plusieurs parties assez bien identifiées. Elle est écrite pour un certain type de lecteur (on n'écrit pas du tout la même introduction, suivant qu'on s'adresse aux chercheurs participant à la conférence, ou au grand public). Dans le cadre d'un article de conférence, il est naturel de s'adresser à un public correspondant, en gros, aux participants à la conférence. L'introduction peut commencer par motiver le lecteur à lire l'article, en présentant le contexte auquel appartient le résultat, les problèmes que le résultat résout, et les raisons pour lesquelles on pense que ces questions sont importantes. En simplifiant, les trois premiers paragraphes de l'introduction de l'article [1] sont organisés comme

suit. Premier paragraphe : on indique que les algorithmes étudiés dans l'article sont très liés à l'étude d'une certaine propriété mathématique. Pour le lecteur non averti, on explique que cette propriété est elle-même liée à une autre propriété mathématique, dont l'importance est mieux connue. Deuxième paragraphe : la propriété en question a de nombreuses applications. Suivent trois applications possibles, énoncées en une phrase chacune. Troisième paragraphe : l'une des questions que l'article résout avait été explicitement posée par les rapporteurs d'un précédent article de revue.

L'introduction peut ensuite rappeler les résultats de l'article, en précisant le résumé. Dans l'article [1], on indique qu'il y a trois résultats. Le premier est donné figure 1 et théorème 4 ; le deuxième, théorème 3 ; et le troisième, théorème 2.

On peut après comparer les résultats avec l'existant. Dans l'introduction de l'article [1], cette comparaison est à peine esquissée, parce qu'elle porte sur des détails qui ont semblé trop techniques à ce stade de l'article. Ces comparaisons avec l'existant sont très importantes puisque ce sont elles qui justifient la nouveauté des résultats. Dans l'article [1], elles sont en fait développées à la fin de la section 4.

Enfin, on peut terminer par un plan de l'article : la section 2 présente ceci, la section 3 présente cela, etc. Dans l'article [1], on ne l'a pas fait (exemple à ne pas suivre).

Les sections suivantes. Elles varient beaucoup d'un article à l'autre. D'une façon générale, il faut partir du principe que le lecteur aura du mal à comprendre l'article, et l'aider autant que possible à comprendre où il est arrivé et à lui expliquer ce qu'il va lire, avant qu'il le lise. Ce type de clarification est d'ailleurs souvent demandé par les rapporteurs. Reprenons l'exemple de l'article [1]. En début de section 3, on écrit que la section est entièrement constituée de notions connues, empruntées à un document bien précisé. Au cours de la section 4, qui contient les principaux résultats, les lemmes 2 et 3 sont présentés comme faciles. Une notation, peu utilisée jusque-là, est rappelée juste avant le théorème 1, et on précise que les lemmes 3, 4 et 5 sont introduits, parce qu'ils contribuent à prouver le théorème 2 et que, parmi eux, le lemme 5 est nouveau.

La conclusion. On peut rappeler les principaux résultats. On peut l'utiliser aussi pour ouvrir l'article en indiquant des prolongations possibles, des problèmes laissés en suspens. Dans le cas de l'article [1], il est clair aujourd'hui qu'une bonne conclusion aurait été possible mais on ne l'a pas faite à l'époque où on a écrit l'article (encore un exemple à ne pas suivre).

La bibliographie. Elle est très importante, puisqu'elle permet de reconnaître le travail des autres chercheurs [1] et de bien mettre en évidence les résultats nouveaux par rapport aux résultats connus (elle facilite ainsi le travail des lecteurs et des rapporteurs de l'article). Elle permet ainsi aux autres chercheurs de mener leur propre recherche bibliographique. Lorsqu'on cite un article, il est bien sûr souhaitable de donner une citation aussi précise que possible, et de préférer « voir [1, Lemma 5, page 66] », par exemple, à « voir [1] ».

1. Le nombre de citations faites à un article entre dans le calcul de certains indicateurs de performance des chercheurs (voir section 3.3).

3.2 Gestation de l'article

Je vais raconter un peu la gestation de l'article [1], qui est assez simple. D'autres articles ont eu des histoires très différentes. J'aurais bien du mal à donner une règle générale.

Une partie importante de l'évaluation des chercheurs est liée au nombre d'articles publiés. La décision d'écrire l'article [1] à ce moment-là plutôt qu'à un autre, est venue d'une considération de ce type. Connaissant les dates limites de soumission d'un article à la conférence CASC, nous avons choisi une question technique, qui semblait raisonnablement simple, laissée en suspens, faute de temps, lors d'un article précédent. Nous avons travaillé la question et, à force de travail, nous avons obtenu une réponse satisfaisante à notre question. Les grandes lignes des sections techniques de l'article ont alors été rédigées, ce qui était indispensable, pour nous assurer, en relisant les enchaînements de preuves, que notre résultat était techniquement correct. Au moment de rédiger l'introduction, la question de la nouveauté du résultat s'est posée. Nous nous sommes lancés dans une recherche bibliographique, en utilisant un célèbre moteur de recherche sur internet et nous nous sommes alors aperçus, deux semaines avant la date limite de soumission, que notre solution était nouvelle, mais tout juste : au moins deux autres articles étaient quasiment arrivés au même résultat, des années avant nous. Pour renforcer l'article, nous avons alors creusé une question annexe que nous souhaitions initialement conserver pour plus tard et, nous avons obtenu un joli résultat théorique, inattendu, qui a alors constitué notre résultat principal. Le temps de restructurer le document et, en particulier, l'introduction, nous étions à deux jours de la date limite de soumission. La proximité de la date limite explique en partie pourquoi on a oublié le plan de l'article dans l'introduction ainsi que la conclusion.

Comme le texte ci-dessus l'indique, j'ai tendance à procéder aux recherches bibliographiques après-coup. L'inconvénient est évident : le risque de passer du temps à réinventer des resultats connus. L'avantage : on a plus de chances d'obtenir un résultat original, puisqu'on évite d'être influencé par l'approche des autres chercheurs et de retomber sur leurs solutions. D'autres chercheurs suivent la démarche inverse. Dans tous les cas, cette recherche est obligatoire. C'est ce positionnement par rapport à l'existant qui différencie la « recherche » de la simple « curiosité scientifique ».

Certains aspects de l'histoire de cet article sont assez généraux. Par exemple, le « résultat inattendu » a été obtenu en travaillant les détails d'un point technique d'un algorithme, en apparence anodin. C'est assez courant, et je pense donc que, pour mener une activité de recherche, dans le cadre de la « science normale », il faut aimer clarifier les détails. En lisant entre les lignes, on comprend aussi que l'activité d'un chercheur peut être, parfois, stressante. Il vaut donc mieux la pratiquer en équipe.

