LES SEPT PROBLÈMES MATHÉMATIQUES DU PRIX DU MILLÉNAIRE

Andrew Wiles, Pierre Deligne, Charles Fefferman,
John Milnor, Stephen Cook, Enrico Bombieri,
Arthur Jaffe et Edward Witten

Traduction par Nicolas Bacaër
avec l'aide de
Michel Balazard, Gérard Besson,
Jean-Louis Colliot-Thélène, Isabelle Gallagher,
Catherine Goldstein, Thibaut Lemoine,
Sylvain Perifel et Claire Voisin

Les sept problèmes mathématiques du prix du millénaire

Andrew Wiles, Pierre Deligne, Charles Fefferman,
John Milnor, Stephen Cook, Enrico Bombieri,
Arthur Jaffe et Edward Witten

Traduction par Nicolas Bacaër
avec l'aide de
Michel Balazard, Gérard Besson,
Jean-Louis Colliot-Thélène, Isabelle Gallagher,
Catherine Goldstein, Thibaut Lemoine,
Sylvain Perifel et Claire Voisin

Textes sources :
www.claymath.org/millennium-problems
Carlson (J.), Jaffe (A.) et Wiles (A.), *The Millennium Prize Problems*, Providence, American Mathematical Society, 2006.

Les textes ont été traduits automatiquement avec le logiciel DeepL puis relus et corrigés par

Nicolas Bacaër (IRD, Bondy)
nicolas.bacaer@ird.fr

Michel Balazard (CNRS, Marseille)
→ L'hypothèse de Riemann

Gérard Besson (CNRS, Grenoble)
→ La conjecture de Poincaré

Jean-Louis Colliot-Thélène (CNRS, Orsay)
→ La conjecture de Hodge

Isabelle Gallagher (École normale supérieure, Paris)
→ L'équation de Navier-Stokes

Catherine Goldstein (CNRS, Paris)
→ La conjecture de Birch et Swinnerton-Dyer

Thibaut Lemoine (Collège de France)
→ La théorie de Yang-Mills quantique

Sylvain Perifel (Université Sorbonne-Paris-Nord)
→ Le problème P = NP

Claire Voisin (CNRS, Paris)
→ La conjecture de Hodge

Couverture : Collège de France

Paris, 2025

Introduction

« Tu ne dois avoir honte d'écrire en ta langue. »

Joachim du Bellay,
La Défense et illustration de la langue française

Le 24 mai de l'an 2000 fut présentée à Paris au Collège de France une liste de sept problèmes mathématiques. Pour chaque problème, l'institut de mathématiques Clay offrait une récompense d'un million de dollars dans le cadre de son « prix du millénaire ». Le choix du lieu pour cette présentation s'inspirait de l'exposé de vingt-trois problèmes que David Hilbert fit à Paris (à la Sorbonne) en 1900 lors du deuxième Congrès international des mathématiciens [3].

L'institut Clay, situé dans le Massachussetts aux États-Unis, demanda également à des mathématiciens de renom de présenter ces sept problèmes de manière détaillée dans des textes en anglais qui furent publiés dans un livre en 2006 et que l'on trouve aussi sous forme de fichiers PDF sur le site de l'institut.

À la différence du texte de Hilbert, publié en allemand dans une revue en 1900 et traduit en français dans les actes du congrès en 1902, les présentations détaillées des sept problèmes ne semblent pas avoir fait l'objet d'une traduction intégrale en français à ce jour [1]. Il existe néanmoins un livre de vulgarisation sur les sept problèmes qui est disponible en français [2]. Rappelons qu'un des problèmes a été résolu, celui qui concerne la conjecture de Poincaré (voir par exemple [1, 4, 5]).

Lors du congrès de 1900 eut lieu une discussion suite à un exposé intitulé « Proposition d'un vœu pour l'adoption d'une langue scientifique universelle », fait par Léopold Leau, un promoteur de l'espéranto. Un mathématicien russe, Alexandre Vassiliev, fit remarquer que la traduction pouvait déjà rendre de nombreux services et fit adopter par la majorité des congressistes le vœu « que les Académies et Sociétés savantes de tous les pays étudient les moyens propres à remédier aux

1. La présentation d'Enrico Bombieri sur l'hypothèse de Riemann a déjà été traduite par Denise Chemla : www.denisevellachemla.eu/transc-Clay-Bombieri-HR-fr.pdf.

maux qui proviennent de la variété croissante des langues employées dans la littérature scientifique ».

De nos jours et plus particulièrement depuis les progrès spectaculaires de la traduction automatique neuronale, la traduction d'un texte mathématique est relativement facile. Avec un logiciel tel que `mathpix`, on peut transformer un fichier PDF en fichier LaTeX. Puis avec un traducteur automatique tel que DeepL, on obtient une première version en français, qu'il faut évidemment relire pour corriger les erreurs de traduction en s'aidant par exemple de Wikipédia pour retrouver la traduction de certains termes techniques. Enfin, la relecture par des spécialistes permet de s'assurer de la qualité du processus de traduction et de corriger certains détails. C'est cette méthode que l'on a employée dans ce recueil pour traduire les sept problèmes du « prix du millénaire ». L'institut Clay autorise la reproduction de ses textes par tout moyen pour l'enseignement ou la recherche dans un cadre non commercial.

On remercie M^mes^ Gallagher, Goldstein et Voisin et MM. Balazard, Besson, Colliot-Thélène, Lemoine et Perifel pour leur relecture attentive.

Nicolas Bacaër
Paris, février 2025

Bibliographie

[1] Bessières (Laurent), Besson (Gérard) et Boileau (Michel), « La preuve de la conjecture de Poincaré d'après Perelman », *Images des mathématiques*, 2006.

[2] Devlin (Keith), *Les Énigmes mathématiques du 3^e^ millénaire* (trad. C. Laroche), Paris, Le Pommier, 2005.

[3] Duporcq (E.), *Compte rendu du deuxième Congrès international des mathématiciens tenu à Paris du 6 au 12 août 1900*, Paris, Gauthier-Villars, 1902.

[4] O'Shea (Donal), *Grigori Perelman face à la conjecture de Poincaré* (trad. J. Randon-Furling), Paris, Dunod, 2007.

[5] Szpiro (George), *La Conjecture de Poincaré* (trad. B. Sigaud), Paris, JC Lattès, 2007.

Sommaire

La conjecture de Birch et Swinnerton-Dyer (par Andrew Wiles)

Une relation polynomiale $f(x, y) = 0$ en deux variables définit une courbe C_0. Si les coefficients du polynôme sont des nombres rationnels, alors on peut chercher des solutions de l'équation $f(x, y) = 0$ avec $x, y \in \mathbb{Q}$, autrement dit des points rationnels sur la courbe. Si l'on considère un modèle projectif non singulier C de la courbe, alors il est classé topologiquement par son genre, que l'on appelle aussi le genre de C_0. Remarquons que $C_0(\mathbb{Q})$ et $C(\mathbb{Q})$ sont soit tous deux finis, soit tous deux infinis. Mordell a conjecturé le résultat profond suivant, que Faltings a démontré en 1983 :

Théorème [9]. *Si le genre de* C_0 *est supérieur ou égal à 2, alors* $C_0(\mathbb{Q})$ *est fini.*

La démonstration n'est pas constructive : on ne possède pas d'algorithme pour trouver les points rationnels. Il existe une borne explicite sur le nombre de solutions, mais cela n'aide pas beaucoup à les trouver.

Le cas des courbes de genre zéro est beaucoup plus facile et a été traité en détail par Hilbert et Hurwitz [12]. Il se ramène explicitement au cas des équations linéaires ou quadratiques. Le premier cas est facile et le second est résolu par le critère de Legendre. En particulier, pour un modèle projectif non singulier C, on trouve que $C(\mathbb{Q})$ est non vide si et seulement si C a des points p-adiques pour tout nombre premier p, et ceci est à son tour déterminé par un nombre fini de congruences. Si $C(\mathbb{Q})$ est non vide, alors C est paramétré par des fonctions rationnelles et il existe une infinité de points rationnels.

Le cas le plus insaisissable est celui du genre 1. Il peut y avoir ou non des solutions rationnelles. On ne connaît aucune méthode pour déterminer dans quel cas se trouve une courbe donnée. De plus, lorsqu'il y a des solutions rationnelles, il peut y en avoir une infinité. Si un modèle projectif non singulier C possède un point rationnel, alors $C(\mathbb{Q})$ a une structure naturelle de groupe abélien dont ce point est l'élément neutre. Dans ce cas, on appelle C une courbe elliptique sur $\mathbb{Q}$. Pour un historique du développement de cette idée, voir [19]. En 1922, Mordell [15] a démontré que ce groupe est de type fini, répondant ainsi

à une hypothèse implicite de Poincaré.

THÉORÈME. *Si* C *est une courbe elliptique sur* $\mathbb{Q}$*, alors*

$$\mathrm{C}(\mathbb{Q}) \simeq \mathbb{Z}^r \oplus \mathrm{C}(\mathbb{Q})^{\mathrm{tors}}$$

pour un certain entier $r \geqslant 0$*, où* $\mathrm{C}(\mathbb{Q})^{\mathrm{tors}}$ *est un groupe abélien fini.*

L'entier r est appelé le rang de C. Il est nul si et seulement si $\mathrm{C}(\mathbb{Q})$ est fini. On peut trouver un modèle affine pour la courbe sous la forme de Weierstrass

$$\mathrm{C} : y^2 = x^3 + ax + b$$

avec $a, b \in \mathbb{Z}$. On note Δ le discriminant de cette cubique et l'on pose

$$\begin{aligned} \mathrm{N}_p &:= \#\left\{\text{solutions de } y^2 \equiv x^3 + ax + b \bmod p\right\}, \\ a_p &:= p - \mathrm{N}_p. \end{aligned}$$

On peut alors définir la série L incomplète de C (incomplète car on omet les facteurs eulériens pour les nombres premiers $p \mid 2\Delta$) par

$$\mathrm{L}(\mathrm{C}, s) := \prod_{p \nmid 2\Delta} \left(1 - a_p p^{-s} + p^{1-2s}\right)^{-1}.$$

On considère cela comme une fonction de la variable complexe s. On sait que ce produit eulérien converge pour $\mathrm{Re}(s) > 3/2$. Une conjecture remontant à Hasse (voir le commentaire sur 1952(d) dans [26]) prédisait que $\mathrm{L}(\mathrm{C}, s)$ devrait avoir un prolongement analytique en une fonction de s sur tout le plan complexe. Ceci a été démontré [1, 24, 25]. On peut maintenant énoncer le problème du prix du millénaire :

CONJECTURE (Birch et Swinnerton-Dyer). *Le développement de Taylor de* $\mathrm{L}(\mathrm{C}, s)$ *en* $s = 1$ *est de la forme*

$$\mathrm{L}(\mathrm{C}, s) = c(s-1)^r + \text{ termes d'ordres supérieurs}$$

avec $c \neq 0$ *et* $r = \mathrm{rang}(\mathrm{C}(\mathbb{Q}))$.

En particulier, cette conjecture affirme que $\mathrm{L}(\mathrm{C}, 1) = 0 \Leftrightarrow \mathrm{C}(\mathbb{Q})$ est infini.

REMARQUES. 1. Il existe une version plus précise de cette conjecture. Dans cette version, il faut définir les facteurs eulériens aux nombres

premiers $p \mid 2\Delta$ pour obtenir la série L complète, $\mathrm{L}^*(\mathrm{C}, s)$. La conjecture prédit alors que $\mathrm{L}^*(\mathrm{C}, s) \sim c^*(s-1)^r$ avec

$$c^* = |\text{Ш}_\mathrm{C}|\, \mathrm{R}_\infty w_\infty \prod_{p|2\Delta} w_p / \left|\mathrm{C}(\mathbb{Q})^{\mathrm{tors}}\right|^2 .$$

Ici, $|\text{Ш}_\mathrm{C}|$ est l'ordre du groupe de Tate-Chafarevitch de la courbe elliptique C, un groupe dont on ne sait pas en général s'il est fini bien que l'on conjecture qu'il l'est. Il compte le nombre de classes d'équivalence d'espaces homogènes de C qui ont des points dans tous les corps locaux. Le terme R_∞ est un déterminant $r \times r$ d'une matrice dont les entrées sont données par un accouplement de hauteur appliqué à un système de générateurs de $\mathrm{C}(\mathbb{Q})/\mathrm{C}(\mathbb{Q})^{\mathrm{tors}}$. Les w_p sont des facteurs locaux élémentaires et w_∞ est un multiple simple de la période réelle de C. Pour une définition précise de ces facteurs, voir [20] ou [22]. On espère qu'une démonstration de la conjecture donnera aussi une démonstration de la finitude de Ш_C.

2. La conjecture peut également être énoncée sur n'importe quel corps de nombres ainsi que pour les variétés abéliennes (voir [20]). Depuis l'énoncé de la conjecture originale, des conjectures beaucoup plus élaborées concernant les valeurs spéciales des fonctions L sont apparues ; elles sont dues à Tate, Lichtenbaum, Deligne, Bloch, Beilinson et d'autres (voir [2, 3, 21]). En particulier, elles relient les rangs des groupes des cycles à l'ordre d'annulation (ou l'ordre des pôles) de certaines fonctions L.

3. Il existe une conjecture analogue pour les courbes elliptiques sur les corps de fonctions. Dans ce cas, Artin et Tate [20] ont démontré que la série L possède un zéro d'ordre au moins r, mais la conjecture elle-même n'a pas été démontrée. Dans le cas d'un corps de fonctions, on sait maintenant qu'elle est équivalente à la finitude du groupe de Tate-Chafarevitch ([20], [17, corollaire 9.7]).

4. Une démonstration de la conjecture sous la forme la plus forte donnerait un moyen effectif pour trouver des générateurs pour le groupe des points rationnels. En fait, on a seulement besoin du fait que le terme Ш_C soit un entier dans l'expression de $\mathrm{L}^*(\mathrm{C}, s)$ ci-dessus, sans aucune interprétation comme l'ordre du groupe de Tate-Chararevitch. Ceci a été démontré par Manin [16] à condition que les courbes elliptiques soient modulaires, une propriété qui est maintenant connue pour toutes les courbes elliptiques grâce à [1,24,25] (une courbe elliptique modulaire apparaît comme un facteur de la jacobienne d'une courbe modulaire).

Les débuts de l'histoire

Les problèmes sur les courbes de genre 1 figurent en bonne place dans les *Arithmétiques* de Diophante. Il est facile de voir qu'une droite rencontre une courbe elliptique en trois points (en comptant la multiplicité) de sorte que si deux des points sont rationnels, le troisième l'est aussi [1]. En particulier, si une tangente est prise en un point rationnel, elle rencontre à nouveau la courbe en un point rationnel. Diophante a implicitement utilisé cette méthode pour obtenir une deuxième solution à partir d'une première. Il n'a cependant pas itéré ce processus. C'est Fermat qui a compris le premier qu'on pouvait parfois obtenir une infinité de solutions de cette manière. Fermat a également introduit une méthode de « descente » qui permet parfois de montrer que le nombre de solutions est fini ou même nul.

Un très vieux problème lié à celui des points rationnels sur les courbes elliptiques est le problème des nombres congruents. Une façon de l'énoncer est de demander quels nombres entiers peuvent apparaître comme l'aire d'un triangle rectangle dont les côtés ont des longueurs rationnelles. Ces nombres entiers sont appelés « nombres congruents ». Par exemple, Fibonacci a été confronté au problème de $n = 5$ à la cour de Frédéric II et a réussi à trouver un tel triangle. Il affirma en outre qu'un tel triangle n'existait pas pour $n = 1$, mais la démonstration était fausse. La première démonstration correcte fut donnée par Fermat. Le problème remontait à des manuscrits arabes du X$^{\text{e}}$ siècle (pour l'histoire, voir [27, chap. I, §7] et [7, chap. XVI]). Il est étroitement lié au problème de la détermination des points rationnels sur la courbe $\mathrm{C}_n : y^2 = x^3 - n^2 x$. En effet,

$$\mathrm{C}_n(\mathbb{Q}) \text{ est infini } \iff n \text{ est un nombre congruent.}$$

En supposant la conjecture de Birch et Swinnerton-Dyer (ou même l'affirmation plus faible que $\mathrm{C}_n(\mathbb{Q})$ est infini $\Leftrightarrow \mathrm{L}(\mathrm{C}_n, 1) = 0$), on peut montrer que tout $n \equiv 5, 6, 7 \bmod 8$ est un nombre congruent. De plus, Tunnell a montré (toujours en supposant la conjecture) que pour n impair et sans facteur carré

$$\begin{aligned} & n \text{ est un nombre congruent} \\ \iff \quad & \#\left\{x, y, z \in \mathbb{Z} : 2x^2 + y^2 + 8z^2 = n\right\} \\ & = 2 \times \#\left\{x, y, z \in \mathbb{Z} : 2x^2 + y^2 + 32z^2 = n\right\} \end{aligned}$$

1. Il semblerait que ce soit Newton qui l'ait explicitement mentionné pour la première fois.

avec un critère similaire si n est pair [23]. Tunnell a démontré l'implication de gauche à droite inconditionnellement à l'aide du théorème principal de [5] décrit ci-dessous.