3.3 Indices de performance

Il existe plusieurs indices de performance, aussi bien pour les chercheurs que pour les journaux. Il est utile de les connaître mais il ne faut pas leur attribuer plus d'importance qu'ils ne devraient en avoir.

3.3.1 L'indice de Hirsch

L'indice de Hirsch [4], ou h-index, d'un chercheur est calculé à partir des citations dont son travail fait l'objet. Un chercheur a un indice de Hirsch h s'il a publié h articles qui ont été cités au moins h fois. Voir [2].

Cet indice a des qualités mais aussi bien des défauts. Voici les principaux : il ne tient pas compte du nombre d'auteurs d'un article et, en particulier, du fait qu'un article peut avoir un auteur principal et plusieurs auteurs secondaires ; il ne tient pas compte du contexte de la citation (les citations négatives sont comptées au même titre que les citations positives) ; un article difficile, contenant un résultat important, peut être « masqué » par un article de vulagrisation, qui synthétise le résultat et recueille toutes les citations ; enfin, il peut être facilement manipulé grâce aux auto-citations.

3.3.2 Le facteur d'impact

Le facteur d'impact d'un journal est calculé à partir des citations faites aux articles publiés dans le journal, durant les deux années précédentes. Un journal a un facteur d'impact f si, sur les deux années écoulées, chacun des articles qu'il a publiés a été cité f fois, en moyenne.

3.4 L'article et les rapports

On donne ci-dessous l'article pris en exemple dans ce chapitre, après correction des erreurs mineures relevées par les rapporteurs. On donne ensuite le courrier d'acceptation de l'article de la part du responsable du comité de programme ainsi que les quatre rapports. On remarque que la feuille de style de l'article est imposée. Les rapports, anonymes, sont composés d'une « note » qui peut varier de 2 (accept) à -2 (reject) ainsi que de commentaires plus détaillés (les commentaires sont particulièrement importants, lorsqu'ils doivent justifier un rejet). Ici, ils sont de longueur moyenne. En lisant entre les lignes, on remarque que plusieurs rapporteurs ont eu des difficultés à comprendre l'article, malgré les efforts des auteurs.

On the Regularity Property of Differential Polynomials Modulo Regular Differential Chains*

François Boulier[1], François Lemaire[1], and Alexandre Sedoglavic[1]

Université Lille I, LIFL, 59655 Villeneuve d'Ascq, France
Francois.Boulier@univ-lille1.fr
Francois.Lemaire@lifl.fr
Alexandre.Sedoglavic@lifl.fr

Abstract. This paper provides an algorithm which computes the normal form of a rational differential fraction modulo a regular differential chain if, and only if, this normal form exists. A regularity test for polynomials modulo regular chains is revisited in the nondifferential setting and lifted to differential algebra. A new characterization of regular chains is provided.

1 Introduction

This paper is concerned by methods for deciding whether a polynomial f (multivariate, over a field, say, $\mathbb{Q}$) is regular (i.e. not a zerodivisor) modulo a polynomial ideal defined by a regular chain C, which is a set of polynomials. For casual readers, this regularity property may seem quite exotic, compared to (say) the membership property to polynomial ideals. It is however very important and is pretty much related to the problem of computing the solutions of the system of polynomial equations $C = 0$. For instance, if f is proved to be a zerodivisor, then a factorization of some element of C is exhibited, which permits to split the set of equations to be solved, into two simpler sets. Moreover, as we shall see, regularity testing is strongly related to the problem of computing normal forms of polynomials modulo the ideal defined by the regular chain C, which are canonical representatives of the residue class ring defined by C. These comments are stated in the nondifferential case, for simplicity. However, they all have a counterpart for polynomial differential equations, i.e. in differential algebra.

Normal forms have many applications. In differential algebra, they make it easier to compute power series solutions, as pointed out in [2]. In both nondifferential and differential algebra, they permit to search linear dependencies between rational fractions modulo regular chains, by searching linear dependencies between their normal forms, modulo "nothing" (one of the key ideas of [10],

* This work has benefited from the support of the French ANR (decision number ANR-2010-BLAN-0109-03). It benefited also of many exchanges with Markus Rosenkranz and Georg Regensburger.

developed in the differential case in [3]). The very same principle, applied on the derivatives of rational differential fractions, may help to find first integrals.

The motivation for this paper comes from very fruitful remarks of a few reviewers of [2]. In that paper, a normal form algorithm is given for rational differential fractions modulo regular differential chains [2, Figure 2, Algorithm NF]. This normal form algorithm ultimately relies on an algorithm for computing the inverse of a nondifferential polynomial, modulo the ideal defined by a nondifferential regular chain. However, the algorithm provided in [2] may fail to compute the inverse, even if the inverse does exist [2, last comments of section 4]. A few reviewers of [2] then asked if it is possible to provide a complete algorithm, based on regular chains related methods[1], for computing normal forms if, and only if, these normal forms exist. In this paper, we provide the following results:

1. a complete normal form algorithm (Figure 1 and Theorem 4) ;
2. a revisited algorithmic characterization of the polynomials which are regular modulo the ideal defined by a nondifferential regular chain (Theorem 1) and its generalization in differential algebra (Theorem 3) ;
3. a new characterization of regular chains (Theorem 2).

The first result is an answer to the reviewers request. The second one improves former results of [18] and [8] in the nondifferential setting. It completes the proof of [14, Theorem 2.4] and extends this theorem in differential algebra. The third one permits to generalize [14, Theorem 2.4] and [1, Theorem 6.1].

2 Basics of Differential Algebra

The reference books are [16] and [13]. A *differential ring* R is a ring endowed with finitely many, say m, abstract *derivations* $\delta_1, \ldots, \delta_m$ i.e. unary operations which satisfy the following axioms, for each derivation δ:

$$\delta(a+b) = \delta(a) + \delta(b), \quad \delta(a\,b) = \delta(a)\,b + a\delta(b), \qquad (\forall\ a,\, b \in R)$$

and which are assumed to commute pairwise. This paper is mostly concerned with a differential polynomial ring R in n *differential indeterminates* $u_1, \ldots, u_n$ with coefficients in a commutative differential field K of characteristic zero, say $K = \mathbb{Q}$. Letting $U = \{u_1, \ldots, u_n\}$, one denotes $R = K\{U\}$, following Ritt and Kolchin. The set of derivations generates a commutative monoid w.r.t. the composition operation. It is denoted:

$$\Theta = \{\delta_1^{a_1} \cdots \delta_m^{a_m} \mid a_1, \ldots, a_m \in \mathbb{N}\}$$

where $\mathbb{N}$ stands for the set of the nonnegative integers. The elements of Θ are the *derivation operators*. The monoid Θ acts multiplicatively on U, giving the infinite set ΘU of the *derivatives*.

[1] Observe that, in principle, each required inverse could be easily obtained by using Rabinowitsch's trick and by running the Buchberger algorithm. However, Gröbner bases are not regular chains related methods. Moreover, the method could be costly.

If A is a finite subset of R, one denotes (A) the smallest ideal containing A w.r.t. the inclusion relation and $[A]$ the smallest differential ideal containing A. Let $\mathfrak{A}$ be an ideal and $S = \{s_1, \ldots, s_t\}$ be a finite subset of R, not containing zero. Then

$$\mathfrak{A} : S^\infty = \{p \in R \mid \exists\ a_1, \ldots, a_t \in \mathbb{N},\ s_1^{a_1} \cdots s_t^{a_t}\, p \in \mathfrak{A}\}$$

is called the *saturation* of $\mathfrak{A}$ by the multiplicative family generated by S. The saturation of a (differential) ideal is a (differential) ideal [13, chapter I, Corollary to Lemma 1].

Fix a *ranking*, i.e. a total ordering over ΘU satisfying some properties [13, chapter I, section 8]. Consider some differential polynomial $p \notin K$. The highest derivative v w.r.t. the ranking such that $\deg(p, v) > 0$ is called the *leading derivative* of p. It is denoted $\mathrm{ld}\, p$. The leading coefficient of p w.r.t. v is called the *initial* of p. The differential polynomial $\partial p/\partial v$ is called the *separant* of p. If C is a finite subset of $R \setminus K$ then I_C denotes its set of initials, S_C denotes its set of separants and $H_C = I_C \cup S_C$.

A differential polynomial q is said to be *partially reduced* w.r.t. p if it does not depend on any proper derivative of the leading derivative v of p. It is said to be *reduced* w.r.t. p if it is partially reduced w.r.t. p and $\deg(q, v) < \deg(p, v)$. A set of differential polynomials of $R \setminus K$ is said to be *autoreduced* if its elements are pairwise reduced. Autoreduced sets are necessarily finite [13, chapter I, section 9]. To each autoreduced set C, one may associate the set $L = \mathrm{ld}\, C$ of the leading derivatives of C and the set $N = \Theta U \setminus \Theta L$ of the derivatives which are not derivatives of any element of L (the derivatives "under the stairs" defined by C).

The following definition is borrowed from [2, Definition 3.1].

Definition 1. *The set $C = \{c_1, \ldots, c_n\}$ is a* regular differential chain *if it satisfies the following conditions:*

- **a** *the elements of C are pairwise partially reduced and have distinct leading derivatives ;*
- **b** *for each $2 \leq k \leq n$, the initial i_k of c_k is regular in $K[N \cup L]/(c_1, \ldots, c_{k-1}) : (i_1 \cdots i_{k-1})^\infty$;*
- **c** *for each $1 \leq k \leq n$, the separant s_k of c_k is regular in $K[N \cup L]/(c_1, \ldots, c_k) : (i_1 \cdots i_k)^\infty$;*
- **d** *for any pair $\{c_k, c_\ell\}$ of elements of C, whose leading derivatives $\theta_k u$ and $\theta_\ell u$ are derivatives of some same differential indeterminate u, the Δ-polynomial*

$$\Delta(c_k, c_\ell) = s_\ell \frac{\theta_{k\ell}}{\theta_k}\, c_k - s_k \frac{\theta_{k\ell}}{\theta_\ell}\, c_\ell\,,$$

where $\theta_{k\ell}$ denotes the least common multiple of θ_k and θ_ℓ, is reduced to zero by C, using Ritt's reduction algorithm [13, chapter I, section 9].

Triangularity plus condition **b** is the *regular chain* condition of [1]. Autoreduced regular differential chains are the same objects as Ritt characteristic sets. See [2, Proposition 3.2].

3 The Normal Form of a Rational Differential Fraction

All the results of this section are borrowed from [2]. Let C be a regular differential chain of R, defining a differential ideal $\mathfrak{A} = [C] : H_C^\infty$. Let $L = \mathrm{ld}\, C$ and $N = \Theta U \setminus \Theta L$. The normal form of a rational differential fraction is introduced in [2, Definition 5.1 and Proposition 5.2], recalled below.

Definition 2. *Let a/b be a rational differential fraction, with b regular modulo $\mathfrak{A}$. A* normal form *of a/b modulo C is any rational differential fraction f/g such that*

1 *f is reduced with respect to C ;*
2 *g belongs to $K[N]$ (and is thus regular modulo $\mathfrak{A}$),*
3 *a/b and f/g are equivalent modulo $\mathfrak{A}$ (in the sense that $a\,g - b\,f \in \mathfrak{A}$).*

Proposition 1. *Let a/b be a rational differential fraction, with b regular modulo $\mathfrak{A}$. The normal form f/g of a/b exists and is unique. In particular,*

4 *a belongs to $\mathfrak{A}$ if and only if its normal form is zero ;*
5 *f/g is a canonical representative of the residue class of a/b in the total fraction ring of $R/\mathfrak{A}$.*

Moreover,

6 *each irreducible factor of g divides the denominator of an inverse of b, or of some initial or separant of C .*

Recall that the normal form algorithm relies on the computation of inverses of differential polynomials, defined below.

Definition 3. *Let f be a nonzero differential polynomial of R. An* inverse *of f is any fraction p/q of nonzero differential polynomials such that $p \in K[N \cup L]$ and $q \in K[N]$ and $f\,p \equiv q \mod \mathfrak{A}$.*

4 On the Regularity Property of Polynomials

Though this section only addresses algebraic (i.e. nondifferential) questions, we state it with the terminology of the differential algebra. Consider a triangular set C in the polynomial ring $S = K[N \cup L]$. The ideal defined by C, in S, is $\mathfrak{B} = (C) : I_C^\infty$. Assume that $C = \{c_1, \ldots, c_n\}$, that the leading derivative (leading variable) of c_k is x_k and that $x_1 < \cdots < x_n$. It is possible to define the *iterated resultant* of any polynomial f w.r.t. C as follows. See [18, Definition 5.2] or [19, Definition 1.2]. See [8, Definition 4] or [15, Definition 1] for a close definition. See [7] for a definition in a more general setting.

$$\mathrm{res}(f, C) = \mathrm{res}(\cdots \mathrm{res}(f, c_n, x_n), \ldots, c_1, x_1) \tag{1}$$

where $\mathrm{res}(f, c_k, x_k)$ denotes the usual resultant of f and c_k w.r.t. x_k. The next lemma is borrowed from [18, Lemma 5.2]. Together with the two following ones, it prepares the proof of Theorem 1.