L'histoire récente

C'est l'article de Poincaré, publié en 1901, qui est à l'origine de la théorie moderne des points rationnels sur les courbes et qui a soulevé pour la première fois la question du nombre minimal de générateurs de $\mathrm{C}(\mathbb{Q})$. La conjecture elle-même a été énoncée pour la première fois sous la forme que l'on a donnée au début des années soixante [4]. Elle a été trouvée expérimentalement en utilisant l'un des premiers ordinateurs EDSAC à Cambridge. Le premier résultat général démontré concernait les courbes elliptiques à multiplication complexe. Les courbes à multiplication complexe appartiennent à un nombre fini de familles comprenant $\{y^2 = x^3 - \mathrm{D}x\}$ et $\{y^2 = x^3 - k\}$ avec $\mathrm{D}, k \neq 0$ variables. Ce théorème a été démontré en 1976 par Coates et Wiles [5]. Il affirme que si C est une courbe à multiplication complexe et si $\mathrm{L}(\mathrm{C}, 1) \neq 0$, alors $\mathrm{C}(\mathbb{Q})$ est fini. En 1983, Gross et Zagier ont montré que si C est une courbe elliptique modulaire avec $\mathrm{L}(\mathrm{C}, 1) = 0$ mais $\mathrm{L}'(\mathrm{C}, 1) \neq 0$, alors une construction antérieure de Heegner donne en fait un point rationnel d'ordre infini. En utilisant de nouvelles idées ainsi que ce résultat, Kolyvagin a montré en 1990 que pour les courbes elliptiques modulaires, si $\mathrm{L}(\mathrm{C}, 1) \neq 0$ alors $r = 0$, et si $\mathrm{L}(\mathrm{C}, 1) = 0$ mais $\mathrm{L}'(\mathrm{C}, 1) \neq 0$ alors $r = 1$. Dans le premier cas, Kolyvagin a eu besoin d'une hypothèse analytique qui a été confirmée peu après ; voir [6] pour l'histoire de cette hypothèse et pour d'autres références. Enfin, comme indiqué dans la remarque 4 ci-dessus, on sait maintenant que toutes les courbes elliptiques sur $\mathbb{Q}$ sont modulaires, ce qui nous permet d'obtenir le résultat suivant :

THÉORÈME. *Si* $\mathrm{L}(\mathrm{C}, s) \sim c(s-1)^m$ *avec* $c \neq 0$ *et* $m = 0$ *ou 1, alors la conjecture est vérifiée.*

Dans les cas où $m = 0$ ou 1, des résultats plus précis sur c (qui dépend bien sûr de la courbe) sont connus grâce aux travaux de Rubin et Kolyvagin.

Les points rationnels sur les variétés de dimension supérieure

On a commencé par discuter des propriétés diophantiennes des courbes et on a vu que le problème qui consiste à donner un critère pour savoir si $C(\mathbb{Q})$ est fini ou non ne se pose que pour les courbes de genre 1. De plus, d'après la conjecture ci-dessus, dans le cas du genre 1, $C(\mathbb{Q})$ est fini si et seulement si $L(C, 1) \neq 0$. En dimension supérieure, si V est une variété algébrique, on conjecture (voir [14]) que si l'on retire de V (l'adhérence de) toutes les sous-variétés qui sont des images de $\mathbb{P}^1$ ou de variétés abéliennes, alors la variété ouverte restante W devrait avoir la propriété que $W(\mathbb{Q})$ est finie. Cela a été démontré par Faltings dans le cas où V est elle-même une sous-variété d'une variété abélienne [10].

Cela suggère que pour trouver un nombre infini de points sur V, il faut chercher des courbes rationnelles ou des variétés abéliennes dans V. Dans ce dernier cas, on peut espérer utiliser des méthodes liées à la conjecture de Birch et Swinnerton-Dyer pour trouver des points rationnels sur la variété abélienne. À titre d'exemple, considérons la conjecture d'Euler de 1769 selon laquelle $x^4+y^4+z^4 = t^4$ n'a pas de solutions non triviales. En trouvant une courbe de genre 1 sur la surface et un point d'ordre infini sur cette courbe, Elkies [8] a trouvé la solution

$$2682440^4 + 15365639^4 + 18796760^4 = 20615673^4.$$

Son raisonnement montre qu'il existe une infinité de solutions de l'équation d'Euler.

En conclusion, bien que l'on ait réussi au cours des cinquante dernières années à limiter le nombre de points rationnels sur les variétés, il n'existe encore pratiquement aucune méthode pour trouver de tels points. Il faut espérer qu'une démonstration de la conjecture de Birch et Swinnerton-Dyer permettra de mieux comprendre ce problème général.

Bibliographie

[1] Breuil (C.), Conrad (B.), Diamond (F.) et Taylor (R.), « On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises », *J. Amer. Math. Soc.*, n° 14, 2001, p. 843-939.

[2] Beilinson (A.), « Notes on absolute Hodge cohomology », dans *Applications of Algebraic K-Theory to Algebraic Geometry and Number Theory*, Providence, American Mathematical Society, 1986, p. 35-68.

[3] Bloch (S.), « Height pairings for algebraic cycles », *J. Pure Appl. Algebra*, n° 34, 1984, p. 119-145.

[4] Birch (B.) et Swinnerton-Dyer (H.), « Notes on elliptic curves II », *J. reine angew. Math.*, n° 218, 1965, p. 79-108.

[5] Coates (J.) et Wiles (A.), « On the conjecture of Birch and Swinnerton-Dyer », *Invent. math.*, n° 39, 1977, p. 223-251.

[6] Darmon (H.), « Wiles' theorem and the arithmetic of elliptic curves », dans *Modular Forms and Fermat's Last Theorem*, Berlin, Springer, 1997, p. 549-569.

[7] Dickson (L.), *History of the Theory of Numbers*, vol. II, Washington, Carnegie Institute, 1919. Réimpr., Providence, American Mathematical Society, 1999.

[8] Elkies (N.), « On $A^4 + B^4 + C^4 = D^4$ », *Math. Comput.*, n° 51, 1988, p. 825-835.

[9] Faltings (G.), « Endlichkeitsätze für abelsche Varietäten über Zahlkörpern », *Invent. math.*, n° 73, 1983, p. 549-576.

[10] Faltings (G.), « The general case of S. Lang's conjecture », dans *Barsotti Symposium in Algebraic Geometry*, Boston, Academic Press, 1994, p. 175-182.

[11] Gross (B.) et Zagier (D.), « Heegner points and derivatives of L-series », *Invent. math.*, n° 84, 1986, p. 225-320.

[12] Hilbert (D.) et Hurwitz (A.),« Über die diophantischen Gleichungen von Geschlecht Null », *Acta Math.*, n° 14, 1890, p. 217-224.

[13] Kolyvagin (V.), « Finiteness of $E(\mathbb{Q})$ and $Ш(E, \mathbb{Q})$ for a class of Weil curves », *Math. USSR Izv.*, n° 32, 1989, p. 523-541.

[14] Lang (S.), *Number Theory III, Encyclopœdia of Mathematical Sciences*, vol. 60, Heidelberg, Springer, 1991.

[15] Mordell (L.), « On the rational solutions of the indeterminate equations of the third and fourth degrees », *Proc. Cambridge Phil. Soc.*, n° 21, 1922-1923, p. 179-192.

[16] Manin (Y.), « Cyclotomic fields and modular curves », *Russ. Math. Surv.*, n° 26-6, 1971, p. 7-78.

[17] Milne (J.), *Arithmetic Duality Theorems*, Boston, Academic Press, 1986.

[18] Poincaré (H.), « Sur les propriétés arithmétiques des courbes algébriques », *J. math. pures appl. (5)*, n° 7, 1901, p. 161-233.

[19] Schappacher (N.), « Développement de la loi de groupe sur une cubique », dans *Séminaire de théorie des nombres (Paris 1988/1989)*, Bâle, Birkhäuser, 1991, p. 159-184.

[20] Tate (J.), « On the conjectures of Birch and Swinnerton-Dyer and a geometric analog », *Séminaire Bourbaki*, n° 306, 1965-1966.

[21] Tate (J.), « Algebraic cycles and poles of zeta functions », dans Schilling (O. F. G.), *Arithmetical Algebraic Geometry : Proceedings of a Conference at Purdue University*, New York, Harper and Row, 1965, p. 93-110.

[22] Tate (J.), « The arithmetic of elliptic curves », *Invent. math.*, n° 23, 1974, p. 179-206.

[23] Tunnell (J.), « A classical diophantine problem and modular forms of weight 3/2 », *Invent. math.*, n° 72, 1983, p. 323-334.

[24] Taylor (R.) et Wiles (A.), « Ring-theoretic properties of certain Hecke algebras », *Ann. Math.*, n° 141, 1995, p. 553-572.

[25] Wiles (A.), « Modular elliptic curves and Fermat's last theorem », *Ann. Math.*, n° 142, 1995, p. 443-551.

[26] Weil (A.), *Collected Papers*, vol. II, New York, Springer, 1979.

[27] Weil (A.), *Basic Number Theory*, Boston, Birkhaüser, 1984.

La conjecture de Hodge (par Pierre Deligne)

Énoncé

Rappelons qu'une structure presque complexe sur une variété X de classe C^∞ et de dimension 2N est une structure de $\mathbb{C}$-module sur le fibré tangent T_X. Une telle structure de module induit une action du groupe $\mathbb{C}^*$ sur T_X, avec $\lambda \in \mathbb{C}^*$ agissant par multiplication par λ. Par transport de structures, le groupe $\mathbb{C}^*$ agit aussi sur toute puissance extérieure $\wedge^n T_X$, ainsi que sur le dual complexifié $\Omega^n := \mathcal{H}om\,(\wedge^n T_X, \mathbb{C})$. Si $p + q = n$, une (p,q)-forme est une section de Ω^n sur laquelle $\lambda \in \mathbb{C}^*$ agit par multiplication par $\lambda^{-p}\bar{\lambda}^{-q}$.

Supposons désormais que X soit analytique complexe. Une (p,q)-forme est alors une forme qui, en coordonnées holomorphes locales, peut s'écrire sous la forme

$$\sum a_{i_1,\ldots,i_p,j_1\ldots j_q}\, dz_{i_1} \wedge \cdots \wedge dz_{i_p} \wedge d\bar{z}_{j_1} \wedge \cdots \wedge d\bar{z}_{j_q}.$$

La décomposition $\Omega^n = \oplus \Omega^{p,q}$ induit une décomposition $d = d' + d''$ de la différentielle extérieure, avec d' (respectivement d'') de degré $(1,0)$ (respectivement $(0,1)$).

Si X est compacte et admet une métrique kählérienne, par exemple si X est une variété algébrique projective non singulière, cette action de $\mathbb{C}^*$ sur les formes induit une action sur la cohomologie. Plus précisément, $H^n(X,\mathbb{C})$ est l'espace des n-formes fermées modulo les formes exactes. Si l'on définit $H^{p,q}$ comme l'espace des (p,q)-formes fermées modulo le $d'd''$ des $(p-1,q-1)$-formes, alors l'application naturelle

$$\bigoplus_{p+q=n} H^{p,q} \to H^n(X,\mathbb{C}) \tag{1}$$

est un isomorphisme. Si l'on choisit une structure kählérienne sur X, on peut donner l'interprétation suivante de la décomposition (1) de $H^n(X,\mathbb{C})$: l'action de $\mathbb{C}^*$ sur les formes commute avec le laplacien et induit donc une action de $\mathbb{C}^*$ sur l'espace $\mathcal{H}^n$ des n-formes harmoniques. On a $\mathcal{H}^n \xrightarrow{\sim} H^n(X,\mathbb{C})$ et $H^{p,q}$ s'identifie à l'espace des (p,q)-formes harmoniques.

Lorsque X varie dans une famille holomorphe, la filtration de Hodge $\mathrm{F}^p := \bigoplus_{a \geqslant p} \mathrm{H}^{a,n-a}$ de $\mathrm{H}^n(\mathrm{X}, \mathbb{C})$ se comporte mieux que la décomposition de Hodge. Localement sur l'espace des paramètres T, $\mathrm{H}^n(\mathrm{X}_t, \mathbb{C})$ est indépendant de $t \in \mathrm{T}$ et la filtration de Hodge peut être considérée comme une filtration variable $\mathrm{F}(t)$ sur un espace vectoriel fixe. Elle varie de façon holomorphe avec t et obéit à la transversalité de Griffiths : au premier ordre autour de $t_0 \in \mathrm{T}$, $\mathrm{F}^p(t)$ reste dans $\mathrm{F}^{p-1}(t_0)$.

Jusqu'à présent, on a calculé la cohomologie en utilisant des formes C^∞. On aurait pu tout aussi bien utiliser des formes avec des coefficients qui sont des fonctions généralisées, c'est-à-dire des courants. Les groupes obtenus $\mathrm{H}^n(\mathrm{X}, \mathbb{C})$ et $\mathrm{H}^{p,q}$ sont les mêmes. Si Z est un sous-espace analytique fermé de X, de codimension complexe p, Z est un cycle entier et définit par dualité de Poincaré une classe $\mathrm{cl}(\mathrm{Z})$ dans $\mathrm{H}^{2p}(\mathrm{X}, \mathbb{Z})$. Le courant d'intégration sur Z est une (p,p)-forme fermée à coefficients qui sont des fonctions généralisées, représentant l'image de $\mathrm{cl}(\mathrm{Z})$ dans $\mathrm{H}^{2p}(\mathrm{X}, \mathbb{C})$. La classe $\mathrm{cl}(\mathrm{Z})$ dans $\mathrm{H}^{2p}(\mathrm{X}, \mathbb{Z})$ est donc de type (p,p), au sens où son image dans $\mathrm{H}^{2p}(\mathrm{X}, \mathbb{C})$ l'est. Les (p,p)-classes rationnelles sont appelées « classes de Hodge ». Elles forment le groupe

$$\mathrm{H}^{2p}(\mathrm{X}, \mathbb{Q}) \cap \mathrm{H}^{p,p}(\mathrm{X}) = \mathrm{H}^{2p}(\mathrm{X}, \mathbb{Q}) \cap \mathrm{F}^p \subset \mathrm{H}^{2p}(\mathrm{X}, \mathbb{C}).$$

Dans [6], Hodge a posé la

Conjecture de Hodge. *Sur une variété algébrique projective non singulière sur* $\mathbb{C}$*, toute classe de Hodge est une combinaison linéaire rationnelle de classes* $\mathrm{cl}(\mathrm{Z})$ *de cycles algébriques.*

Remarques

(i) Par le théorème de Chow, sur une variété projective complexe, les fermés algébriques sont identiques aux sous-espaces analytiques fermés.

(ii) Sur une variété projective non singulière X sur $\mathbb{C}$, le groupe des combinaisons linéaires entières des classes $\mathrm{cl}(\mathrm{Z})$ de cycles algébriques coïncide avec le groupe des combinaisons linéaires entières des produits de classes de Chern de fibrés vectoriels algébriques (ou, de manière équivalente d'après GAGA, analytiques). Pour exprimer $\mathrm{cl}(\mathrm{Z})$ en termes de classes de Chern, on résout le faisceau structurel $\mathcal{O}_\mathrm{Z}$ par un complexe fini de fibrés vectoriels. Le fait que les classes de Chern soient des cycles algébriques vient essentiellement du fait

que les fibrés vectoriels possèdent de nombreuses sections méromorphes.

(iii) Un cas particulier de (ii) est que les combinaisons linéaires entières des classes de diviseurs (= cycles de codimension 1) sont simplement les premières classes de Chern des fibrés en droites. Si $Z^+ - Z^-$ est le diviseur d'une section méromorphe de $\mathcal{L}$, alors $c_1(\mathcal{L}) = \mathrm{cl}\,(Z^+) - \mathrm{cl}\,(Z^-)$. C'est le point de départ de la démonstration donnée par Kodaira et Spencer [7] de la conjecture de Hodge pour H^2 : une classe $c \in H^2(X, \mathbb{Z})$ de type $(1,1)$ a pour image 0 dans le quotient $H^{0,2} = H^2(X, \mathcal{O})$ de $H^2(X, \mathbb{C})$, et la suite exacte longue de cohomologie définie par la suite exacte exponentielle

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathcal{O} \xrightarrow{\exp(2\pi i \quad)} \mathcal{O}^* \longrightarrow 0$$

montre que c est la première classe de Chern d'un fibré en droites.

(iv) La relation entre les cycles algébriques et les fibrés vectoriels algébriques est également à la base du théorème d'Atiyah et Hirzebruch [2] selon lequel la conjecture de Hodge ne peut pas s'appliquer intégralement. Dans la suite spectrale d'Atiyah-Hirzebruch de la cohomologie à la K-théorie topologique,

$$E_2^{pq} = H^p\left(X, K^q\left(P^t\right)\right) \Longrightarrow K^{p+q}(X)$$

la filtration de $K^n(X)$ qui en résulte est par le

$$F^pK^n(X) = \mathrm{Ker}\,(K^n(X) \to K^n((p-1)\text{-squelette, dans toute triangulation}))\ .$$

De manière équivalente, une classe c est dans F^p si pour un sous-espace topologique Y de codimension p, c est l'image d'une classe $\tilde{c}$ dont le support est dans Y. Si Z est un cycle algébrique de codimension p, une résolution de $\mathcal{O}_Z$ définit une classe de K-théorie à support dans Z : $c_Z \in K^0(X, X - Z)$. Son image dans $F^pK^0(X)$ est la classe de Z dans $H^{2p}(X, \mathbb{Z})$. Cette dernière est donc dans le noyau des différentielles successives d_r de la suite spectrale.

On ne connaît pas de contre-exemple à l'affirmation selon laquelle les (p,p)-classes entières tuées par tous les d_r sont des combinaisons linéaires entières des classes cl(Z). On n'a aucune idée des classes qui devraient être effectives, c'est-à-dire de la forme cl(Z), plutôt qu'une différence de telles classes.

Sur une variété de Stein X, tout fibré vectoriel complexe topologique peut être muni une structure holomorphe et, au moins pour X du type

d'homotopie d'un CW-complexe fini, il s'ensuit que toute classe dans $\mathrm{H}^{2p}(\mathrm{X},\mathbb{Z})$ qui est dans le noyau de tous les d_r est une combinaison linéaire à coefficients dans $\mathbb{Z}$ de classes de cycles analytiques.

(v) L'hypothèse de la conjecture de Hodge selon laquelle X doit être algébrique ne peut pas être affaiblie en X simplement kählérienne. Voir l'appendice de Zucker à [11] pour des contre-exemples où X est un tore complexe.

(vi) Lorsque Hodge a formulé sa conjecture, il n'avait pas réalisé qu'elle ne pouvait être valable que rationnellement (c'est-à-dire après une tensorisation avec $\mathbb{Q}$). Il a également proposé une autre conjecture, qui caractérise le sous-espace de $\mathrm{H}^n(\mathrm{X},\mathbb{Z})$ engendré par les images des classes de cohomologie ayant un support dans un sous-espace analytique fermé approprié de codimension complexe k. Grothendieck a observé que cette autre conjecture est trivialement fausse et en a donné une version corrigée dans [5].

La jacobienne intermédiaire

La classe de cohomologie d'un cycle algébrique Z de codimension p se relève naturellement dans un groupe $\mathrm{J}_p(\mathrm{X})$, extension du groupe des classes de type (p,p) dans $\mathrm{H}^{2p}(\mathrm{X},\mathbb{Z})$ par la jacobienne intermédiaire

$$\mathrm{J}_p(\mathrm{X})^0 := \mathrm{H}^{2p-1}(\mathrm{X},\mathbb{Z})\backslash\mathrm{H}^{2p-1}(\mathrm{X},\mathbb{C})/\mathrm{F}^p.$$

Cela signifie que la classe admet une représentation entière (en cohomologie singulière) ainsi qu'une description analytique comme (p,q)-courant fermé donnant une classe d'hypercohomologie dans $\mathbb{H}^{2p}$ du sous-complexe $\mathrm{F}^p\Omega^*_{\mathrm{hol}} := (0 \to \cdots \to 0 \to \Omega^p_{\mathrm{hol}} \to \cdots)$ du complexe de de Rham holomorphe, avec une compréhension au niveau des cocycles de la raison pour laquelle les deux coïncident dans $\mathrm{H}^{2p}(\mathrm{X},\mathbb{C})$. « Compréhension » signifie une cochaîne dans un complexe qui calcule $\mathrm{H}^*(\mathrm{X},\mathbb{C})$, dont le cobord est la différence entre les cocycles provenant des constructions entières, respectivement analytiques. En effet, $\mathrm{J}_p(\mathrm{X})$ est l'hypercohomologie $\mathbb{H}^{2p}$ du noyau homotopique de l'application différence $\mathbb{Z} \oplus \mathrm{F}^p\Omega^*_{\mathrm{hol}} \to \Omega^*$.

En général, en utilisant le fait que tous les cycles algébriques sur X s'inscrivent dans un nombre dénombrable de familles algébriques, on vérifie que le sous-groupe $\mathrm{A}_p(\mathrm{X})$ de $\mathrm{J}_p(\mathrm{X})$ engendré par les cycles algébriques est extension d'un groupe dénombrable par sa composante

connexe $A^0_p(X)$, et que pour toute sous-structure de Hodge H_{alg} de type $\{(p-1,p),(p,p-1)\}$ de $H^{2p-1}(X)$, $A^0_p(X)$ est $H_{alg_{\mathbb{Z}}} \backslash H_{alg_{\mathbb{C}}}/F^p$. On entend par « sous-structure de Hodge » le sous-groupe du réseau dont la complexification est la somme de ses intersections avec les $H^{a,b}$. La conjecture de Hodge (appliquée au produit de X et d'une variété abélienne convenable) prédit que H_{alg} est la plus grande sous-structure de Hodge de $H^{2p-1}(X)$ de type $\{(p-1,p)(p,p-1)\}$.

Aucune conjecture ne permet de prédire de quel sous-groupe de $J_p(X)$ est le groupe $A_p(X)$. On connaît des cas où $A_p(X)/A^0_p(X)$ est de rang infini (voir par exemple l'article [9] et les références qu'il contient). Ceci a rendu généralement inapplicables les méthodes introduites par Griffiths (voir par exemple Zucker [11]) pour démontrer la conjecture de Hodge par récurrence sur la dimension de X, en utilisant un pinceau de Lefschetz de sections hyperplanes de X. En effet, la méthode requiert non seulement la conjecture de Hodge pour les sections hyperplanes H, mais aussi que tout $J_p(H)$ provienne de cycles algébriques.

La détection des classes de Hodge

Soit $(X_s)_{s\in S}$ une famille algébrique de variétés algébriques projectives non singulières : les fibres d'une application projective et lisse $f : X \to S$. On suppose qu'elle est définie sur la clôture algébrique $\overline{\mathbb{Q}}$ de $\mathbb{Q}$ dans $\mathbb{C}$. On ne connaît pas d'algorithme pour décider si une classe de cohomologie entière donnée d'une fibre typique X_0 est quelque part sur S de type (p,p). La conjecture de Hodge implique que le lieu où cela se produit est une union dénombrable de sous-variétés algébriques de S (c'est connu, voir [4]) et défini sur $\overline{\mathbb{Q}}$ (inconnu).

La conjecture de Hodge n'est pas encore établie, même dans les exemples suivants.

Exemple 1. Pour X de dimension complexe N, la diagonale Δ de $X \times X$ est un cycle algébrique de codimension N. La décomposition de Hodge étant compatible avec Künneth, les composantes de Künneth $cl(\Delta)_{a,b} \in H^a(X) \otimes H^b(X) \subset H^{2N}(X \times X)$ $(a+b=2N)$ de $cl(\Delta)$ sont des classes de Hodge.

Exemple 2. Si $\eta \in H^2(X,\mathbb{Z})$ est la classe de cohomologie d'une section hyperplane de X, le *cup*-produit itéré $\eta^p : H^{N-p}(X,\mathbb{C}) \to H^{N+p}(X,\mathbb{C})$ est un isomorphisme (théorème de Lefschetz difficile démontré par Hodge, voir [10, IV.6]). Soit $\mathfrak{z} \in H^{N-p}(X,\mathbb{C})\otimes$

$\mathrm{H}^{\mathrm{N}-p}(\mathrm{X},\mathbb{C}) \subset \mathrm{H}^{2\mathrm{N}-2p}(\mathrm{X}\times\mathrm{X})$ la classe telle que l'isomorphisme inverse $(\eta^p)^{-1}$ soit $c \mapsto \mathrm{pr}_{1!}\,(\mathfrak{z} \cup \mathrm{pr}_2^*\, c)$. La classe $\mathfrak{z}$ est de Hodge.

Les motifs

Les variétés algébriques admettent une multitude de théories cohomologiques, reliées sur $\mathbb{C}$ par des isomorphismes de comparaison. Les structures résultantes sur $\mathrm{H}^*(\mathrm{X},\mathbb{Z})$ doivent être considérées comme analogues à la structure de Hodge. Exemples : si X est défini sur un sous-corps K de $\mathbb{C}$, avec une clôture algébrique $\bar{\mathrm{K}}$ dans $\mathbb{C}$, $\mathrm{Gal}(\bar{\mathrm{K}}/\mathrm{K})$ agit sur $\mathrm{H}^*(\mathrm{X},\mathbb{Z}) \otimes \mathbb{Z}_\ell$ et $\mathrm{H}^*(\mathrm{X},\mathbb{C}) = \mathrm{H}^*(\mathrm{X},\mathbb{Z}) \otimes \mathbb{C}$ possède une K-structure naturelle $\mathrm{H}_{\mathrm{DR}}(\mathrm{X}$ sur $\mathrm{K})$, compatible avec la filtration de Hodge. Ces théories cohomologiques donnent lieu à des conjectures parallèles à la conjecture de Hodge, déterminant l'espace vectoriel engendré par des classes de cycles algébriques. Exemple : la conjecture de Tate [8]. Ces conjectures sont ouvertes même pour H^2.

La théorie des motifs de Grothendieck vise à comprendre le parallélisme entre ces théories cohomologiques. Les progrès sont bloqués par le manque de méthodes pour construire des cycles algébriques intéressants. Si les cycles des exemples 1 et 2 du §4 étaient algébriques, les motifs de Grothendieck sur $\mathbb{C}$ formeraient une catégorie abélienne semi-simple avec un produit tensoriel, et seraient la catégorie des représentations d'un certain schéma en groupes pro-réductif. Si l'on suppose l'algébricité de ces cycles, la conjecture de Hodge complète est équivalente à un foncteur naturel de la catégorie des motifs vers la catégorie des structures de Hodge qui est pleinement fidèle.

Substituts et formes affaiblies

En désespoir de cause, des efforts ont été faits pour trouver des substituts à la conjecture de Hodge. Sur les variétés abéliennes, les classes de Hodge partagent au moins de nombreuses propriétés avec les classes de cohomologie des cycles algébriques : elles sont « absolument de Hodge » [3], voire « motivées » [1]. Cela suffit pour certaines applications, par exemple pour la démonstration des relations algébriques entre les périodes et les quasi-périodes des variétés abéliennes prédites par la conjecture de Hodge [3], mais ne permet pas de réduction modulo p. Les corollaires suivants de la conjecture de Hodge seraient particulièrement intéressants. Soit A une variété abélienne sur la clôture algébrique $\mathbb{F}$

d'un corps fini $\mathbb{F}_q$. On relève A de deux manières différentes à la caractéristique 0, à des variétés abéliennes complexes A_1 et A_2 définies sur $\overline{\mathbb{Q}}$. On choisit des classes de Hodge $\mathfrak{z}_1$ et $\mathfrak{z}_2$ sur A_1 et A_2, de dimension complémentaire. En interprétant $\mathfrak{z}_1$ et $\mathfrak{z}_2$ comme des classes de cohomologie ℓ-adiques, on peut définir le nombre d'intersection κ de la réduction de $\mathfrak{z}_1$ et $\mathfrak{z}_2$ sur $\mathbb{F}$. Est-ce que κ est un nombre rationnel ? Si $\mathfrak{z}_1$ et $\mathfrak{z}_2$ étaient $\mathrm{cl}(Z_1)$ et $\mathrm{cl}(Z_2)$, on pourrait choisir de définir Z_1 et Z_2 sur $\overline{\mathbb{Q}}$ et κ serait le nombre d'intersection des réductions de Z_1 et Z_2. Même question pour le nombre d'intersection de la réduction de $\mathfrak{z}_1$ sur $\mathbb{F}$ avec la classe d'un cycle algébrique sur A.

Bibliographie

[1] André (Y.), « Pour une théorie inconditionnelle des motifs », *Publ. math. IHES*, n° 83, 1996, p. 5-49.

[2] Atiyah (M. F.) et Hirzebruch (F.), « Analytic cycles on complex manifolds », *Topology*, n° 1, 1962, p. 25-45.

[3] Deligne (P.), « Cycles de Hodge absolus et périodes des intégrales des variétés abéliennes » (rédigé par J.-L. Brylinski), *Mémoires SMF*, n° 2, 1980, p. 23-33.

[4] Deligne (P.), Cattani (E.) et Kaplan (A.), « On the locus of Hodge classes », *J. AMS*, n° 8, 1995, p. 483-505.

[5] Grothendieck (A.), « Hodge's general conjecture is false for trivial reasons », *Topology*, n° 8, 1969, p. 299-303.

[6] Hodge (W. V. D.), « The topological invariants of algebraic varieties », dans *Proceedings ICM 1950*, Providence, American Mathematical Society, 1952, p. 181-192.

[7] Kodaira (K.) et Spencer (D. C.), « Divisor classes on algebraic varieties », *Proc. Nat. Acad. Sci.*, n° 39, 1953, p. 872-877.

[8] Tate (J.), « Algebraic cycles and poles of zeta functions », dans *Arithmetic Algebraic Geometry*, New York, Harper and Row, 1965, p. 93-110.

[9] Voisin (C.), « The Griffiths group of a general Calabi-Yau threefold is not finitely generated », *Duke Math. J.*, n° 102, 2000, p. 151-186.

[10] Weil (A.), *Introduction à l'étude des variétés kählériennes*, Paris, Hermann, 1958.

[11] Zucker (S.), « The Hodge conjecture for cubic fourfolds », *Comp. Math.*, n° 34, 1977, p. 199-209.

Existence et régularité pour l'équation de Navier-Stokes (par Charles L. Fefferman)

Les équations d'Euler et de Navier-Stokes décrivent le mouvement d'un fluide dans $\mathbb{R}^n$ ($n = 2$ ou 3). Ces équations ont comme inconnues le vecteur vitesse $u(x,t) = (u_i(x,t))_{1\leqslant i\leqslant n} \in \mathbb{R}^n$ et la pression $p(x,t) \in \mathbb{R}$, définis pour toute position $x \in \mathbb{R}^n$ et tout temps $t \geqslant 0$. On se limite ici aux fluides incompressibles qui remplissent $\mathbb{R}^n$. Les équations de Navier-Stokes sont alors données par

$$\frac{\partial}{\partial t}u_i + \sum_{j=1}^{n} u_j \frac{\partial u_i}{\partial x_j} = \nu\Delta u_i - \frac{\partial p}{\partial x_i} + f_i(x,t) \quad (x \in \mathbb{R}^n, t \geqslant 0)\,, \tag{1}$$

$$\operatorname{div} u = \sum_{i=1}^{n} \frac{\partial u_i}{\partial x_i} = 0 \quad (x \in \mathbb{R}^n, t \geqslant 0) \tag{2}$$

avec des conditions initiales

$$u(x,0) = u^\circ(x) \quad (x \in \mathbb{R}^n)\,. \tag{3}$$

Ici, $u^\circ(x)$ est un champ de vecteurs à divergence nulle de classe C^∞ sur $\mathbb{R}^n$, $f_i(x,t)$ sont les composantes d'une force donnée, appliquée de l'extérieur (par exemple la gravité), ν est un coefficient strictement positif (la viscosité), et $\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$ est le laplacien pour les variables d'espace. Les équations d'Euler sont les équations (1), (2) et (3) avec ν égal à zéro.

L'équation (1) est simplement la loi de Newton $f = ma$ pour un élément fluide soumis à la force externe $f = (f_i(x,t))_{1\leqslant i\leqslant n}$ et aux forces qui résultent de la pression et du frottement. L'équation (2) indique que le fluide est incompressible. Pour que les solutions soient physiquement raisonnables, on veut s'assurer que $u(x,t)$ ne devienne pas très grand lorsque $|x| \to \infty$. Par conséquent, on limitera notre attention aux forces f et aux conditions initiales u° qui satisfont, pour tout α et K,

$$|\partial_x^\alpha u^\circ(x)| \leqslant \mathrm{C}_{\alpha\mathrm{K}}(1+|x|)^{-\mathrm{K}} \text{ sur } \mathbb{R}^n, \tag{4}$$

et, pour tout α, m et K,

$$|\partial_x^\alpha \partial_t^m f(x,t)| \leqslant \mathrm{C}_{\alpha m\mathrm{K}}(1+|x|+t)^{-\mathrm{K}} \text{ sur } \mathbb{R}^n \times [0,\infty[. \tag{5}$$

On accepte une solution de (1)-(2)-(3) comme physiquement raisonnable seulement si elle vérifie

$$p, u \in \mathrm{C}^\infty\left(\mathbb{R}^n \times [0, \infty[\right) \tag{6}$$

et

$$\int_{\mathbb{R}^n} |u(x,t)|^2 dx < \mathrm{C} \quad \text{pour tout } t \geqslant 0 \quad \text{(énergie bornée)}. \tag{7}$$

Alternativement, pour exclure les problèmes à l'infini, on peut rechercher des solutions spatialement périodiques de (1)-(2)-(3). Ainsi, on suppose que $u^\circ(x)$ et $f(x,t)$ vérifient

$$u^\circ\left(x + e_j\right) = u^\circ(x), \quad f\left(x + e_j, t\right) = f(x,t), \tag{8}$$

pour tout $1 \leqslant j \leqslant n$, où e_j est le j-ième vecteur unité de $\mathbb{R}^n$.

Au lieu de (4) et (5), on suppose que u° est une fonction régulière et que, pour tout α, m et K,

$$|\partial_x^\alpha \partial_t^m f(x,t)| \leqslant \mathrm{C}_{\alpha m \mathrm{K}} (1 + |t|)^{-\mathrm{K}} \text{ sur } \mathbb{R}^3 \times [0, \infty[, \tag{9}$$

On accepte alors une solution de (1)-(2)-(3) comme physiquement raisonnable si elle vérifie

$$u(x,t) = u\left(x + e_j, t\right) \text{ sur } \mathbb{R}^3 \times [0, \infty[\text{ pour tout } 1 \leqslant j \leqslant n \tag{10}$$

et

$$p, u \in \mathrm{C}^\infty\left(\mathbb{R}^n \times [0, \infty[\right). \tag{11}$$

Un problème fondamental en analyse est de décider si de telles solutions régulières et physiquement raisonnables existent pour les équations de Navier-Stokes. Pour donner une marge de manœuvre raisonnable aux candidats tout en conservant le cœur du problème, on demande une démonstration de l'une des quatre affirmations suivantes :

(A) **Existence et régularité des solutions de Navier-Stokes sur $\mathbb{R}^3$.** Soient $\nu > 0$ et $n = 3$. Soit $u^\circ(x)$ un champ de vecteurs régulier, à divergence nulle, qui vérifie (4). Supposons que la fonction $f(x,t)$ soit identiquement nulle. Il existe alors des fonctions régulières $p(x,t)$ et $u_i(x,t)$ sur $\mathbb{R}^3 \times [0, \infty[$ qui vérifient (1)-(2)-(3), (6) et (7).