Lemma 1. *Let f be any polynomial. There exist polynomials p, $q_1, \ldots, q_n$ such that*

$$p\, f = q_1\, c_1 + q_2\, c_2 + \cdots + q_n\, c_n + \mathrm{res}(f,\, C)\,. \tag{2}$$

Proof. By [17, section 5.8, identity (5.21)], there exist two polynomials p_n and g_n such that

$$p_n\, f = g_n\, c_n + \mathrm{res}(f,\, c_n,\, x_n)\,. \tag{3}$$

There exist two polynomials p_{n-1} and g_{n-1} such that

$$p_{n-1}\ \mathrm{res}(f,\, c_n,\, x_n) = g_{n-1}\, c_{n-1} + \mathrm{res}(\mathrm{res}(f,\, c_n,\, x_n),\, c_{n-1},\, x_{n-1}) \tag{4}$$

hence such that

$$p_{n-1}\, p_n\, f = p_{n-1}\, g_n\, c_n + g_{n-1}\, c_{n-1} + \mathrm{res}(\mathrm{res}(f,\, c_n,\, x_n),\, c_{n-1},\, x_{n-1})\,. \tag{5}$$

Continuing, we obtain (2).

The two following lemmas are easy.

Lemma 2. *Let f, g be two polynomials. Then* $\mathrm{res}(f\, g,\, C) = \mathrm{res}(f,\, C)\ \mathrm{res}(g,\, C)$.

Proof. By induction on the number n of elements of C. If $n = 1$ then the lemma is the well-known multiplicativity property of resultants. See [9, section 3.1, exercises 3, 8 and 10] or [7, page 349]. If $n > 1$, assume inductively that the lemma holds for $C_{n-1} = \{c_1, \ldots, c_{n-1}\}$. Then $\mathrm{res}(f\, g,\, C) = \mathrm{res}(\mathrm{res}(f\, g,\, c_n,\, x_n),\, C_{n-1})$. Then, by the induction hypothesis and the multiplicativity property of resultants, $\mathrm{res}(f\, g,\, C)$ is equal to $\mathrm{res}(\mathrm{res}(f,\, c_n,\, x_n),\, C_{n-1})\ \mathrm{res}(\mathrm{res}(g,\, c_n,\, x_n),\, C_{n-1})$, which, in turn, is equal to $\mathrm{res}(f,\, C)\ \mathrm{res}(g,\, C)$.

Lemma 3. *Let $2 \leq k < n$ be an index and f be any polynomial such that* $\deg(f,\, x_\ell) = 0$, *for $k < \ell \leq n$. There exists a positive integer m such that* $\mathrm{res}(f,\, C) = \mathrm{res}(f,\, C_k)^m$.

Proof. It is an easy consequence of Lemma 2 and of the fact that, if $\deg(f, x) = 0$ and $\deg(g, x) > 0$ then $\mathrm{res}(f,\, g,\, x) = f^{\deg(g,\, x)}$.

In the sequel, a polynomial $f \in S$ is said to be regular modulo $\mathfrak{B}$ (recall $\mathfrak{B} = (C) : I_C^\infty$) if it is a regular element of the ring $S/\mathfrak{B}$. Regular elements and zerodivisors of a ring are defined as in [21, chapter I, § 5].

Theorem 1. *Assume C is a regular chain. A polynomial f is regular modulo $\mathfrak{B}$ if, and only if,* $\mathrm{res}(f,\, C) \neq 0$. *Together with the iterated resultant $q = \mathrm{res}(f,\, C)$, one can compute a polynomial p such that*

$$p\, f = q \mod \mathfrak{B}$$

If f is a zerodivisor modulo $\mathfrak{B}$ then $q = 0$, else p/q is an inverse of f modulo $\mathfrak{B}$.

Proof. The triangularity of C ensures that $\mathrm{res}(f, C) \in K[N]$. Thus, if the first part of the Theorem is proved, the second one follows immediately by Lemma 1.

In order to prove the first part of the Theorem, we first show that we can reduce our problem to the zerodimensional case[2]. Denote $S_0 = K(N)[L]$, and $\mathfrak{B}_0 = (C) : I_C^\infty$ in the ring S_0. By [5, Theorem 1.6], the multiplicative family of the nonzero elements of $K[N]$, is regular modulo $\mathfrak{B}$. Thus the ring $S_0/\mathfrak{B}_0$ is a subring of the total ring of fractions of $S/\mathfrak{B}$ [21, chapter IV, § 9]. Thus, f is regular modulo $\mathfrak{B}$ in S if, and only if, $f/1$ is regular modulo $\mathfrak{B}_0$ in S_0 [21, chapter I, § 19, Corollary 1].

Assume C is a regular chain in S. Then it is a zerodimensional regular chain in S_0. By [8, Lemma 4], an element $f/1$ is regular modulo $\mathfrak{B}_0$ in S_0 if, and only if, $\mathrm{res}(f, C) \neq 0$. Therefore, a polynomial f is regular modulo $\mathfrak{B}$ if, and only if, $\mathrm{res}(f, C) \neq 0$.

The next three lemmas prepare Theorem 2, which gives a necessary and sufficient condition that a triangular set C needs to satisfy in order to be a regular chain. Thus, recall that C is only supposed to be a triangular set.

Lemma 4. *Assume $\mathfrak{B}$ is proper. Let f be any polynomial. If $\mathrm{res}(f, C) \neq 0$ then f is regular modulo $\mathfrak{B}$.*

Proof. Let $\mathfrak{p}$ be any associated prime ideal of $\mathfrak{B}$ (such a $\mathfrak{p}$ exists for $\mathfrak{B}$ is proper). Take Formula (2) modulo $\mathfrak{p}$. The triangularity of C implies that $\mathrm{res}(f, C) \in K[N]$. By [5, Theorem 1.6] and the hypothesis, $\mathrm{res}(f, C) \neq 0 \mod \mathfrak{p}$. Since the elements of C are zero modulo $\mathfrak{p}$, the polynomial f is nonzero modulo $\mathfrak{p}$, i.e. is regular modulo $\mathfrak{B}$ by [21, chapter IV, § 6, Corollary 3].

The following lemma is new.

Lemma 5. *Assume $\mathfrak{B}$ is proper. Assume that, for any polynomial f which is regular modulo $\mathfrak{B}$, we have $\mathrm{res}(f, C) \neq 0$. Then C is a regular chain.*

Proof. The initials of the elements of $C = \{c_1, \ldots, c_n\}$ are regular modulo $\mathfrak{B}$ by [21, chapter IV, § 6, Corollary 3, and § 10, Theorem 17] and the fact that $\mathfrak{B}$ is proper. Thus, by assumption, for each $1 \leq k \leq n$, we have $\mathrm{res}(i_k, C) \neq 0$, where i_k denotes the initial of c_k. Thus, by Lemma 3 and the fact that $\deg(i_k, x_\ell) = 0$ for $k \leq \ell \leq n$, we have $\mathrm{res}(i_k, C_{k-1}) \neq 0$, where $C_{k-1} = \{c_1, \ldots, c_{k-1}\}$. Then, by Lemma 4, the initial i_k is regular modulo $(C_{k-1}) : I_{C_{k-1}}^\infty$. Thus C is a regular chain.

The following lemma is part of [8, Theorem 1].

Lemma 6. *Let h denote the product of the initials of the elements $c_2, \ldots, c_n$ of C. If $\mathrm{res}(h, C) \neq 0$ then C is a regular chain.*

[2] We are actually proving a very close variant of [5, Theorem 1.1].

Proof. Denote $C_k = \{c_1, \ldots, c_k\}$, for $1 \leq k \leq n$. By Lemma 2, Lemma 3 and the hypothesis, $\mathrm{res}(i_k, C_{k-1}) \neq 0$, for all $2 \leq k \leq n$. The ideal $(C_1) : I_{C_1}^{\infty}$ is proper. By Lemma 4, and the fact that $\mathrm{res}(i_2, C_1) \neq 0$, the initial i_2 is regular modulo $(C_1) : I_{C_1}^{\infty}$. The set C_2 is thus a regular chain and $(C_2) : I_{C_2}^{\infty}$ is proper. By Lemma 4, and the fact that $\mathrm{res}(i_3, C_2) \neq 0$, the initial i_3 is regular modulo $(C_2) : I_{C_2}^{\infty}$. Continuing, one concludes that C is a regular chain.

In the following theorem, the implication $2 \Rightarrow 1$ is new. The equivalence between the other points is a consequence of [8, Theorem 1] and [14, Theorem 2.4].

Theorem 2. *Let C be a triangular set. The three following properties are equivalent.*

1. *C is a regular chain ;*
2. *a polynomial f is regular modulo $\mathfrak{B}$ if, and only if,* $\mathrm{res}(f, C) \neq 0$ *;*
3. $\mathrm{res}(h, C) \neq 0$*, where h denotes the product of the initials of the elements $c_2, \ldots, c_n$ of C.*

Proof. The implication $1 \Rightarrow 2$ is a corollary to Theorem 1. The implication $2 \Rightarrow 1$: since $\mathrm{res}(1, C) \neq 0$ for any triangular set C, Property 2 implies that $\mathfrak{B}$ is proper ; the implication is thus a corollary to Lemma 5. The implication $3 \Rightarrow 1$ is Lemma 6. The implication $2 \Rightarrow 3$: Property 2 implies that $\mathfrak{B}$ is proper ; thus h is regular modulo $\mathfrak{B}$ by [21, chapter IV, § 6, Corollary 3, and § 10, Theorem 17]. Thus $\mathrm{res}(h, C) \neq 0$.

Comparison of Theorem 1 with other works. Inspecting the proof of Lemma 6, we see that Property 3 is equivalent to [19, Definition 1.3 (normal ascending chains)], which refers to [20], i.e. that $\mathrm{res}(i_k, C) \neq 0$ for $2 \leq k \leq n$, where i_k denotes the initial of c_k. Therefore, normal ascending chains and regular chains are exactly the same objects.

An algorithm for computing the inverse of a polynomial modulo a regular chain can be found in [15, Algorithm 3]. This algorithm relies on the hypothesis that the polynomial to be inverted is regular modulo the ideal defined by the chain. It relies on a different method (linear system solving) and is not proved.

Another algorithm for computing the inverse of a polynomial modulo a regular chain can be found in [6, Algorithm Invert]. It is based on a Gröbner basis computation. It is based on Kalkbrener's definition of regular chains [12] and thus computes an inverse of a polynomial modulo the intersection of all the prime ideals which contain the ideal defined by the chain, which have dimension $|N|$ and do not meet the multiplicative family M generated by the nonzero elements of $K[N]$, i.e. modulo the radical of the ideal defined by the regular chain. However, [6] misses [5, Theorem 1.6] which implies that $\mathfrak{B}$ has the same set of associated prime ideals as its radical, hence that the computed inverse also is an inverse modulo $\mathfrak{B}$.

Theorem 1 enhances [8, Lemma 4] which is stated in the zerodimensional case only, and does not provide the inverse computation.

Theorem 1 enhances also [18, Proposition 5.3]. Indeed, this Proposition states that $\mathrm{res}(f, C) \neq 0$ if, and only if, the polynomial f does not annihilate on any "regular zero" of C, where "regular zeros" are defined as generic zeros of the associated prime ideals of $\mathfrak{B}$ which have dimension $|N|$ and do not meet the multiplicative family M generated by the nonzero elements of $K[N]$ (see [18, Definition 5.1]). However, [18] misses [5, Theorem 1.6] which states that this property is held by all the associated prime ideals of $\mathfrak{B}$. See the comments on [6, Algorithm Invert].

The fact that a polynomial f is regular modulo $\mathfrak{B}$ if, and only if, $\mathrm{res}(f, C) \neq 0$ is already stated in [14, Theorem 2.4]. However, the proof of that Theorem just refers to [8] and [18] and thus misses the use of [5, Theorem 1.6].

Relationship between Theorem 1 and [5, Theorem 1.6]. Theorem 1 implies "easily" [5, Theorem 1.6] in the particular case of regular chains, i.e. that the associated prime ideals of $\mathfrak{B}$ have dimension $|N|$ and do not meet the multiplicative family M generated by the nonzero elements of $K[N]$. This remark is interesting for [5, Theorem 1.6] is one of the most difficult results of the regular chains theory. See [5]. It stresses, moreover, the strong relationship between the two theorems.

Proof. The regular chain C is a triangular set. Thus, for any nonzero $f \in K[N]$ the iterated resultant $\mathrm{res}(f, C)$ also is a nonzero element of $K[N]$. Thus, by Theorem 1, for any associated prime ideal $\mathfrak{p}$ of $\mathfrak{B}$, we have $\mathfrak{p} \cap M = \varnothing$ whence $\dim \mathfrak{p} \geq |N|$. Since the initials of the element of C do not lie in $\mathfrak{p}$, the derivatives $x_1, \ldots, x_n$ are algebraically dependent over N modulo $\mathfrak{p}$ and $\dim \mathfrak{p} \leq |N|$. Therefore, $\dim \mathfrak{p} = |N|$.

Observe that Theorem 1 does not hold for general triangular sets, while [5, Theorem 1.6] does. This claim is easily proved by an example. Take $f = x - 1$ and $C = \{x^2 - 1, (x - 1)\, y^2 - 2\}$ with $x < y$. The set C is triangular but is not a regular chain, for the initial $x - 1$ of the second element of C is not regular modulo the ideal defined by the first element. The ideal $\mathfrak{B}$ is generated by $\{x + 1, y^2 + 1\}$. It is prime, hence equal to its unique associated prime, if we assume $K = \mathbb{Q}$. The polynomial f is regular modulo $\mathfrak{B}$. However, $\mathrm{res}(f, C) = 0$.