(B) **Existence et régularité des solutions de Navier-Stokes sur $\mathbb{R}^3/\mathbb{Z}^3$.** Soient $\nu > 0$ et $n = 3$. Soit $u^\circ(x)$ un champ de vecteurs régulier, à divergence nulle, qui vérifie (8). Supposons que la fonction $f(x,t)$ soit identiquement nulle. Il existe alors des fonctions régulières $p(x,t)$ et $u_i(x,t)$ sur $\mathbb{R}^3 \times [0,\infty[$ qui vérifient (1)-(2)-(3), (10) et (11).

(C) **Explosion des solutions de Navier-Stokes sur $\mathbb{R}^3$.** Soient $\nu > 0$ et $n = 3$. Il existe alors un champ de vecteurs régulier et à divergence nulle $u^\circ(x)$ sur $\mathbb{R}^3$ et un champ de vecteurs $f(x,t)$ régulier sur $\mathbb{R}^3 \times [0,\infty[$ qui vérifient (4) et (5), pour lesquels il n'existe pas de solutions (p,u) de (1)-(2)-(3), (6) et (7) sur $\mathbb{R}^3 \times [0,\infty[$.

(D) **Explosion des solutions de Navier-Stokes sur $\mathbb{R}^3/\mathbb{Z}^3$.** Soient $\nu > 0$ et $n = 3$. Il existe alors un champ de vecteurs régulier et à divergence nulle $u^\circ(x)$ sur $\mathbb{R}^3$ et un champ de vecteurs $f(x,t)$ régulier sur $\mathbb{R}^3 \times [0,\infty[$ qui vérifient (8) et (9), pour lesquels il n'existe pas de solutions (p,u) de (1)-(2)-(3), (10) et (11) sur $\mathbb{R}^3 \times [0,\infty[$.

Ces problèmes sont également ouverts et très importants pour les équations d'Euler ($\nu = 0$), bien que l'équation d'Euler ne figure pas sur la liste des problèmes du prix du millénaire de l'institut Clay.

Rappelons les principaux résultats partiels connus concernant les équations d'Euler et de Navier-Stokes, et concluons par quelques remarques sur l'importance de la question.

En dimension 2, les analogues des affirmations (A) et (B) sont connus depuis longtemps (Ladyjenskaïa [4]), y compris pour le cas plus difficile des équations d'Euler. Cela ne donne aucune indication sur le cas de la dimension 3, puisque les principales difficultés sont absentes en dimension 2. En dimension 3, on sait que (A) et (B) sont valables à condition que la vitesse initiale u° satisfasse une condition de petitesse. Pour des données initiales $u^\circ(x)$ qui ne sont pas supposées petites, on sait que (A) et (B) sont valables (également pour $\nu = 0$) si l'intervalle de temps $[0,\infty[$ est remplacé par un petit intervalle de temps $[0, \mathrm{T}[$, avec T qui dépend des données initiales. Pour une donnée initiale $u^\circ(x)$, le maximum admissible pour T est appelé le « temps d'explosion ». Soit (A) et (B) sont valables, soit il existe une fonction $u^\circ(x)$ régulière et à divergence nulle pour laquelle (1)-(2)-(3) ont une solution avec un temps d'explosion fini. Pour les équations de Navier-Stokes ($\nu > 0$), s'il existe une solution avec un temps d'explosion fini T, alors la vitesse

$(u_i(x,t))_{1\leqslant i\leqslant 3}$ devient non bornée au voisinage du temps d'explosion.

On sait que d'autres choses désagréables se produisent au temps d'explosion T, si $T<\infty$. Pour les équations d'Euler ($\nu=0$), s'il existe une solution (avec $f\equiv 0$, par exemple) avec un temps d'explosion fini T, alors la vorticité $\omega(x,t)=\mathrm{rot}_x\, u(x,t)$ vérifie

$$\int_0^T \left\{ \sup_{x\in\mathbb{R}^3} |\omega(x,t)| \right\} dt = \infty \quad \text{(Beale-Kato-Majda)}$$

de sorte que la vorticité explose rapidement.

De nombreux calculs numériques semblent présenter une explosion pour les solutions des équations d'Euler, mais l'extrême instabilité numérique des équations rend très difficile d'en tirer des conclusions fiables.

Les résultats ci-dessus sont très bien présentés dans le livre de Bertozzi et Majda [1].

Depuis Leray [5], des progrès importants ont été réalisés dans la compréhension des solutions faibles des équations de Navier-Stokes. Pour parvenir à l'idée d'une solution faible d'une équation aux dérivées partielles (EDP), on intègre l'équation multipliée une fonction test, puis on procède (formellement) à une intégration par parties pour que les dérivées tombent sur la fonction test. Si par exemple (1) et (2) sont vérifiés, alors, pour tout champ de vecteurs régulier $\theta(x,t)=(\theta_i(x,t))_{1\leqslant i\leqslant n}$ à support compact dans $\mathbb{R}^3\times]0,\infty[$, une intégration formelle par parties donne

$$-\iint_{\mathbb{R}^3\times\mathbb{R}} u\cdot\frac{\partial\theta}{\partial t}dxdt-\sum_{ij}\iint_{\mathbb{R}^3\times\mathbb{R}} u_iu_j\frac{\partial\theta_i}{\partial x_j}dxdt \tag{12}$$

$$=\nu\iint_{\mathbb{R}^3\times\mathbb{R}} u\cdot\Delta\theta dxdt+\iint_{\mathbb{R}^3\times\mathbb{R}} f\cdot\theta dxdt+\iint_{\mathbb{R}^3\times\mathbb{R}} p\cdot(\mathrm{div}\,\theta)dxdt.$$

Notons que (12) a un sens pour $u\in \mathrm{L}^2$, $f\in \mathrm{L}^1$ et $p\in \mathrm{L}^1$, alors que (1) n'a de sens que si $u(x,t)$ est deux fois différentiable en x. De même, si $\varphi(x,t)$ est une fonction régulière, à support compact dans $\mathbb{R}^3\times]0,\infty[$, alors une intégration formelle par parties et (2) impliquent

$$\iint_{\mathbb{R}^3\times\mathbb{R}} u\cdot\nabla_x\varphi dxdt=0. \tag{13}$$

Une solution de (12)-(13) est appelée une solution faible des équations de Navier-Stokes. Une vieille idée en analyse est de démontrer

l'existence et la régularité des solutions d'une EDP en construisant d'abord une solution faible, puis en montrant que toute solution faible est régulière. On a essayé ce programme pour les équations de Navier-Stokes avec un succès partiel. Leray a montré dans [5] que les équations de Navier-Stokes (1)-(2)-(3) en dimension 3 ont toujours une solution faible (p, u) avec des propriétés de croissance appropriées. L'unicité des solutions faibles de l'équation de Navier-Stokes n'est pas connue. Pour l'équation d'Euler, l'unicité des solutions faibles est étonnamment fausse. Scheffer [8] et plus tard Schnirelman [9] ont présenté des solutions faibles des équations d'Euler sur $\mathbb{R}^2 \times \mathbb{R}$ à support compact en espace et en temps. Cela correspond à un fluide qui part du repos à l'instant $t = 0$, commence à se mouvoir à l'instant $t = 1$ sans stimulation extérieure, et revient au repos à l'instant $t = 2$, son mouvement étant toujours confiné dans une boule $\mathrm{B} \subset \mathbb{R}^3$.

Scheffer [7] a appliqué des idées de la théorie géométrique de la mesure pour démontrer un théorème de régularité partielle pour des solutions faibles appropriées des équations de Navier-Stokes. Caffarelli, Kohn et Nirenberg [2] ont amélioré les résultats de Scheffer, tandis que F.-H. Lin [6] a simplifié la démonstration des résultats de Caffarelli, Kohn et Nirenberg [2]. Le théorème de régularité partielle de [2] et [6] concerne un analogue parabolique de la dimension de Hausdorff de l'ensemble singulier d'une solution faible appropriée de Navier-Stokes. Ici, l'ensemble singulier d'une solution faible u est constitué de tous les points $(x^\circ, t^\circ) \in \mathbb{R}^3 \times \mathbb{R}$ tels que la solution u soit non bornée dans tout voisinage de (x°, t°). [Si la force f est régulière et si (x°, t°) n'appartient pas à l'ensemble singulier, alors il n'est pas difficile de montrer que u peut être corrigé sur un ensemble de mesure nulle pour devenir régulier dans un voisinage de (x°, t°).]

Pour définir l'analogue parabolique de la dimension de Hausdorff, on utilise des cylindres paraboliques $\mathrm{Q}_r = \mathrm{B}_r \times \mathrm{I}_r \subset \mathbb{R}^3 \times \mathbb{R}$, où $\mathrm{B}_r \subset \mathbb{R}^3$ est une boule de rayon r et où $\mathrm{I}_r \subset \mathbb{R}$ est un intervalle de longueur r^2. Étant donné $\mathrm{E} \subset \mathbb{R}^3 \times \mathbb{R}$ et $\delta > 0$, on pose

$$\mathcal{P}_{\mathrm{K},\delta}(\mathrm{E}) = \inf \left\{ \sum_{i=1}^{\infty} r_i^{\mathrm{K}} : \mathrm{Q}_{r_1}, \mathrm{Q}_{r_2}, \cdots \text{ recouvrent E, avec } r_i < \delta \right\}$$

et l'on définit

$$\mathcal{P}_{\mathrm{K}}(\mathrm{E}) = \lim_{\delta \to 0+} \mathcal{P}_{\mathrm{K},\delta}(\mathrm{E})$$

Les principaux résultats de [2] et [6] peuvent être énoncés *grosso*

modo comme suit.

THÉORÈME. *(A) Soit u une solution faible des équations de Navier-Stokes qui vérifie des conditions de croissance appropriées. Soit* E *l'ensemble singulier de u. Alors $\mathcal{P}_1(\mathrm{E}) = 0$.*

(B) Étant donné un champ de vecteurs à divergence nulle $u^\circ(x)$ et une force $f(x,t)$ qui vérifie (4) et (5), il existe une solution faible des équations de Navier-Stokes (1)-(2)-(3) qui vérifie les conditions de croissance de (A).

En particulier, l'ensemble singulier de u ne peut pas contenir une courbe spatio-temporelle de la forme $\big\{(x,t) \in \mathbb{R}^3 \times \mathbb{R} : x = \phi(t)\big\}$. Il s'agit du meilleur théorème de régularité partielle connu à ce jour pour l'équation de Navier-Stokes. Il semble très difficile d'aller plus loin.

Permettez-moi de conclure par quelques mots sur l'importance des problèmes posés ici. Les fluides sont importants et difficiles à comprendre. Il existe de nombreux problèmes et conjectures fascinants sur le comportement des solutions des équations d'Euler et de Navier-Stokes (voir par exemple Bertozzi-Majda [1] ou Constantin [3]). Comme on ne sait même pas si ces solutions existent, notre compréhension est à un niveau très primitif. Les méthodes standard des EDP semblent inadéquates pour résoudre le problème. Au lieu de cela, on a probablement besoin d'idées nouvelles et profondes.

Bibliographie

[1] BERTOZZI (A.) et MAJDA (A.), *Vorticity and Incompressible Flows*, Cambridge University Press, 2002.

[2] CAFFARELLI (L.), KOHN (R.) et NIRENBERG (L.), « Partial regularity of suitable weak solutions of the Navier-Stokes equations », *Comm. Pure Appl. Math.*, n° 35, 1982, p. 771-831.

[3] CONSTANTIN (P.), « Some open problems and research directions in the mathematical study of fluid dynamics », dans *Mathematics Unlimited – 2001 and Beyond*, Berlin, Springer, 2001, p. 353-360.

[4] LADYZHENSKAYA (O.), *The Mathematical Theory of Viscous Incompressible Flows*, 2e éd., New York, Gordon and Breach, 1969.

[5] LERAY (J.), « Sur le mouvement d'un liquide visqueux emplissent l'espace », *Acta Math.*, n° 63, 1934, p. 193-248.

[6] LIN (F.-H.), « A new proof of the Caffarelli-Kohn-Nirenberg theorem », *Comm. Pure. Appl. Math.*, n° 51, 1998, p. 241-257.

[7] Scheffer (V.), « Turbulence and Hausdorff dimension », dans *Turbulence and the Navier-Stokes Equations*, Berlin, Springer, 1976, p. 94-112.

[8] Scheffer (V.), « An inviscid flow with compact support in spacetime », *J. Geom. Anal.*, n° 3, 1993, p. 343-401.

[9] Shnirelman (A.), « On the nonuniqueness of weak solutions of the Euler equation », *Comm. Pure Appl. Math.*, n° 50, 1997, p. 1260-1286.

Erratum

La condition supplémentaire $p\,(x + e_j, t) = p(x, t)$ devrait figurer dans l'équation (8).

La conjecture de Poincaré (par John Milnor)

Introduction

La topologie des variétés à deux dimensions ou des surfaces était bien comprise au début du XX^{e} siècle. En fait, il existe une liste simple de toutes les surfaces lisses, compactes, sans bord et orientables possibles. Toute surface de ce type a un genre bien défini $g \geqslant 0$, qui peut être décrit intuitivement comme le nombre de trous. Deux surfaces de ce type sont homéomorphes si et seulement si elles ont le même genre [1]. La question correspondante en dimensions supérieures est beaucoup plus difficile.

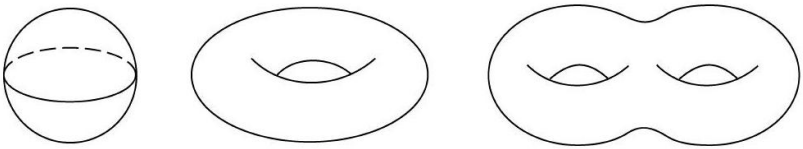

Fig. 1 – *Esquisses de surfaces lisses de genre 0, 1 et 2.*

Henri Poincaré fut peut-être le premier à essayer de faire une étude similaire des variétés tridimensionnelles. L'exemple le plus élémentaire d'une telle variété est la sphère unité à trois dimensions, c'est-à-dire le lieu de tous les points (x, y, z, w) dans l'espace euclidien à quatre dimensions qui ont une distance à l'origine exactement égale à 1 : $x^2 + y^2 + z^2 + w^2 = 1$. Il remarqua qu'une caractéristique distinctive de la sphère bidimensionnelle est que toute courbe fermée simple de la sphère peut être déformée continûment en un point sans quitter celle-ci. En 1904, il posa la même question en dimension 3. Dans un langage plus moderne, elle peut être formulée comme suit [2].

1. Pour les définitions et autres documents de référence, voir par exemple [21] ou [29], ainsi que [48].

2. Voir [36, p. 498 et 370]. Pour Poincaré, les variétés étaient toujours lisses ou polyédrales, de sorte que son terme « homéomorphisme » désignait un homéomorphisme lisse ou linéaire par morceaux.

Question. *Si une variété tridimensionnelle* M^3 *compacte sans bord a la propriété que toute courbe fermée simple à l'intérieur de la variété peut être déformée continûment en un point, s'ensuit-il que* M^3 *est homéomorphe à la sphère* S^3 *?*

Il ajouta avec beaucoup de clairvoyance : « mais cette question nous entraînerait trop loin ». Depuis lors, l'hypothèse selon laquelle toute variété de dimension 3 fermée simplement connexe est homéomorphe à la sphère de dimension 3 est connue sous le nom de « conjecture de Poincaré ». Elle a inspiré les topologues depuis lors. Les tentatives pour la démontrer ou l'infirmer ont conduit à de nombreuses avancées dans notre compréhension de la topologie des variétés.

Les premiers faux pas

Dès le début, la nature apparemment simple de cet énoncé a conduit les mathématiciens à des excès. Quatre ans plus tôt, en 1900, Poincaré lui-même fut le premier à se tromper en énonçant un théorème faux que l'on peut formuler ainsi.

Théorème faux. *Toute variété polyédrale compacte ayant l'homologie d'une sphère de dimension n est en fait homéomorphe à la sphère de dimension n.*

Mais son article de 1904 fournit un magnifique contre-exemple à cette affirmation basé sur le concept de groupe fondamental qu'il avait introduit plus tôt (voir [36, p. 189-192 et 193-288]). Cet exemple peut être décrit géométriquement de la façon suivante. Considérons tous les icosaèdres réguliers possibles inscrits dans la sphère unité à deux dimensions. Pour spécifier un icosaèdre particulier dans cette famille, on doit fournir trois paramètres. Par exemple, deux paramètres sont nécessaires pour spécifier un seul sommet sur la sphère, puis un autre paramètre pour spécifier la direction vers un sommet voisin. Ainsi, chacun de ces icosaèdres peut être considéré comme un « point » unique dans la variété tridimensionnelle M^3 constituée de tous les icosaèdres de ce type [3]. Cette variété répond au critère préliminaire de Poincaré : par

3. En langage plus technique, ce M^3 peut être défini comme le quotient $SO(3)/I_{60}$, où $SO(3)$ est le groupe de toutes les rotations de l'espace euclidien tridimensionnel et où I_{60} est le sous-groupe constitué des soixante rotations qui envoient un icosaèdre standard dans lui-même. Le groupe fondamental $\pi_1(M^3)$, constitué de

les méthodes de la théorie de l'homologie, il ne peut être distingué de la sphère tridimensionnelle. Cependant, Poincaré put démontrer qu'il ne s'agit pas d'une sphère en construisant une courbe fermée simple qui ne peut pas être déformée en un point à l'intérieur de M^3. La construction n'est pas difficile : choisir un icosaèdre quelconque et considérer ses images par les rotations autour d'un sommet d'angles $0 \leqslant \theta \leqslant 2\pi/5$. Cela définit une courbe fermée simple dans M^3 qui ne peut pas être déformée en un point.