Computational comment. For computational purposes, it is desirable to avoid computing the resultant with respect to x_k of polynomials which do not both depend on x_k, as in [8, Definition 4] and [15, Definition 1]. The iterated resultant is then defined as follows:

$$\mathrm{res}(f, C) = \overline{\mathrm{res}}(\cdots \overline{\mathrm{res}}(f, c_n, x_n), \ldots, c_1, x_1)\,, \tag{6}$$

where $\overline{\mathrm{res}}(f, c_k, x_k)$ is equal to $\mathrm{res}(f, c_k, x_k)$ if $\deg(f, x_k) > 0$ else is equal to f. Lemma 1 still holds with this definition of iterated resultants. By Lemma 2 and Lemma 3, the vanishing conditions of the iterated resultant $\mathrm{res}(f, C)$ are the same with Formula (1) as with Formula (6). Therefore, Theorems 1 and 2 also hold with Formula (6).

Computation of algebraic inverses and normal forms. Consider the triangular set $C = \{(x-1)(x-2), y^2-1\}$ for the ordering $y > x$. Since the initials are equal to 1, it is a regular chain. Consider the polynomial $f = (x-1)y + (x-2)$. We have $pf = -1 = \mathrm{res}(f, C)$, where $p = (-yx + y + x - 2)(2x - 3)$. Thus f is regular modulo the ideal $\mathfrak{B} = (C) : I_C^\infty$. Its inverse is $-p$ modulo $\mathfrak{B}$. Observe that the function [2, Inverse] would have failed to compute the inverse of f, since it would have tried to invert the initial $x-1$ of f modulo $\mathfrak{B}$, which is a zerodivisor modulo $\mathfrak{B}$, before computing the remainder of y^2-1 by f in the algorithm provided in [2, Figure 5]. Therefore, $\mathrm{NF}(1/f, C)$ succeeds with the new algorithm, given in Figure 1, while it fails with the old one, because of the inverse computation of f, w.r.t. C.

5 On the Regularity Property of Differential Polynomials

In this section, C denotes a regular differential chain of the differential polynomial ring R, defining a differential ideal $\mathfrak{A} = [C] : H_C^\infty$. Let $L = \mathrm{ld}\, C$ and $N = \Theta U \setminus \Theta L$.

The following Theorem provides an algorithm for deciding if a differential polynomial is regular modulo a differential ideal defined by a regular differential chain, and, if it is, for computing an inverse of it.

Theorem 3. *Let f be any differential polynomial, r be its partial remainder w.r.t. C and h a product of initials and separants of C such that $hf = r \mod \mathfrak{A}$. Together with the iterated resultant $q = \mathrm{res}(r, C)$, it is possible to compute a polynomial p such that*

$$pr = q \mod (C) : I_C^\infty$$

If f is a zerodivisor modulo $\mathfrak{A}$ then $q = 0$, else hp/q is an inverse of f modulo $\mathfrak{A}$.

Proof. The key arguments are the following: on the one hand, by [4, Corollary 4 to Theorem 3], a differential polynomial is regular modulo $\mathfrak{A}$ if, and only if, its partial remainder with respect to C is regular modulo $\mathfrak{B} = (C) : H_C^\infty$; on the other hand, $\mathfrak{B} = (C) : I_C^\infty$ by [11, Lemma 6.1].

If f is a zerodivisor modulo $\mathfrak{A}$, then r is a zerodivisor modulo $\mathfrak{B}$ and $q = 0$ by Theorem 1. Assume f is regular modulo $\mathfrak{A}$. Then r is regular modulo $\mathfrak{B}$ and q is a nonzero element of $K[N]$ by Theorem 1. Since $\mathfrak{B} \subset \mathfrak{A}$, we have $hpf = q \mod \mathfrak{A}$. Thus hp/q is an inverse of f modulo $\mathfrak{A}$.

A complete algorithm for computing the normal form of a rational differential fraction is presented in Figure 1. This algorithm is obtained from [2, The NF function, Figure 2] by udpating the method applied for computing inverses.

Theorem 4. *Let a/b be a rational differential fraction and C be a regular differential chain. If b is a zerodivisor modulo $\mathfrak{A}$, then $\mathrm{NF}(a/b, C)$ raises an error, else $\mathrm{NF}(a/b, C)$ returns the normal form of a/b modulo C.*

Proof. The Theorem is simply a restatement of [2, Proposition 5.3], taking into account the fact that inverses are computed using a method (Theorem 3) which succeeds if, and only if, the polynomial to be inverted is invertible.

Comment. The algorithm presented in Figure 1 has a drawback with respect to [2, The NF function, Figure 2]: if the denominator of the rational fraction is a zerodivisor, the algorithm does not exhibit a factorization of some element of C. This drawback may be easily overcome if one computes resultants by means of pseudoremainder sequences.

---

function NF(a/b, C)
Parameters
  a/b is a rational differential fraction such that a, $b \in R$.
  C is a regular differential chain, defining a differential ideal $\mathfrak{A}$.
Result
  if b is regular modulo $\mathfrak{A}$, then the normal form of a/b modulo $\mathfrak{A}$, else an error
begin
Regularity test and inverse computation of the denominator
  Apply Theorem 3 over b:
    if b is a zerodivisor modulo $\mathfrak{A}$ then
      error "the denominator is a zerodivisor modulo $\mathfrak{A}$"
    end if
    Denote $p_\mathrm{b}/q_\mathrm{b}$ an inverse of b modulo $\mathfrak{A}$
Inverse computation of the separants (they are necessarily regular)
  Apply Theorem 1 over each separant s_i of $C = \{c_1, \ldots, c_n\}$ and
                denote p_i/q_i an inverse of s_i modulo $\mathfrak{A}$
  $(f_{n+2},\, g_{n+2}) := (p_\mathrm{b}\, a,\, q_\mathrm{b})$
  Using Ritt's partial reduction algorithm, compute $d_1, \ldots, d_n \in \mathbb{N}$ and
                $r_{n+1} \in K[N \cup L]$ such that $s_1^{d_1} \cdots s_n^{d_n}\, f_{n+2} \equiv r_{n+1} \mod \mathfrak{A}$
  $f_{n+1} := p_1^{d_1} \cdots p_n^{d_n}\, r_{n+1}$
  $g_{n+1} := q_1^{d_1} \cdots q_n^{d_n}\, g_{n+2}$
  Denote $x_i = \mathrm{ld}\, c_i$ $(1 \leq i \leq n)$ and assume $x_n > \cdots > x_1$
  for ℓ from n to 1 by -1 do
    $r_\ell := \mathrm{prem}(f_{\ell+1},\, c_\ell,\, x_\ell)$
    Let i_ℓ denote the initial of c_ℓ
    Let $d_\ell \in \mathbb{N}$ be such that $i_\ell^{d_\ell}\, f_{\ell+1} \equiv r_\ell \mod (c_\ell)$
Inverse computation of an initial (it is necessarily regular)
    Apply Theorem 1 over i_ℓ and denote p_ℓ/q_ℓ an inverse of i_ℓ modulo $\mathfrak{A}$
    $f_\ell := p_\ell^{d_\ell}\, r_\ell$
    $g_\ell := q_\ell^{d_\ell}\, g_{\ell+1}$
  end do
  return f_1/g_1
the rational fraction may be reduced by means of a gcd computation
of multivariate polynomials over the field K
end

Fig. 1. The NF function

Bibliography

[1] Philippe Aubry, Daniel Lazard, and Marc Moreno Maza. On the Theories of Triangular Sets. *Journal of Symbolic Computation*, 28:105–124, 1999.

[2] François Boulier and François Lemaire. A Normal Form Algorithm for Regular Differential Chains. *Mathematics in Computer Science*, 4(2):185–201, 2010. 10.1007/s11786-010-0060-3.

[3] François Boulier. Efficient computation of regular differential systems by change of rankings using Kähler differentials. Technical report, Université Lille I, 59655, Villeneuve d'Ascq, France, November 1999. Ref. LIFL 1999–14, presented at the MEGA 2000 conference. `http://hal.archives-ouvertes.fr/hal-00139738`.

[4] François Boulier, Daniel Lazard, François Ollivier, and Michel Petitot. Computing representations for radicals of finitely generated differential ideals. *Applicable Algebra in Engineering, Communication and Computing*, 20(1):73–121, 2009. (1997 Techrep. IT306 of the LIFL).

[5] François Boulier, François Lemaire, and Marc Moreno Maza. Well known theorems on triangular systems and the D^5 principle. In *Proceedings of Transgressive Computing 2006*, pages 79–91, Granada, Spain, 2006. `http://hal.archives-ouvertes.fr/hal-00137158`.

[6] Driss Bouziane, Abdelillah Kandri Rody, and Hamid Maârouf. Unmixed–Dimensional Decomposition of a Finitely Generated Perfect Differential Ideal. *Journal of Symbolic Computation*, 31:631–649, 2001.

[7] Laurent Busé and Bernard Mourrain. Explicit factors of some iterated resultants and discriminants. *Mathematics of Computation*, 78:345–386, 2009.

[8] Changbo Chen, François Lemaire, Marc Moreno Maza, and Wei Pan. Comprehensive Triangular Decompositions. In *Proceedings of CASC'07*, pages 73–101, 2007.

[9] David Cox, John Little, and Donal O'Shea. *Using Algebraic Geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer Verlag, New York, 2nd edition, 2005.

[10] Jean-Charles Faugère, Patricia Gianni, Daniel Lazard, and Teo Mora. Efficient computation of Gröbner bases by change of orderings. *Journal of Symbolic Computation*, 16:329–344, 1993.

[11] Évelyne Hubert. Factorization free decomposition algorithms in differential algebra. *Journal of Symbolic Computation*, 29(4,5):641–662, 2000.

[12] Mickael Kalkbrener. A Generalized Euclidean Algorithm for Computing Triangular Representations of Algebraic Varieties. *Journal of Symbolic Computation*, 15:143–167, 1993.

[13] Ellis Robert Kolchin. *Differential Algebra and Algebraic Groups*. Academic Press, New York, 1973.

[14] François Lemaire, Marc Moreno Maza, Wei Pan, and Yuzhen Xie. When does (T) equal sat(T) ? In *Proceedings of thee International Symposium on Symbolic and algebraic computation*, pages 207–214. ACM Press, 2008.

[15] Banghe Li and Dingkang Wang. *An Algorithm for Transforming Regular Chain into Normal Chain*, volume 5081 of *Lecture Notes in Computer Science*, pages 236–245. Springer Berlin / Heidelberg, 2008.

[16] Joseph Fels Ritt. *Differential Algebra*. Dover Publications Inc., New York, 1950.

[17] Bruno Louis van der Waerden. *Algebra*. Springer Verlag, Berlin, seventh edition, 1966.

[18] Dongming Wang. Computing Triangular Systems and Regular Systems. *Journal of Symbolic Computation*, 30:221–236, 2000.

[19] Lu Yang, Xiaorong Hou, and Bican Xia. A complete algorithm for automated discovering of a class of inequality-type theorems. *Science in China Series F: Information Sciences*, 44(1):33–49, 2001.

[20] Lu Yang, Jingzhong Zhang, and Xiaorong Hou. An Efficient Decomposition Algorithm for Geometry Theorem Proving Without Factorization. *Proceedings of ASCM*, pages 33–41, 1995.

[21] Oscar Zariski and Pierre Samuel. *Commutative Algebra*. Van Nostrand, New York, 1958. Also volumes 28 and 29 of the *Graduate Texts in Mathematics*, Springer Verlag.

Date: Sun, 22 May 2011 08:36:52 +0100
From: CASC 2011 <casc2011@easychair.org>
To: <Francois.Boulier@lifl.fr>
Subject: CASC 2011 notification for paper 7

Dear Authors,

I am happy to inform you that your submission 7

François Boulier: On the Regularity Property of Differential Polynomials Modulo Regular D
ifferential Chains

has been accepted for presentation at the CASC 2011 conference.

This decision is based upon the referee reports which are attached.

Please prepare the final version of your paper following very carefully the referee sugge
stions and using the Springer lncs style package for Latex2e, which you can download from
 http://www14.in.tum.de/CASC2011/call.html

Send the final version of your paper (LaTeX source, all necessary style and graphics file
s, and pdf) as a zip archive to me <koepf@mathematik.uni−kassel.de> until June 10, 2011 (
sharp deadline).

Thank you again for submitting your work to the conference, and best wishes,

Wolfram Koepf, PC Chair CASC 2011

---------- REVIEW 1 ----------
PAPER: 7
TITLE: On the Regularity Property of Differential Polynomials Modulo Regular Differential
 Chains
AUTHORS: François Boulier, François Lemaire and Alexandre Sedoglavic

OVERALL RATING: 2 (accept)

The article is very interesting. It is a prolongation of the previous paper [2] by the au
thors which was published last year. I recommend to accept this paper, but I have some re
marks that should be taken into account.

My suggestion is to explain the link between notions of regular differential chain and Ri
tt characteristic set. Such explanations are contained in [2], but for unfamiliar readers
 it would be better to repeat them. It is also would be better to say something about imp
lementations of the algorithm given in Figure 1. Finally, I suggest to describe more care
fully the applications of the algorithm.