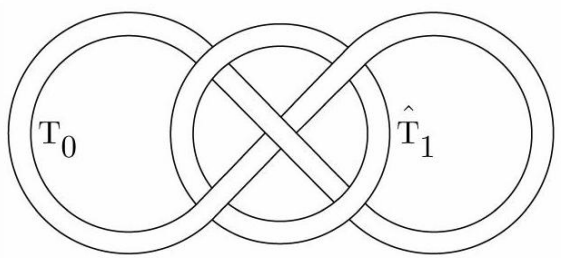

FIG. 2 – *L'entrelacs de Whitehead.*

Le théorème faux suivant fut énoncé par Henry Whitehead en 1934 [52]. Dans le cadre d'une prétendue démonstration de la conjecture de Poincaré, il affirma de façon nette que toute variété tridimensionnelle ouverte qui est contractile (c'est-à-dire qui peut être continûment déformée en un point) est homéomorphe à l'espace euclidien. En suivant les traces de Poincaré, il améliora ensuite considérablement notre compréhension de la topologie des variétés en découvrant un contre-exemple à son propre théorème. Son contre-exemple peut être brièvement décrit comme suit. Commençons avec deux tores solides disjoints T_0 et $\widehat{T}_1$ dans la sphère de dimension 3 qui sont plongés comme le montre la figure 2 : chacun d'eux pris individuellement n'est pas noué, mais les deux sont enlacés avec un enlacement nul. Puisque $\widehat{T}_1$ n'est pas noué, son complémentaire $T_1 = S^3 \backslash \operatorname{int}(\widehat{T}_1)$ est un autre tore solide non noué qui contient T_0. Choisissons un homéomorphisme h de la sphère de dimension 3 qui envoie T_0 sur ce plus grand tore solide T_1. On peut alors construire par récurrence des tores solides

$$T_0 \subset T_1 \subset T_2 \subset \cdots$$

toutes les classes d'homotopie des lacets dans M^3, est un groupe parfait d'ordre 120.

dans S^3 en posant $T_{j+1} = h(T_j)$. L'union $M^3 = \bigcup T_j$ de cette suite croissante est le contre-exemple de Whitehead requis, une variété contractile qui n'est pas homéomorphe à l'espace euclidien. Pour voir que $\pi_1(M^3) = 0$, remarquons que tout lacet dans T_0 peut être réduit à un point (après s'être peut-être croisé) à l'intérieur du tore solide plus grand T_1. Mais tout lacet dans M^3 doit être contenu dans un certain T_j, et peut donc être réduit à un point dans $T_{j+1} \subset M^3$. En revanche, M^3 n'est pas homéomorphe à l'espace euclidien tridimensionnel puisque, si $K \subset M^3$ est un sous-ensemble compact suffisamment grand pour contenir T_0, on peut démontrer que la différence $M^3 \backslash K$ n'est pas simplement connexe.

Depuis lors, de nombreuses fausses démonstrations de la conjecture de Poincaré ont été proposées, certaines d'entre elles reposant sur des erreurs plutôt subtiles et difficiles à détecter. Pour une présentation savoureuse de certains pièges de la topologie tridimensionnelle, voir [4].

Dimensions supérieures

La fin des années cinquante et le début des années soixante ont vu une avalanche de progrès avec la découverte que les variétés de dimensions supérieures sont en fait plus faciles à traiter que les variétés tridimensionnelles. L'une des raisons est que le groupe fondamental joue un rôle important dans toutes les dimensions même lorsqu'il est trivial et les relations entre les générateurs du groupe fondamental correspondent à des disques bidimensionnels appliqués dans la variété. En dimension 5 ou plus, ces disques peuvent être mis en position générale de manière à ce qu'ils soient disjoints les uns des autres, sans auto-intersections. Mais en dimension 3 ou 4, il n'est pas toujours possible d'éviter les intersections, ce qui entraîne de sérieuses difficultés.

Stephen Smale annonça une démonstration de la conjecture de Poincaré en dimensions supérieures en 1960 [41]. Il fut rapidement suivi par John Stallings, qui utilisait une méthode complètement différente [43], et par Andrew Wallace, qui avait travaillé selon des méthodes assez similaires à celles de Smale [51].

Je décrirai d'abord le résultat de Stallings, dont l'hypothèse est plus faible et la démonstration plus facile, mais dont la conclusion est également plus faible. Il supposait que la dimension était supérieure ou égale à sept, mais Christopher Zeeman a par la suite étendu l'argument aux dimensions 5 et 6 [54].

Théorème de Stallings-Zeeman. *Si* M^n *est un complexe simplicial fini de dimension* $n \geqslant 5$ *qui a le type d'homotopie*[4] *de la sphère* S^n *et qui est localement homéomorphe de façon linéaire par morceaux à l'espace euclidien* $\mathbb{R}^n$*, alors* M^n *est homéomorphe à* S^n *avec un homéomorphisme qui est linéaire par morceaux sauf en un seul point. En d'autres termes, le complémentaire* $\mathrm{M}^n \backslash$*(point) est homéomorphe de façon linéaire par morceaux à* $\mathbb{R}^n$*.*

La méthode de démonstration consiste à repousser toutes les difficultés vers un seul point. Il ne peut donc y avoir de contrôle à proximité de ce point.

La démonstration de Smale et la démonstration très proche donnée peu après par Wallace dépendaient plutôt de méthodes différentiables, en construisant une variété par récurrence en commençant par une boule de dimension n et en ajoutant successivement des anses. Ici, une k-anse peut être ajoutée à une variété à bord M^n en attachant d'abord une cellule de dimension k, en utilisant un homéomorphisme d'attachement de la sphère bordante de dimension $(k-1)$ dans le bord de M^n, puis en épaississant et en lissant les coins de façon à obtenir une variété à bord plus grande. La démonstration se fait en réarrangeant et en supprimant de telles anses (voir la présentation dans [24]).

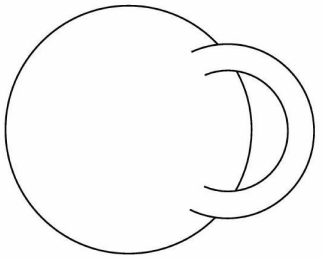

FIG. 3 – *Une boule tridimensionnelle à laquelle est attachée une 1-anse.*

Théorème de Smale. *Si* M^n *est une sphère d'homotopie différentiable de dimension* $n \geqslant 5$*, alors* M^n *est homéomorphe à* S^n*. En fait,* M^n *est difféomorphe à une variété obtenue en collant les bords de deux*

4. Pour vérifier qu'une variété M^n a le même type d'homotopie que la sphère S^n, il faut vérifier non seulement qu'elle est simplement connexe, $\pi_1(\mathrm{M}^n) = 0$, mais aussi qu'elle a la même homologie que la sphère. L'exemple du produit $\mathrm{S}^2 \times \mathrm{S}^2$ montre qu'il ne suffit pas de supposer que $\pi_1(\mathrm{M}^n) = 0$ lorsque $n > 3$.

n-boules fermées avec un difféomorphisme approprié.

Ceci a également été démontré par Wallace, au moins pour $n \geqslant 6$. Il faut noter que le cas de la dimension 5 est particulièrement difficile.

Le cas beaucoup plus difficile de la dimension 4 a dû attendre vingt ans jusqu'aux travaux de Michael Freedman [8]. Dans ce cas, les méthodes différentiables utilisées par Smale et Wallace et les méthodes linéaires par morceaux utilisées par Stallings et Zeeman ne fonctionnent pas du tout. Freedman a utilisé des méthodes totalement non différentiables, non seulement pour démontrer la conjecture de Poincaré en dimension 4 pour les variétés topologiques, mais aussi pour donner une classification complète de toutes les variétés topologiques de dimension 4 fermées et simplement connexes. Le groupe de cohomologie intégrale H^2 d'une telle variété est un groupe abélien libre. Freedman n'a eu besoin que de deux invariants : le *cup*-produit $\beta : \mathrm{H}^2 \otimes \mathrm{H}^2 \to \mathrm{H}^4 \cong \mathbb{Z}$ est une forme bilinéaire symétrique de déterminant ± 1, tandis que l'invariant de Kirby-Siebenmann κ est un entier modulo 2 qui s'annule si et seulement si la variété produit $\mathrm{M}^4 \times \mathbb{R}$ peut être munie d'une structure différentielle.

Théorème de Freedman. *Deux variétés de dimension 4 fermées simplement connexes sont homéomorphes si et seulement si elles ont la même forme bilinéaire β et le même invariant de Kirby-Siebenmann κ. Tout β peut être réalisé par une telle variété. Si $\beta(x \otimes x)$ est impair pour un certain $x \in \mathrm{H}^2$, alors n'importe quelle valeur de κ peut être réalisée aussi. Cependant, si $\beta(x \otimes x)$ est toujours pair, alors κ est déterminé par β, en étant congruent à un huitième de la signature de β.*

En particulier, si M^4 est une sphère d'homotopie, alors $\mathrm{H}^2 = 0$ et $\kappa = 0$, donc M^4 est homéomorphe à S^4. Il faut noter que les théories linéaires par morceaux ou différentiables en dimension 4 sont beaucoup plus difficiles. On ne sait pas si toute sphère de dimension 4 d'homotopie lisse est difféomorphe à S^4. On ne sait pas quelles variétés de dimension 4 avec $\kappa = 0$ possèdent effectivement des structures différentielles. On ne sait pas non plus quand cette structure est essentiellement unique. Les principaux résultats sur ces questions sont dus à Simon Donaldson [7]. Pour donner une idée des complications, Freedman a montré en utilisant les travaux de Donaldson que $\mathbb{R}^4$ admet un nombre non dénombrable de structures différentielles non équivalentes (voir [12]).

En dimension 3, les différences entre les théories topologiques, linéaires par morceaux et différentiables disparaissent (voir [18, 26, 28]). Mais les difficultés avec le groupe fondamental deviennent sérieuses.

La conjecture de géométrisation de Thurston

Dans le cas bidimensionnel, toute surface compacte lisse a une belle structure géométrique, celle d'une sphère dans le cas du genre 0, celle d'un tore plat dans le cas du genre 1, et celle d'une surface à courbure négative constante lorsque le genre est supérieur ou égal à 2. Une conjecture d'une grande portée formulée par William Thurston en 1983 affirme que quelque chose de similaire est vrai en dimension 3 [46]. Cette conjecture affirme que toute variété tridimensionnelle compacte et orientable peut être découpée le long de sphères de dimension 2 et de tores de manière à se décomposer en morceaux essentiellement uniques, dont chacun possède une structure géométrique simple. Il existe huit géométries tridimensionnelles possibles dans le programme de Thurston. Six d'entre elles sont maintenant bien comprises [5] et les connaissances sur la géométrie à courbure négative constante ont beaucoup progressé [6]. Cependant, la huitième géométrie, qui correspond à une courbure positive constante, reste largement inexplorée. Pour cette géométrie, on a l'extension suivante de la conjecture de Poincaré :

Conjecture d'« elliptisation » de Thurston. *Toute variété de dimension 3 fermée avec un groupe fondamental fini possède une métrique à courbure positive constante ; elle est donc homéomorphe à un quotient* S^3/Γ*, où* $\Gamma \subset \mathrm{SO}(4)$ *est un groupe fini de rotations qui agit librement sur* S^3*.*

La conjecture de Poincaré correspond au cas particulier où le groupe $\Gamma \cong \pi_1(\mathrm{M}^3)$ est trivial. Les sous-groupes possibles $\Gamma \subset \mathrm{SO}(4)$ ont été classifiés il y a longtemps [19] (voir aussi [23]).

Approches par la géométrie différentielle et les équations différentielles [7]

Ces dernières années, le problème de la géométrisation (et donc la conjecture de Poincaré) a fait l'objet de plusieurs approches basées sur l'étude de la géométrie de l'espace de dimension infinie constitué de

5. Voir par exemple [3, 6, 9, 13, 38–40, 49].
6. Voir [22, 27, 30, 44, 47]. Les articles pionniers [14, 50] ont servi de base à une grande partie de ce travail.
7. Ajouté en 2004.

toutes les métriques riemanniennes sur une variété tridimensionnelle lisse donnée.

Par définition, la longueur d'un chemin γ sur une variété riemannienne se calcule avec le tenseur métrique g_{ij} comme l'intégrale

$$\int_\gamma ds = \int_\gamma \sqrt{\sum g_{ij} dx^i dx^j}$$

À partir des dérivées première et seconde de ce tenseur métrique, on peut calculer le tenseur de Ricci R_{ij} et la courbure scalaire R. À titre d'exemple, pour l'espace euclidien plat, on obtient $\mathrm{R}_{ij} = \mathrm{R} = 0$, tandis que pour une sphère tridimensionnelle de rayon r, on obtient $\mathrm{R}_{ij} = 2g_{ij}/r^2$ et une courbure scalaire $\mathrm{R} = 6/r^2$.

Une approche de Michael Anderson, inspirée des idées de Hidehiko Yamabe [53], étudie la courbure scalaire totale $\iiint_{\mathrm{M}^3} \mathrm{R}\, d\mathrm{V}$ comme une fonctionnelle sur l'espace de toutes les métriques riemanniennes lisses à volume unité. Les points critiques de cette fonctionnelle sont les métriques de courbure constante (voir [1]).

Une approche différente, initiée par Richard Hamilton, étudie le flot de Ricci [15–17], c'est-à-dire les solutions de l'équation différentielle

$$\frac{dg_{ij}}{dt} = -2\mathrm{R}_{ij}.$$

En d'autres termes, la métrique évolue avec le temps de sorte que les distances diminuent dans les directions où la courbure est positive. Il s'agit essentiellement d'une équation parabolique qui se comporte comme l'équation de la chaleur étudiée par les physiciens : si l'on chauffe l'une des extrémités d'une tige froide, la chaleur se répandra instantanément dans toute la tige jusqu'à ce qu'elle atteigne une température uniforme. De même, un espoir naïf pour les variétés de dimension 3 avec un groupe fondamental fini aurait pu être que la courbure positive s'étende sous l'effet du flot de Ricci jusqu'à ce que la variété atteigne une courbure constante à la limite (après remise à l'échelle à une taille constante). Si l'on part d'une variété de dimension 3 avec une courbure de Ricci positive, Hamilton a pu réaliser ce programme et construire une métrique à courbure constante, résolvant ainsi un cas très particulier de la conjecture d'elliptisation. Il y a cependant dans le cas général de très sérieuses difficultés, car ce flot peut évoluer vers des singularités [8].

8. Grisha Perelman, de Saint-Pétersbourg, a mis en ligne sur arXiv.org trois prépublications qui tendent à résoudre ces difficultés et qui pré-

Je tiens à remercier les nombreux mathématiciens qui m'ont aidé à rédiger ce rapport.

Mai 2000, révisé en juin 2004.

Bibliographie

[1] Anderson (M. T.), « Scalar curvature, metric degenerations and the static vacuum Einstein equations on 3-manifolds », *Geom. Funct. Anal.*, n° 9, 1999, p. 855-963 et n° 11, 2001, p. 273-381. Voir aussi : « Scalar curvature and the existence of geometric structures on 3-manifolds », *J. reine angew. Math.*, n° 553, 2002, p. 125-182 et n° 563, 2003, p. 115-195.

[2] Anderson (M. T.), « Geometrization of 3-manifolds via the Ricci flow », *Notices AMS*, n° 51, 2004, p. 184-193.

[3] Auslander (L.) et Johnson (F. E. A.), « On a conjecture of C. T. C. Wall », *J. Lond. Math. Soc.*, n° 14, 1976, p. 331-332.

[4] Bing (R. H.), « Some aspects of the topology of 3-manifolds related to the Poincaré conjecture », dans Saaty (T. L.) *Lectures on Modern Mathematics II*, New York, Wiley, 1964.

[5] Birman (J.), « Poincaré's conjecture and the homeotopy group of a closed, orientable 2-manifold », *J. Austral. Math. Soc.*, n° 17, 1974, p. 214-221.

[6] Casson (A.) et Jungreis (D.), « Convergence groups and Seifert fibered 3-manifolds », *Invent. math.*, n° 118, 1994, p. 441-456.

[7] Donaldson (S. K.), « Self-dual connections and the topology of smooth 4-manifolds », *Bull. Amer. Math. Soc.*, n° 8, 1983, p. 81-83.

[8] Freedman (M. H.), « The topology of four-dimensional manifolds », *J. Diff. Geom.*, n° 17, 1982, p. 357-453.

[9] Gabai (D.), « Convergence groups are Fuchsian groups », *Ann. Math.*, n° 136, 1992, p. 447-510.

[10] Gabai (D.), « Valentin Poenaru's program for the Poincaré conjecture », dans *Geometry, Topology, & Physics, Conf. Proc. Lecture Notes Geom. Topology, VI*, Cambridge, MA, International Press, 1995, p. 139-166.

[11] Gillman (D.) et Rolfsen (D.), « The Zeeman conjecture for standard spines is equivalent to the Poincaré conjecture », *Topology*, n° 22, 1983, p. 315-323.

tendent même démontrer la conjecture de géométrisation complète [32–34]. Ces prépublications ont suscité beaucoup d'intérêt (voir [2, 25] ainsi que le site web http://www.math.lsa.umich.edu/research/ricciflow/perelman.html organisé par B. Kleiner et J. Lott). Cependant, tous les détails n'ont pas été publiés.

[12] Gompf (R.), « An exotic menagerie », *J. Differ. Geom.*, n° 37, 1993, p. 199-223.

[13] Gordon (C.) et Heil (W.), « Cyclic normal subgroups of fundamental groups of 3-manifolds », *Topology*, n° 14, 1975, p. 305-309.

[14] Haken (W.), « Über das Homöomorphieproblem der 3-Mannigfaltigkeiten I », *Math. Z.*, n° 80, 1962, p. 89-120.

[15] Hamilton (R. S.), « Three-manifolds with positive Ricci curvature », *J. Differ. Geom.*, n° 17, 1982, p. 255-306.

[16] Hamilton (R. S.), « The formation of singularities in the Ricci flow », dans *Surveys in Differential Geometry*, vol. II, Cambridge, MA, International Press, 1995, p. 7-136.

[17] Hamilton (R. S.), « Non-singular solutions of the Ricci flow on three-manifolds », *Comm. Anal. Geom.*, n° 7, 1999, p. 695-729.

[18] Hirsch (M.), « Obstruction theories for smoothing manifolds and maps », *Bull. Amer. Math. Soc.*, n° 69, 1963, p. 352-356.

[19] Hopf (H.), « Zum Clifford-Kleinschen Raumproblem », *Math. Ann.*, n° 95, 1925-1926, p. 313-319.

[20] Jakobsche (W.), « The Bing-Borsuk conjecture is stronger than the Poincaré conjecture », *Fund. Math.*, n° 106, 1980, p. 127-134.

[21] Massey (W. S.), *Algebraic Topology : An Introduction*, New York, Harcourt Brace, 1967 ; New York, Springer, 1977 ; ou *A Basic Course in Algebraic Topology*, New York, Springer, 1991.

[22] McMullen (C.), « Riemann surfaces and geometrization of 3-manifolds », *Bull. Amer. Math. Soc.*, n° 27, 1992, p. 207-216.

[23] Milnor (J.), « Groups which act on S^n without fixed points », *Amer. J. Math.*, n° 79, 1957, p. 623-630.

[24] Milnor (J.), avec Siebenmann (L.) et Sondow (J.), *Lectures on the h-Cobordism Theorem*, Princeton University Press, 1965.

[25] Milnor (J.), « Towards the Poincaré conjecture and the classification of 3-manifolds », *Notices AMS*, n° 50, 2003, p. 1226-1233.

[26] Moise (E. E.), *Geometric Topology in Dimensions 2 and 3*, New York, Springer, 1977.

[27] Morgan (J.), « On Thurston's uniformization theorem for three-dimensional manifolds », dans Bass (H.) et Morgan (J.), *The Smith Conjecture*, New York, Academic Press, 1984, p. 37-125.

[28] Munkres (J.), « Obstructions to the smoothing of piecewise-differentiable homeomorphisms », *Ann. Math.*, n° 72, 1960, p. 521-554.

[29] Munkres (J.), *Topology : A First Course*, Englewood Cliffs, Prentice-Hall, 1975.

[30] Otal (J.-P.), « The hyperbolization theorem for fibered 3-manifolds » (traduit de la version originale de 1996 en français par L. D. Kay), Providence, American Mathematical Society, Paris, Société Mathématique de France, 2001.

[31] Papakyriakopoulos (C.), « A reduction of the Poincaré conjecture to group theoretic conjectures », *Ann. Math.*, n° 77, 1963, p. 250-305.

[32] Perelman (G.), « The entropy formula for the Ricci flow and its geometric applications », arXiv : math.DG/0211159v1, 11/11/2002.

[33] Perelman (G.), « Ricci flow with surgery on three-manifolds », arXiv : math.DG/0303109, 10/03/2003.

[34] Perelman (G.), « Finite extinction time for the solutions to the Ricci flow on certain threemanifolds », arXiv : math.DG/0307245, 17/07/2003.

[35] Poénaru (V.), « A program for the Poincaré conjecture and some of its ramifications », in *Topics in Low-dimensional Topology*, River Edge, World Scientific Publishing, 1999, p. 65-88.

[36] Poincaré (H.), *Œuvres*, tome VI, Paris, Gauthier-Villars, 1953.

[37] Rourke (C.), « Algorithms to disprove the Poincaré conjecture », *Turk. J. Math.*, n° 21, 1997, p. 99-110.

[38] Scott (P.), « A new proof of the annulus and torus theorems », *Amer. J. Math.*, n° 102, 1980, p. 241-277.

[39] Scott (P.), « There are no fake Seifert fibre spaces with infinite π_1 », *Ann. Math.*, n° 117, 1983, p. 35-70.

[40] Scott (P.), « The geometries of 3-manifolds », *Bull. Lond. Math. Soc.*, n° 15, 1983, p. 401-487.

[41] Smale (S.), « Generalized Poincaré's conjecture in dimensions greater than four », *Ann. Math.*, n° 74, 1961, p. 391-406. Voir aussi : *Bull. Amer. Math. Soc.*, n° 66, 1960, p. 373-375.

[42] Smale (S.), « The story of the higher dimensional Poincaré conjecture (What actually happened on the beaches of Rio) », *Math. Intell.*, n° 12-2, 1990, p. 44-51.

[43] Stallings (J.), « Polyhedral homotopy spheres », *Bull. Amer. Math. Soc.*, n° 66, 1960, p. 485-488.

[44] Sullivan (D.), « Travaux de Thurston sur les groupes quasi-fuchsiens et sur les variétés hyperboliques de dimension 3 fibrées sur le cercle », *Séminaire Bourbaki 554*, New York, Springer, 1981.

[45] Thickstun (T. L.), « Open acyclic 3-manifolds, a loop theorem and the Poincaré conjecture », *Bull. Amer. Math. Soc. (N.S.)*, n° 4, 1981, p. 192-194.

[46] Thurston (W. P.), « Three dimensional manifolds, Kleinian groups and hyperbolic geometry », dans *The Mathematical Heritage of Henri Poincaré*, 1983, 1re partie. Aussi dans *Bull. Amer. Math. Soc.*, n° 6, 1982, p. 357-381.

[47] Thurston (W. P.), « Hyperbolic structures on 3-manifolds, I, deformation of acyclic manifolds », *Ann. Math.*, n° 124, 1986, p. 203-246.

[48] Thurston (W. P.), *Three-Dimensional Geometry and Topology*, vol. I, Princeton University Press, 1997.

[49] Tukia (P.), « Homeomorphic conjugates of Fuchsian groups », *J. reine angew. Math.*, n° 391, 1988, p. 1-54.

[50] Waldhausen (F.), « On irreducible 3-manifolds which are sufficiently large », *Ann. Math.*, n° 87, 1968, p. 56-88.

[51] Wallace (A.), « Modifications and cobounding manifolds, II », *J. Math. Mech.*, n° 10, 1961, p. 773-809.

[52] Whitehead (J. H. C.), *Mathematical Works*, vol. II, New York, Pergamon Press, 1962. Voir p. 21-50.

[53] Yamabe (H.), « On a deformation of Riemannian structures on compact manifolds », *Osaka Math. J.*, n° 12, 1960, p. 21-37.

[54] Zeeman (E. C.), « The Poincaré conjecture for $n \geqslant 5$ », dans *Topology of 3-Manifolds and Related Topics*, Englewood Cliffs, Prentice-Hall, 1962, p. 198-204. Voir aussi *Bull. Amer. Math. Soc.*, n° 67, 1961, p. 270.

Note : Pour une collection représentative d'approches de la conjecture de Poincaré, voir [5, 10, 11, 20, 31, 35, 37, 45].

Le problème P = NP (par Stephen Cook)

Énoncé du problème

Le problème **P = NP** consiste à déterminer si tout langage accepté par un algorithme non déterministe en temps polynomial est également accepté par un algorithme (déterministe) en temps polynomial. Pour définir précisément le problème, il est nécessaire de donner un modèle formel d'ordinateur. Le modèle d'ordinateur standard dans la théorie de la calculabilité est la machine de Turing, introduite par Alan Turing en 1936 [37]. Bien que ce modèle ait été introduit avant la construction d'ordinateurs physiques, il continue néanmoins d'être accepté comme le modèle d'ordinateur approprié pour définir la notion de fonction calculable.

De manière informelle, la classe **P** est la classe des problèmes de décision qui peuvent être résolus par un algorithme dont le nombre d'étapes est polynomial en la longueur de l'entrée. Turing ne s'intéressait pas à l'efficacité de ses machines mais plutôt à leur capacité à simuler des algorithmes arbitraires en un temps suffisant. Il s'avère cependant que les machines de Turing peuvent généralement simuler des modèles d'ordinateurs plus efficaces (par exemple, des machines équipées de nombreuses bandes ou d'une mémoire vive non limitée) en élevant au maximum le temps de calcul au carré ou au cube. Ainsi, **P** est une classe robuste qui a des définitions équivalentes sur une large classe de modèles informatiques. On suit ici la pratique habituelle et l'on définit la classe **P** en termes de machines de Turing.

Formellement, les éléments de la classe **P** sont des langages. Soit Σ un alphabet fini (c'est-à-dire un ensemble fini non vide) avec au moins deux éléments. Soit Σ^* l'ensemble des suites finies sur Σ, appelées « mots ». Alors un langage sur Σ est un sous-ensemble L de Σ^*. Toute machine de Turing M possède un alphabet d'entrée associé Σ. Pour tout mot w dans Σ^*, il existe un calcul associé à M avec l'entrée w. Les notions de machine de Turing et de calcul sont définies formellement dans l'annexe. On dit que M accepte w si ce calcul se termine dans l'état d'acceptation. Notons que M n'accepte pas w si ce calcul se termine dans l'état de rejet ou si le calcul ne se termine pas. Le langage accepté

par M, noté L(M), a un alphabet associé Σ ; il est défini par

$$\mathrm{L(M)} = \{w \in \Sigma^* \mid \mathrm{M \text{ accepte } } w\}\,.$$

On note $t_{\mathrm{M}}(w)$ le nombre d'étapes du calcul de M sur l'entrée w (voir l'annexe). Si ce calcul ne s'arrête jamais, alors $t_{\mathrm{M}}(w) = \infty$. Pour $n \in \mathbb{N}$, on note $\mathrm{T_M}(n)$ le temps d'exécution de M dans le pire des cas, c'est-à-dire,

$$\mathrm{T_M}(n) = \max\{t_{\mathrm{M}}(w) \mid w \in \Sigma^n\}\,,$$

où Σ^n est l'ensemble de tous les mots sur Σ de longueur n. On dit que M s'exécute en temps polynomial s'il existe k tel que pour tout n, $\mathrm{T_M}(n) \leqslant n^k + k$. On définit maintenant la classe **P** des langages par

$$\mathbf{P} = \{\mathrm{L} \mid \mathrm{L = L(M) \text{ pour une machine de Turing M qui fonctionne en temps polynomial}}\}.$$

La notation **NP** signifie « temps polynomial non déterministe », car **NP** était défini à l'origine en termes de machines non déterministes (c'est-à-dire des machines qui ont plus d'un mouvement possible à partir d'une configuration donnée). Cependant, il est maintenant d'usage de donner une définition équivalente en utilisant la notion de relation, qui est simplement une relation binaire $\mathrm{R} \subset \Sigma^* \times \Sigma_1^*$ pour des alphabets finis Σ et Σ_1. On associe à chacune de ces relations R un langage $\mathrm{L_R}$ sur $\Sigma \cup \Sigma_1 \cup \{\#\}$ défini par

$$\mathrm{L_R} = \{w\#y \mid \mathrm{R}(w, y)\},$$

où le symbole # n'est pas dans Σ. On dit que R est en temps polynomial si $\mathrm{L_R} \in \mathbf{P}$.

On définit maintenant la classe **NP** des langages par la condition qu'un langage L sur Σ est dans **NP** si et seulement s'il existe $k \in \mathbb{N}$ et une relation en temps polynomial R telle que pour tout $w \in \Sigma^*$,

$$w \in \mathrm{L} \iff \exists y \left(|y| \leqslant |w|^k \text{ et } \mathrm{R}(w, y)\right),$$

où $|w|$ et $|y|$ désignent respectivement les longueurs de w et y.

Énoncé du problème. *Est-ce que* $\mathbf{P} = \mathbf{NP}$ *?*

Il est facile de voir que la réponse est indépendante de la taille de l'alphabet Σ (on suppose $|\Sigma| \geqslant 2$), puisque les mots sur un alphabet

de n'importe quelle taille fixée peuvent être efficacement codés par des mots sur un alphabet binaire. Pour $|\Sigma| = 1$, le problème reste ouvert, bien qu'il soit possible que $\mathbf{P} = \mathbf{NP}$ dans ce cas mais pas dans le cas général.

Il est trivial de montrer que $\mathbf{P} \subset \mathbf{NP}$, puisque pour tout langage L sur Σ, si $\mathrm{L} \in \mathbf{P}$ alors on peut définir la relation en temps polynomial $\mathrm{R} \subset \Sigma^* \cup \Sigma^*$ par

$$\mathrm{R}(w, y) \iff w \in \mathrm{L}$$

pour tout $w, y \in \Sigma^*$.

Voici deux exemples simples qui utilisent la notation décimale pour coder les entiers naturels. L'ensemble des carrés parfaits est dans $\mathbf{P}$, puisque la méthode de Newton peut être utilisée pour approximer efficacement les racines carrées. L'ensemble des nombres composés est dans $\mathbf{NP}$, où (en désignant la notation décimale d'un entier naturel c par $\bar{c}$) la relation en temps polynomial associée R est donnée par

$$\mathrm{R}(\bar{a}, \bar{b}) \iff 1 < b < a \text{ et } b \mid a \tag{1}$$

On a récemment montré que l'ensemble des nombres composés est également dans $\mathbf{P}$ [1], ce qui répond à une ancienne question ouverte.

Histoire et importance

L'importance de la question $\mathbf{P} = \mathbf{NP}$ provient du succès des théories en cryptographie basées sur la complexité et la **NP**-complétude, ainsi que des conséquences pratiques potentiellement stupéfiantes d'une démonstration constructive de $\mathbf{P} = \mathbf{NP}$.

La théorie de la **NP**-complétude trouve ses racines dans la théorie de la calculabilité, dont l'origine remonte aux travaux de Turing, Church, Gödel et d'autres dans les années trente. Les précurseurs en calculabilité des classes $\mathbf{P}$ et $\mathbf{NP}$ sont respectivement les classes de langages décidables et récursivement énumérables (**RE**). On dit qu'un langage L est récursivement énumérable (ou semi-décidable) si $\mathrm{L} = \mathrm{L(M)}$ pour une machine de Turing M. On dit que L est décidable si $\mathrm{L} = \mathrm{L(M)}$ pour une machine de Turing M qui vérifie la condition que M s'arrête pour tous les mots d'entrée w. Il existe une définition équivalente de l'énumérabilité récursive qui met en évidence son analogie avec $\mathbf{NP}$, à savoir que L est récursivement énumérable s'il existe une relation calculable $\mathrm{R}(x, y)$ telle que $\mathrm{L} = \{x \mid \exists y \mathrm{R}(x, y)\}$.

En utilisant la notation $\langle M \rangle$ pour désigner un mot décrivant une machine de Turing M, on définit le problème de l'arrêt (*halting problem*) HP ainsi :

$$\mathrm{HP} = \{\langle M \rangle \mid \text{M est une machine de Turing qui s'arrête sur l'entrée } \langle M \rangle\}.$$

Turing a utilisé un simple argument diagonal pour montrer que HP n'est pas décidable. En revanche, il n'est pas difficile de montrer que HP est récursivement énumérable.

Une notion d'une importance capitale dans la théorie de la calculabilité est la notion de réductibilité, définie par Turing *grosso modo* ainsi : un langage L_1 est réductible à un langage L_2 s'il existe une machine de Turing avec oracle M qui accepte L_1, où M est autorisé à poser des questions d'appartenance de la forme $x \in L_2$ auxquelles un « oracle » pour L_2 répond correctement. Plus tard, la notion plus restreinte de réductibilité *many-one* ($\leqslant_m$) a été introduite et définie ainsi :

Définition 1. Supposons que L_i soit un langage sur $\Sigma_i, i = 1, 2$. Alors $L_1 \leqslant_m L_2$ si et seulement s'il existe une fonction calculable (totale) $f : \Sigma_1^* \to \Sigma_2^*$ telle que $x \in L_1 \iff f(x) \in L_2$ pour tout $x \in \Sigma_1^*$.

Il est facile de voir que si $L_1 \leqslant_m L_2$ et si L_2 est décidable, alors L_1 est décidable. Ce fait fournit un outil important pour montrer l'indécidabilité ; par exemple, si $\mathrm{HP} \leqslant_m L$, alors L est indécidable.

La notion de **NP**-complétude est basée sur la notion suivante issue de la théorie de la calculabilité :

Définition 2. Un langage L est **RE**-complet si L est récursivement énumérable et $L' \leqslant_m L$ pour tout langage récursivement énumérable L'.

Il est facile de montrer que HP est **RE**-complet. Il s'avère que la plupart des langages récursivement énumérables indécidables « naturels » sont en fait **RE**-complets. Puisque $\leqslant_m$ est transitif, pour montrer qu'un langage récursivement énumérable L est **RE**-complet, il suffit de montrer que $\mathrm{HP} \leqslant_m L$.

La notion de calcul en temps polynomial a été introduite dans les années soixante par Cobham [8] et Edmonds [13] dans le cadre des premiers développements de la théorie de la complexité algorithmique, bien que von Neumann [38] ait fait dès 1953 la distinction entre les algorithmes en temps polynomial et les algorithmes en temps exponentiel. Edmonds a qualifié les algorithmes en temps polynomial de « bons algorithmes » et les a associés aux algorithmes efficaces.