---------- REVIEW 2 ----------
PAPER: 7
TITLE: On the Regularity Property of Differential Polynomials Modulo Regular Differential
 Chains
AUTHORS: François Boulier, François Lemaire and Alexandre Sedoglavic

OVERALL RATING: 1 (weak accept)

This paper is a continuation of a former paper [2] by the first two coauthors, which was
about computing normal forms of regular differential chains.

This paper gives an algorithm for computing normal forms of rational differential fractio
n modulo a regular differential chain. This algorithm is in some sense complete, compared
 with the normal form algorithms presented in [2], since the latter algorithm (or specifi
cally, the involved inverse function) may fail to return a desired normal form (or invers
e), even it does exist. In other sense, the algorithm in [2] is not complete.

Section 2 introduces some notion and definitions, Section 3 recalls results from [2]. It
seems the main results occur as Theorems 1−−4 in Sections 4 and 5.

For Theorem 1, the authors also pointed out this result has already been given in their I
SSAC2008 paper, while the proof in the current paper being revised a bit.

In the introduction part, the authors pointed out that the normal form algorithm in their
 former paper [2] relies on computing the inverse of a "nondifferential" polynomial, modu
lo the ideal defined by a "nondifferential" regular chain. However, I don’t see a clear p

resentation of this point in the paper. As you claimed, Theorem 3 generaizes is an easy c
onsequence of Theorem 1. Why? Where you have do the transformation from differential to n
ondifferential case? Could you make it more clearly ?

The main algorithm is a slight update of the normal form algorithm in [2].

For the above reasons, I am little concerned that the new results of this paper are not s
ufficient enough and most of the results directly come from their previous work or other
peoples' work. Or, perhaps the author need improve the paper on the structure and the wri
ting style.

One typo:
line −3 on page 3: in the expression $U=[A]:H_A^{infty}$, the A should be C

---------- REVIEW 3 ----------
PAPER: 7
TITLE: On the Regularity Property of Differential Polynomials Modulo Regular Differential
Chains
AUTHORS: François Boulier, François Lemaire and Alexandre Sedoglavic

OVERALL RATING: 2 (accept)

This paper deals with algorithms for differential algebra problems. More precisely, an al
gorithm for computing
the normal form of a rational differential fraction modulo a regular differential chain i
s given in Figure 1. This is an
amelioration of the one given in [2] which was not complete. The paper contains two other
contributions stated
on the top of page 2 (items 2 and 3).
These interesting results seem to me correct (I have not found any wrong statement), prov
ide a real contribution to this topic which
naturally fits the topics of the CASC conference. For these reasons I recommend to accept
the paper.
Below is a list of comments that could be taken into account while preparing a final vers
ion. In particular, in my opinion, Section 4 could be made clearer.

. Introduction, line 11: "the ideal defined BY the regular chain C"
. Page 2, Section 2, First Formula: δ is not defined so I think it is better to re
place it by δ_i for $i=1,\ldots,m$
. Page 2, Section 2, line 6: "mostly concerned WITH"?
. Page 2, Section 2, line 12: remove "the", i.e., "for the set of nonnegative integers".
. Page 3, at the bottom: precise what is A?
. Page 4, definition 2, point 3: it could be useful to precise what do
you mean by "equivalent modulo U"
. Page 4, before Lemma 1: it could be clearer to explain what are the
different points you are going to develop in the following of the
section. This way, the reader would have in mind where you are going
which is not so clear ...
. Page 4, Lemma 1: "There exist" without "s"
. Page 5: I'm not sure that it is really interesting to give the
proofs of the lemmas. I would rather suggest to emphasize on the
algorithmic part of Lemma 1 ... what are the computational
difficulties? Is−this implemented? Where?
. Page 5: Before and in Theorem 1: precise what is B?
. At this point it is not really clear whether Theorem 1 is new or
not?
. Page 6, the sentence before Lemma 4 is not clear at all. I think
that you want to say that Theorem 1 is true for regular chains but you
want an analog for triangular chains which motivates the following?
. Page 7, top of the page. The sentence before Theorem 2 is not
clear at all. The word "equivalence" in "The equivalence between the
other points is .." is confusing as I think that [12, Theorem 2.4] only gives the
implication 1 => 2 and no equivalence.
. Page 7, proof of Theorem 2. Here again it is not so clear. We only have
to prove 1 => 2 => 3 => 1. It could be clearer not to talk about 2 =>
1 even if it is the new point. This could be said in the text but not
in the proof.
. Page 7, in the middle of the page there is two times "It is based on
..." in two lines. Change one of them.
. Page 8, Section 5, in the first sentence, precise what is A?
. Page 9, it could be useful here to talk about implementations of the
new algorithm and to compare it to the one of [2]. Do you have an

implementation of the algorithm given in Figure 1?

---------- REVIEW 4 ----------
PAPER: 7
TITLE: On the Regularity Property of Differential Polynomials Modulo Regular Differential Chains
AUTHORS: François Boulier, François Lemaire and Alexandre Sedoglavic

OVERALL RATING: 2 (accept)

The authors present a complete normal form algorithm for the computation of the normal form of a rational differential equation modulo a regular differential chain, provided the normal form exists. The algorithm improves a previous one proposed by the first two authors in [2]. They also obtain a new characterization of regular chains.

The paper is interesting and it completes the results from [2]. It includes an improved algorithm and give new results on regular chains.

Suggestions for the authors:

Describe an example for which the algorithm in figure 1 works while the corresponding algorithm in [2] fail to compute the inverse of a nondifferential polynomial.

Bibliographie

[1] François Boulier, François Lemaire, and Alexandre Sedoglavic. On the Regularity Property of Differential Polynomials Modulo Regular Differential Chains. In *Proceedings of Computer Algebra in Scientific Computing, LNCS 6885*, pages 61–72, Kassel, Germany, 2011. `http://hal.archives-ouvertes.fr/hal-00599440`.

[2] Jean-Paul Delahaye. Mesurer les chercheurs. *Pour la Science*, 401 :82–87, mars 2011.

[3] Irving R. Epstein and John A. Pojman. *An Introduction to Nonlinear Chemical Dynamics*. Oxford University Press, 1998.

[4] Jorge E. Hirsch. An index to quantify an individual's scientific research output. *PNAS*, 102(46) :16569–16572, 2005.

[5] J. L. Rabinowitsch. Zum Hilbertschen Nullstellensatz. *Mathematische Annalen*, 102(1) :520, 1929.

[6] Jeremy Stribling, Max Krohn, and Dan Aguayo. SCIgen - An Automatic CS Paper Generator. `http://pdos.csail.mit.edu/scigen`.