Dans la théorie de la complexité, il est désormais commun d'associer temps polynomial et efficacité. Ouvrons une parenthèse pour discuter de ce point. Il n'est bien sûr pas littéralement vrai que tout algorithme en temps polynomial peut être exécuté de manière efficace sur de petites entrées ; par exemple, un programme informatique qui nécessite n^{100} étapes ne pourrait jamais être exécuté sur une entrée même aussi petite que $n = 10$. Voici une affirmation plus défendable (voir [10]) :

THÈSE D'EFFICACITÉ. *Un problème naturel a un algorithme efficace si et seulement s'il a un algorithme en temps polynomial.*

Les exemples de problèmes naturels pour lesquels il existe des algorithmes efficaces et en temps polynomial abondent : l'arithmétique des nombres entiers, l'algèbre linéaire, les réseaux de flot, la programmation linéaire, de nombreux problèmes de graphes (connexité, chemin le plus court, arbre couvrant de poids minimal, couplage biparti), etc. D'autre part, les résultats profonds de Robertson et Seymour [29] fournissent une source potentielle de contre-exemples à la thèse : ils démontrent que toute famille de graphes fermée par passage aux mineurs peut être reconnue en temps polynomial (en fait, en temps $\mathrm{O}\left(n^3\right)$), mais les algorithmes fournis par leur méthode ont des constantes si énormes qu'ils ne sont pas efficaces. Cependant, tout contre-exemple potentiel peut être éliminé en trouvant un algorithme efficace pour lui. Par exemple, on connaît un algorithme de reconnaissance efficace pour la classe des graphes planaires, mais aucun n'est actuellement connu pour la classe des graphes plongeables dans $\mathbb{R}^3$ sans que deux cycles ne soient entrelacés. Les deux exemples sont des familles fermées par passage aux mineurs. Bien sûr, les adjectifs « naturel » et « efficace » dans la thèse ci-dessus doivent être expliqués. En général, on ne considère pas naturelle une classe avec un paramètre, comme par exemple l'ensemble des graphes plongeables dans une surface de genre k, avec $k > 1$.

Mentionnons deux problèmes liés au sens « seulement si » de la thèse. Le premier concerne les algorithmes aléatoires. À la fin de la troisième section, on évoquera la possibilité d'utiliser une source de bits aléatoire pour réduire considérablement le temps de reconnaissance requis pour certains langages. Notons toutefois qu'il n'est pas certain qu'une source véritablement aléatoire existe dans la nature. La deuxième préoccupation concerne les ordinateurs quantiques. Ce modèle d'ordinateur incorpore l'idée de superposition des états de la

mécanique quantique et permet une accélération potentiellement exponentielle de certains calculs par rapport aux machines de Turing. Par exemple, Shor [32] a montré qu'un algorithme pour ordinateur quantique est capable de factoriser des nombres entiers en temps polynomial, alors qu'on ne connaît aucun algorithme de factorisation des nombres entiers en temps polynomial pour les machines de Turing. Les physiciens ont jusqu'à présent été incapables de construire un ordinateur quantique capable de traiter plus d'une demi-douzaine de bits, de sorte que cette menace pour la thèse d'efficacité reste hypothétique à l'heure actuelle.

Pour revenir au traitement historique de la théorie de la complexité, l'auteur [9] a introduit en 1971 une notion de **NP**-complétude en tant qu'analogue en temps polynomial de la **RE**-complétude, sauf que la réduction utilisée était un analogue en temps polynomial de la réductibilité de Turing plutôt que de la réductibilité *many-one*. Les principaux résultats de [9] sont que plusieurs problèmes naturels, y compris la satisfaisabilité et le problème 3-SAT (définis ci-dessous) et le problème de l'isomorphisme de sous-graphes sont **NP**-complets. Un an après, Karp [21] a utilisé ces résultats de complétude pour montrer que vingt autres problèmes naturels sont **NP**-complets, démontrant ainsi avec force l'importance du sujet. Karp a également introduit la notation désormais standard **P** et **NP** et a redéfini la **NP**-complétude en utilisant l'analogue en temps polynomial de la réductibilité *many-one*, une définition qui est devenue standard. Parallèlement, Levin [23], indépendamment de Cook et Karp, a défini la notion de « problème de recherche universel », similaire au problème **NP**-complet, et en a donné six exemples, dont la satisfaisabilité.

Les définitions standard concernant la **NP**-complétude sont des analogues proches des définitions 1 et 2 ci-dessus.

Définition 3. Supposons que L_i soit un langage sur $\Sigma_i, i = 1, 2$. Alors $\mathrm{L}_1 \leqslant_p \mathrm{L}_2$ (L_1 est réductible en temps polynomial à L_2) si et seulement s'il existe une fonction calculable en temps polynomial $f : \Sigma_1^* \to \Sigma_2^*$ telle que $x \in \mathrm{L}_1 \iff f(x) \in \mathrm{L}_2$ pour tout $x \in \Sigma_1^*$.

Définition 4. Un langage L est **NP**-complet si et seulement si L est dans **NP** et $\mathrm{L}' \leqslant_p \mathrm{L}$ pour tout langage L' dans **NP**.

La proposition suivante est facile à démontrer : la partie (b) utilise la transitivité de $\leqslant_p$ et la partie (c) découle de la partie (a).

Proposition 1.
(a) *Si* $L_1 \leqslant_p L_2$ *et* $L_2 \in \mathbf{P}$, *alors* $L_1 \in \mathbf{P}$.
(b) *Si* L_1 *est* **NP***-complet,* $L_2 \in \mathbf{NP}$ *et* $L_1 \leqslant_p L_2$, *alors* L_2 *est* **NP***-complet.*
(c) *Si* L *est* **NP***-complet et* $L \in \mathbf{P}$, *alors* $\mathbf{P} = \mathbf{NP}$.

Remarquons que les parties (a) et (b) ont des analogues proches dans la théorie de la calculabilité. L'analogue de la partie (c) est simplement que si L est **RE**-complet, alors L est indécidable. La partie (b) est la méthode de base pour montrer que de nouveaux problèmes sont **NP**-complets. La partie (c) explique pourquoi c'est probablement une perte de temps de chercher un algorithme en temps polynomial pour un problème **NP**-complet.

Dans la pratique, un membre de **NP** se présente sous la forme d'un problème de décision, et le langage correspondant s'entend comme l'ensemble des mots codant les instances OUI du problème de décision à l'aide de méthodes de codage standard. Ainsi, le problème de la satisfaisabilité est le suivant : étant donné une formule propositionnelle F, déterminer si F est satisfaisable. Pour montrer que ce problème est dans **NP**, on définit la relation en temps polynomial $R(x, y)$, qui est vraie si x code une formule propositionnelle F et y code une assignation de vérité aux variables de F qui rendent F vrai. [9] a démontré que ce problème était **NP**-complet, essentiellement en montrant que pour toute machine de Turing à temps polynomial M qui reconnaît une relation $R(x, y)$ pour un langage **NP** L, il existe un algorithme en temps polynomial qui prend en entrée un mot x et produit une formule propositionnelle F_x (qui décrit le calcul de M sur l'entrée (x, y) avec des variables représentant le mot inconnu y) telle que F_x soit satisfaisable si et seulement si M accepte l'entrée (x, y) pour un certain y avec $|y| \leqslant |x|^{O(1)}$.

Un cas particulier important de satisfaisabilité est le problème 3-SAT, dont on a également montré dans [9] qu'il était **NP**-complet. Par exemple, la formule

$$(P \vee Q \vee R) \wedge (\bar{P} \vee Q \vee \bar{R}) \wedge (P \vee \bar{Q} \vee S) \wedge (\bar{P} \vee \bar{R} \vee \bar{S}) \qquad (2)$$

est une instance OUI de 3-SAT puisque l'affectation τ vérifie la formule, où $\tau(P) = \tau(Q) =$ Vrai et $\tau(R) = \tau(S) =$ Faux.

Plusieurs centaines de problèmes **NP**-complets ont été identifiés, notamment le problème de la somme de sous-ensembles — étant donné

un ensemble fini d'entiers strictement positifs et une cible T, existe-t-il un sous-ensemble dont la somme est égale à T ? — et de nombreux problèmes de graphes : étant donné un graphe G, G possède-t-il un cycle hamiltonien ? Est-ce que G a une clique composée de la moitié des sommets ? Les sommets de G peuvent-ils être colorés avec trois couleurs avec des couleurs distinctes pour les sommets adjacents ? Ces problèmes donnent lieu à de nombreux problèmes d'ordonnancement et de routage d'importance industrielle. Le livre [15] constitue une excellente référence en la matière, avec trois cents problèmes **NP**-complets répertoriés en annexe.

À chaque problème de décision dans **NP** est associé un problème de recherche : étant donné un mot x, trouver un mot y qui vérifie la relation $\mathrm{R}(x, y)$ pour le problème (ou établir que x est une instance NON du problème). On dit qu'un tel y est un certificat pour x. Dans le cas d'un problème **NP**-complet, il est facile de voir que le problème de recherche peut être efficacement réduit au problème de décision correspondant. En fait, si $\mathbf{P} = \mathbf{NP}$, le problème de recherche associé à chaque problème **NP** possède un algorithme en temps polynomial. Par exemple, un algorithme pour le problème de décision « satisfaisabilité » peut être utilisé pour trouver une affectation de vérité τ qui vérifie une formule satisfaisable donnée F en définissant, pour chaque variable P de F à tour de rôle, P comme Vrai dans F ou Faux dans F, selon le cas qui maintient F satisfaisable.

L'ensemble des complémentaires des langages **NP** est noté co**NP**. On conjecture généralement que le complémentaire d'un langage **NP**-complet n'est pas dans **NP**, sinon $\mathbf{NP} = \mathrm{co}\mathbf{NP}$. L'ensemble TAUT des tautologies (formules propositionnelles vraies sous toutes les affectations) est l'exemple standard d'un langage co**NP**-complet. La conjecture $\mathbf{NP} \neq \mathrm{co}\mathbf{NP}$ est équivalent à l'affirmation qu'aucun système de preuve formelle (convenablement défini) pour les tautologies n'a de preuves courtes (polynomiales) pour toutes les tautologies [12]. Ce fait a motivé le développement d'une riche théorie de la complexité des preuves propositionnelles [22], dont l'un des objectifs est de démontrer des bornes inférieures superpolynomiales sur la longueur des preuves pour les systèmes de preuves propositionnelles standard.

Il existe des exemples intéressants de problèmes **NP** dont on ne sait pas s'ils sont soit dans **P**, soit **NP**-complets. Un exemple est le problème de l'isomorphisme de graphes : étant donné deux graphes non orientés, déterminer s'ils sont isomorphes.

Un autre exemple jusqu'à récemment était l'ensemble des nombres

composés. Comme mentionné dans la première section, cet ensemble est dans **NP** avec la relation (1). Mais on sait maintenant qu'il est aussi dans **P** [1]. Cependant, le problème de recherche associé à la relation (1) est équivalent au problème de la factorisation des entiers. On pense qu'il est peu probable qu'il puisse être résolu en temps polynomial. En fait, un algorithme de factorisation efficace casserait le système de chiffrement à clé publique RSA [28] couramment utilisé pour permettre des transactions financières (supposées) sécurisées sur Internet.

Il existe un problème de décision **NP** dont la complexité est équivalente à celle de la factorisation des entiers, à savoir

$$\mathrm{L}_{\mathrm{fact}} = \{\langle a, b\rangle \mid \exists d(1 < d < a \text{ et } d|b)\}.$$

Étant donné un entier $b > 1$, le plus petit diviseur premier de b peut être trouvé avec environ $\log_2 b$ requêtes à $\mathrm{L}_{\mathrm{fact}}$, en utilisant la recherche dichotomique. Il est facile de voir que le complémentaire de $\mathrm{L}_{\mathrm{fact}}$ est aussi dans **NP** : un certificat montrant que $\langle a, b\rangle$ n'est pas dans $\mathrm{L}_{\mathrm{fact}}$ pourrait être la décomposition complète en nombres premiers de b. Ainsi $\mathrm{L}_{\mathrm{fact}}$ est un bon exemple d'un problème dans **NP** qui a peu de chances d'être soit dans **P**, soit **NP**-complet.

La théorie de la complexité algorithmique joue un rôle important dans la cryptographie moderne [16]. La sécurité d'Internet, y compris la plupart des transactions financières, dépend d'hypothèses de la théorie de la complexité telles que la difficulté de factoriser des nombres entiers ou de casser le cryptosystème DES (*Data Encryption Standard*, norme de chiffrement de données). Si **P** = **NP**, ces hypothèses sont toutes fausses. Plus précisément, on pourrait utiliser un algorithme qui résoudrait 3-SAT en n^2 étapes pour factoriser des nombres à deux cents chiffres en quelques minutes.

Bien qu'un algorithme pratique pour résoudre un problème **NP**-complet (montrant que **P** = **NP**) aurait des conséquences dévastatrices pour la cryptographie, il aurait également des conséquences pratiques étonnantes d'une nature plus positive, et pas seulement en raison des solutions efficaces aux nombreux problèmes **NP**-difficiles importants pour l'industrie. Par exemple, il transformerait les mathématiques en permettant à un ordinateur de trouver une preuve formelle de tout théorème ayant une preuve de longueur raisonnable, puisque les preuves formelles peuvent facilement être reconnues en temps polynomial. De tels théorèmes pourraient bien inclure tous les problèmes du prix de l'institut de mathématiques Clay. Bien que les preuves formelles puissent ne pas être initialement intelligibles pour les humains, le problème de

trouver des preuves intelligibles se réduirait à celui de trouver un algorithme de reconnaissance pour les preuves intelligibles. Des remarques similaires s'appliquent à diverses activités humaines créatives, telles que la conception d'ailes d'avion, la création de théories physiques ou même la composition musicale. Dans chaque cas, la question est de savoir dans quelle mesure il est possible de trouver un algorithme efficace pour reconnaître un bon résultat. Il s'agit d'un problème fondamental en intelligence artificielle, dont la solution elle-même serait facilitée par le solveur **NP** en permettant de tester facilement les théories de reconnaissance.

Même si $\mathbf{P} \neq \mathbf{NP}$, il se peut que tout problème **NP** soit susceptible d'être résolu par un algorithme en temps polynomial qui fonctionne sur « la plupart » des entrées, ce qui pourrait rendre la cryptographie impossible et apporter la plupart des avantages d'un monde dans lequel $\mathbf{P} = \mathbf{NP}$. Cela sert également de motivation à la théorie de Levin ([24] et [18]) sur la complétude en moyenne, dans laquelle la question $\mathbf{P} = \mathbf{NP}$ est remplacée par la question de savoir si tout problème **NP** avec une distribution de probabilité raisonnable sur ses entrées peut être résolu en temps polynomial en moyenne.

Dans [34], Smale inclut la question $\mathbf{P} = \mathbf{NP}$ comme le problème n° 3 des problèmes mathématiques du siècle prochain. Cependant, Smale s'intéresse non seulement à la version classique de la question mais aussi à une version exprimée en termes de corps des nombres complexes. Dans ce cas, les machines de Turing doivent être remplacées par un modèle de machines capables de faire de l'arithmétique exacte et des tests de zéro sur des nombres complexes arbitraires. La question $\mathbf{P} = \mathbf{NP}$ est remplacée par une question liée au théorème des zéros de Hilbert (*Nullstellensatz*) : existe-t-il un algorithme en temps polynomial qui, étant donné un ensemble de k polynômes à plusieurs variables sur $\mathbb{C}$, détermine s'ils ont un zéro commun ? Voir [4] pour un développement de la théorie de la complexité dans ce cadre.

Les livres de Papadimitriou [25] et de Sipser [33] constituent de bonnes introductions à la théorie de la complexité.

La conjecture et les tentatives de démonstration

La plupart des théoriciens de la complexité pensent que $\mathbf{P} \neq \mathbf{NP}$. Cela s'explique peut-être en partie par les conséquences potentiellement étonnantes de $\mathbf{P} = \mathbf{NP}$ mentionnées ci-dessus, mais il existe de meilleures raisons. On les explique en examinant tour à tour les deux

possibilités : $\mathbf{P} = \mathbf{NP}$ et $\mathbf{P} \neq \mathbf{NP}$.

Supposons tout d'abord que $\mathbf{P} = \mathbf{NP}$ et examinons comment l'on pourrait le démontrer. La façon la plus évidente est de présenter un algorithme en temps polynomial pour 3-SAT ou l'un des quelque mille autres problèmes **NP**-complets connus. D'ailleurs, de nombreuses fausses démonstrations ont été présentées sous cette forme. Il existe une boîte à outils standard [7] pour concevoir des algorithmes en temps polynomial, qui comprend l'algorithme glouton, la programmation dynamique, la réduction à la programmation linéaire, etc. En raison de leur importance dans l'industrie, un grand nombre de programmeurs et d'ingénieurs ont tenté de trouver des algorithmes efficaces pour les problèmes **NP**-complets au cours des trente dernières années, sans succès. Il existe une motivation similaire pour casser les schémas cryptographiques qui supposent $\mathbf{P} \neq \mathbf{NP}$ pour leur sécurité.

Bien entendu, il est possible qu'un argument non constructif, comme les démonstrations de Robertson et Seymour mentionnées plus haut en liaison avec la thèse d'efficacité, montre que $\mathbf{P} = \mathbf{NP}$ sans produire d'algorithme efficace pour les problèmes **NP**-complets standard. Cependant, à l'heure actuelle, la meilleure borne supérieure obtenue pour un algorithme de résolution de 3-SAT est d'environ $1{,}5^n$, où n est le nombre de variables dans la formule d'entrée.

Supposons maintenant que $\mathbf{P} \neq \mathbf{NP}$ et examinons comment on pourrait le démontrer. Deux méthodes générales ont été essayées : la diagonalisation avec réduction et les bornes inférieures pour les circuits booléens.

La méthode de diagonalisation avec réduction a été utilisée avec succès dans la théorie de la calculabilité pour démontrer l'indécidabilité d'un grand nombre de problèmes, à commencer par le problème de l'arrêt. Elle a également été utilisée avec succès en théorie de la complexité pour démontrer des bornes inférieures superexponentielles pour des problèmes décidables très difficiles. Par exemple, l'arithmétique de Presburger, la théorie du premier ordre des entiers naturels munis de l'addition, est une théorie décidable pour laquelle Fischer et Rabin [14] ont démontré que toute machine de Turing décidant de la théorie doit utiliser au moins $2^{2^{cn}}$ étapes dans le pire des cas pour un certain $c > 0$. En général, les bornes inférieures qui utilisent la diagonalisation et la réduction « relativisent », c'est-à-dire qu'elles continuent à s'appliquer dans un cadre où l'instance du problème et la machine de Turing qui le résout peuvent faire des requêtes d'appartenance à un ensemble A arbitraire, appelé « oracle ». Cependant, [3] a montré qu'il existe un

oracle A pour lequel $\mathbf{P} = \mathbf{NP}$, ce qui suggère que la diagonalisation avec réduction ne peut pas être utilisée pour séparer ces deux classes. (Il existe des résultats non relativisants en théorie de la complexité, comme on le verra plus loin.) Il est intéressant de noter que par rapport à un oracle générique, $\mathbf{P} \neq \mathbf{NP}$ [5, 11].

Un circuit booléen est un graphe orienté acyclique fini dans lequel chaque nœud interne (ou porte) est étiqueté avec un connecteur booléen, généralement {ET, OU, NON}. Les nœuds d'entrée sont étiquetés avec des variables $x_1, \cdots, x_n$. Pour chaque affectation de 0 ou 1 à chaque variable, le circuit calcule une valeur booléenne à chaque porte, y compris la porte de sortie de la manière évidente. Il n'est pas difficile de voir que si L est un langage sur $\{0, 1\}$ qui est dans $\mathbf{P}$, alors il existe une famille de taille polynomiale de circuits booléens $\langle \mathrm{B_n} \rangle$ telle que B_n ait n entrées, et pour tout mot binaire w de longueur n, lorsque w est appliqué aux n nœuds d'entrée de B_n, le bit de sortie de B_n est 1 si et seulement si $w \in \mathrm{L}$. Dans ce cas, on dit que $\langle \mathrm{B}_n \rangle$ calcule L.

Ainsi, pour démontrer $\mathbf{P} \neq \mathbf{NP}$, il suffit de démontrer une borne inférieure superpolynomiale sur la taille de toute famille de circuits booléens résolvant un problème **NP**-complet particulier, tel que 3-SAT. En 1949, Shannon [31] a démontré que pour presque toutes les fonctions booléennes $f : \{0, 1\}^n \to \{0, 1\}$, tout circuit booléen qui calcule f nécessite au moins $2^n/n$ portes. Malheureusement, son argument de comptage ne donne aucune indication sur la manière de démontrer des bornes inférieures pour des problèmes dans **NP**. Des bornes inférieures exponentielles pour les problèmes **NP** ont été démontrées pour des modèles de circuits restreints, y compris les circuits monotones [2, 26] et les circuits de profondeur bornée avec des portes dont l'arité entrante est non bornée [6, 17, 35]. Cependant, toutes les tentatives visant à trouver des bornes inférieures, même superlinéaires, pour les circuits booléens non restreints pour des fonctions booléennes « explicitement données » se sont soldées par un échec total ; la meilleure borne inférieure démontrée jusqu'à présent est de l'ordre de $4n$. Razborov et Rudich [27] expliquent cet échec en soulignant que toutes les méthodes utilisées jusqu'à présent peuvent être classées comme des « preuves naturelles », et que les preuves naturelles pour les bornes inférieures des circuits généraux sont vouées à l'échec en supposant une certaine conjecture de la théorie de la complexité affirmant qu'il existe des générateurs de nombres pseudo-aléatoires forts. Puisque de tels générateurs ont été construits en supposant la difficulté de la factorisation des entiers, on peut en déduire le résultat surprenant qu'une preuve naturelle pour une borne

inférieure pour les circuits généraux donnerait lieu à un algorithme de factorisation plus efficace que ce qui est actuellement connu.

L'incapacité de la théorie de la complexité à démontrer des bornes inférieures intéressantes sur un modèle général de calcul est beaucoup plus profonde que l'incapacité à démontrer $\mathbf{P} \neq \mathbf{NP}$. Il est cohérent avec les connaissances actuelles que non seulement la satisfaisabilité pourrait avoir un algorithme en temps polynomial, mais qu'elle pourrait avoir un algorithme en temps linéaire sur une machine de Turing à plusieurs bandes. Il en va de même pour tous les 21 problèmes mentionnés dans l'article original de Karp [21]. Il existe des séparations de classes de complexité dont on sait qu'elles existent mais que l'on ne peut pas démontrer. Par exemple, considérons la suite d'inclusions de classes de complexité

$$\mathbf{LOGSPACE} \subset \mathbf{P} \subset \mathbf{NP} \subset \mathbf{PSPACE}\,.$$

Un simple argument diagonal montre que le premier est un sous-ensemble strict du dernier, mais on ne sait démontrer aucune séparation entre deux classes successives. Un autre exemple est **LINEAR-SIZE**, la classe des langages sur $\{0,1\}$ qui peuvent être calculés par une famille $\langle \mathrm{B}_n \rangle$ de circuits booléens de taille $\mathrm{O}(n)$. On ne sait pas si **P** ou **NP** est un sous-ensemble de **LINEAR-SIZE**, bien que Kannan [20] ait démontré qu'il existe des langages dans la hiérarchie polynomiale (une généralisation de **NP**) qui ne sont pas dans **LINEAR-SIZE**. Puisque si $\mathbf{P} = \mathbf{NP}$, la hiérarchie polynomiale se réduit à **P**, on a :

Proposition 2. *Si* $\mathbf{P} \subset \mathbf{LINEAR\text{-}SIZE}$, *alors* $\mathbf{P} \neq \mathbf{NP}$.

Cette proposition pourrait être interprétée comme une méthode pour démontrer $\mathbf{P} \neq \mathbf{NP}$, mais une opinion plus courante est que l'hypothèse est fausse.

Une question fondamentale en théorie de la complexité est de savoir si une source de bits aléatoires peut être utilisée pour accélérer considérablement la reconnaissance de certains langages, à condition que l'on soit prêt à accepter une faible probabilité d'erreur. La classe **BPP** est constituée de tous les langages L qui peuvent être reconnus par un algorithme aléatoire en temps polynomial avec au plus une probabilité d'erreur exponentiellement faible sur chaque entrée. Bien sûr, $\mathbf{P} \subset \mathbf{BPP}$, mais on ne sait pas si l'inclusion est stricte. L'ensemble des nombres premiers est dans **BPP** [36], mais on ne savait pas avant [1] s'il était dans **P**. Une raison de penser que $\mathbf{BPP} = \mathbf{P}$ est que les

algorithmes aléatoires sont souvent exécutés avec succès en utilisant un générateur de nombres pseudo-aléatoires déterministe comme source de bits « aléatoires ».

Il existe un lien indirect mais étonnant entre les deux questions $\mathbf{P} = \mathbf{BPP}$ et $\mathbf{P} = \mathbf{NP}$. Soit $\mathbf{E}$ la classe des langages reconnaissables en temps exponentiel, c'est-à-dire la classe des langages L tels que $\mathrm{L} = \mathrm{L(M)}$ pour une machine de Turing M avec $\mathrm{T_M}(n) = \mathrm{O}(2^{cn})$ pour un certain $c > 0$. Soit $\mathbf{A}$ l'affirmation qu'un langage dans $\mathbf{E}$ nécessite une complexité de circuit exponentielle :

Assertion A. *Il existe* $\mathrm{L} \in \mathbf{E}$ *et* $\epsilon > 0$ *tels que, pour toute famille de circuits* $\langle \mathrm{B}_n \rangle$ *calculant* L *et pour tout* n *suffisamment grand,* B_n *possède au moins* $2^{\epsilon n}$ *portes.*

Proposition 3. Si $\mathbf{A}$, alors $\mathbf{BPP} = \mathbf{P}$. Si **non A**, alors $\mathbf{P} \neq \mathbf{NP}$.

La première implication est un joli théorème d'Impagliazzo et Wigderson [19] et la seconde est une observation étonnante de V. Kabanets qui renforce la proposition 2. En fait, Kabanets conclut que $\mathbf{P} \neq \mathbf{NP}$ à partir d'une hypothèse plus faible que **non A**, à savoir que tout langage dans $\mathbf{E}$ peut être calculé par une famille de circuits booléens $\langle \mathrm{B}_n \rangle$ telle que pour au moins un n, B_n ait moins de portes que le maximum nécessaire pour calculer n'importe quelle fonction booléenne $f : \{0,1\}^n \to \{0,1\}$. Mais il n'y a pas de consensus sur la véracité de cette hypothèse.

On doit souligner que la proposition 3 relativise. En particulier, par rapport à tout oracle **PSPACE**-complet, l'assertion A est valable et $\mathbf{BPP} = \mathbf{P} = \mathbf{NP}$. Une construction non relativisante sera donc nécessaire pour démontrer $\mathbf{P} \neq \mathbf{NP}$ en donnant de petits circuits pour les langages de $\mathbf{E}$. Des constructions non relativisantes ont déjà été utilisées avec succès, par exemple pour démontrer que **IP** (« interactif en temps polynomial ») contient tout **PSPACE** [30]. Dans cette construction et dans d'autres, une technique clé consiste à représenter les fonctions booléennes par des polynômes à plusieurs variables sur des corps finis.

Annexe : définition de la machine de Turing

Une machine de Turing M consiste en un contrôle avec un nombre fini d'états (c'est-à-dire un programme fini) attaché à une tête de lecture/écriture qui se déplace sur une bande infinie. La bande est divisée en carrés, chacun étant capable de stocker un symbole d'un alphabet

fini Γ qui inclut le symbole blanc b. Chaque machine M dispose d'un alphabet d'entrée particulier Σ, qui est un sous-ensemble de Γ qui n'inclut pas le symbole blanc b. À chaque étape d'un calcul, M se trouve dans un état q d'un ensemble fini spécifié Q d'états possibles. Initialement, un mot d'entrée sur Σ est écrit sur des cases adjacentes de la bande ; toutes les autres cases sont blanches (contiennent b) ; la tête balaie le symbole le plus à gauche du mot d'entrée et M est dans l'état initial q_0. À chaque étape, M est dans un état q et la tête parcourt une case de la bande contenant un symbole s. L'action effectuée dépend de la paire (q, s) ; elle est spécifiée par la fonction de transition de la machine (ou programme) δ. L'action consiste à imprimer un symbole sur la case parcourue, à déplacer la tête d'une case vers la gauche ou vers la droite, et à prendre un nouvel état.

Formellement, une machine de Turing M est un quadruplet $\langle \Sigma, \Gamma, \mathrm{Q}, \delta \rangle$, où Σ, Γ et Q sont des ensembles finis non vides avec $\Sigma \subset \Gamma$ et $b \in \Gamma - \Sigma$. L'ensemble d'états Q contient trois états spéciaux q_0, q_{accept} et q_{rejet}. La « fonction de transition » δ vérifie

$$\delta : (\mathrm{Q} - \{q_{\text{accept}}, q_{\text{rejet}}\}) \times \Gamma \to \mathrm{Q} \times \Gamma \times \{-1, 1\}.$$

Si $\delta(q, s) = (q', s', h)$, l'interprétation est que si M est dans l'état q en train de parcourir le symbole s, alors q' est le nouvel état, s′ est le symbole imprimé, et la tête de la bande se déplace vers la gauche ou la droite d'une case selon que h vaut -1 ou 1.

On suppose que les ensembles Q et Γ sont disjoints.

Une « configuration » de M est un mot xqy avec x et $y \in \Gamma^*$, y n'étant pas le mot vide, et $q \in \mathrm{Q}$.

L'interprétation de la configuration xqy est que M est dans l'état q avec xy sur sa bande, avec sa tête balayant le symbole le plus à gauche de y.

Si C et C′ sont des configurations, alors $\mathrm{C} \xrightarrow{\mathrm{M}} \mathrm{C}'$ si $\mathrm{C} = xqsy$ et $\delta(q, s) = (q', s', h)$ et si l'une des conditions suivantes est remplie :

$\mathrm{C}' = xs'q'y$ et $h = 1$ et y est non vide.

$\mathrm{C}' = xs'q'b$ et $h = 1$ et y est vide.

$\mathrm{C}' = x'q'as'y$ et $h = -1$ et $x = x'a$ pour un certain $a \in \Gamma$.

$\mathrm{C}' = q'bs'y$ et $h = -1$ et x est vide.

Une configuration xqy est une « configuration d'arrêt » si $q \in \{q_{\text{accept}}, q_{\text{rejet}}\}$. Notons que pour toute configuration C qui n'est pas une configuration d'arrêt, il existe une configuration unique C′ telle que $\mathrm{C} \xrightarrow{\mathrm{M}} \mathrm{C}'$.

Le « calcul » de M sur l'entrée $w \in \Sigma^*$ est la suite unique $C_0, C_1, \ldots$ de configurations telle que $C_0 = q_0 w$ (ou $C_0 = q_0 b$ si w est vide) et $C_i \xrightarrow{M} C_{i+1}$ pour tout i avec C_{i+1} dans le calcul. Soit la suite est infinie, soit elle se termine par une configuration d'arrêt. Si le calcul est fini, le nombre d'étapes est inférieur d'une unité au nombre de configurations. Sinon, le nombre d'étapes est infini. On dit que M « accepte » w si le calcul est fini et la configuration finale contient l'état q_{accept}.

Remerciements. Je remercie Avi Wigderson et Hugh Woodin pour leurs nombreuses suggestions qui ont permis d'améliorer une version antérieure de ce texte.

Bibliographie

[1] Agrawal (M.), Kayal (N.) et Saxena (N.), « Primes is in P », *Ann. Math.*, n° 160, 2004, p. 781–793.

[2] Alon (N.) et Boppana (R. B.), « The monotone circuit complexity of boolean functions », *Combinatorica*, n° 7, 1987, p. 1–22.

[3] Baker (T.), Gill (J.) et Solovay (R.), « Relativizations of the P = ? NP question », *SIAM J. Comput.*, n° 4-4, 1975.

[4] Blum (L.), Cucker (F.), Shub (M.) et Smale (S.), *Complexity and Real Computation*, New York, Springer, 1998.

[5] Blum (M.) et Impagliazzo (R.), « Generic oracles and oracle classes », dans Chandra (A. K.), *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, Los Angeles, IEEE Computer Society Press, 1987, p. 118–126.

[6] Boppana (R. B.) et Sipser (M.), « The complexity of finite functions », dans Van Leeuwen (J.), *Handbook of Theoretical Computer Science, Volume A : Algorithms and Complexity*, Cambridge, MA, Elsevier et MIT Press, 1990, p. 759–804.

[7] Cormen (T.), Leiserson (C.), Rivest (R.) et Stein (C.), *Introduction to Algorithms*, 2e éd., New York, McGraw Hill, 2001.

[8] Cobham (A.), « The intrinsic computational difficulty of functions », dans Bar-Hille (Y.) *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, Amsterdam, Elsevier/North-Holland, 1964, p. 24–30.

[9] Cook (S.), « The complexity of theorem-proving procedures », dans *Conference Record of Third Annual ACM Symposium on Theory of Computing*, New York, ACM, 1971, p. 151–158.

[10] Cook (S.), « Computational complexity of higher type functions », dans *Proceedings of the International Congress of Mathematicians, Kyoto, Japan*, Berlin, Springer, 1991, p. 55–69.

[11] Cook (S.), Impagliazzo (R.) et Yamakami (T.), « A tight relationship between generic oracles and type-2 complexity theory », *Inform. Comput.*, n° 137, 1997, p. 159–170.

[12] Cook (S.) et Reckhow (R.), « The relative efficiency of propositional proof systems », *J. Symbolic Logic*, n° 44, 1979, p. 36–50.

[13] Edmonds (J.), « Minimum partition of a matroid into independent subsets », *J. Res. Nat. Bur. Standards Sect. B*, n° 69, 1965, p. 67–72.

[14] Fischer (M. J.) et Rabin (M. O.), « Super-exponential complexity of Presburger arithmetic », dans *Complexity of Computation 7*, Providence, American Mathematical Society, 1974, p. 27–41.

[15] Garey (M. R.) et Johnson (D. S.), *Computers and Intractability, a Guide to the Theory of NP-Completeness*, San Francisco, W. H. Freeman and Co., 1979.

[16] Goldreich (O.), *The Foundations of Cryptography*, vol. I, Cambridge University Press, 2000.

[17] Hastad (J.), « Almost optimal lower bounds for small depth circuits », dans *Randomness and Computation, Advances in Computing Research 5*, Greenwich, JAI Press Inc., 1989, p. 143–170.

[18] Impagliazzo (R.), « A personal view of average-case complexity », dans *10th IEEE Annual Conference on Structure in Complexity Theory*, Washington, IEEE Computer Society Press, 1995, p. 134–147.

[19] Impagliazzo (R.) et Wigderson (A.), « P = BPP if E requires exponential circuits : Derandomizing the XOR lemma », dans *ACM Symposium on Theory of Computing (STOC)*, New York, ACM, 1997, p. 220–229.

[20] Kannan (R.), « Circuit-size lower bounds and non-reducibility to sparse sets », *Inform. Control*, n° 55, 1982, p. 40–56.

[21] Karp (R. M.), « Reducibility among combinatorial problems », dans Miller (R. E.) et Thatcher (J. W.), *Complexity of Computer Computations*, New York, Plenum Press, 1972, p. 85–103.

[22] Krajicek (J.), *Bounded Arithmetic, Propositional Logic, and Complexity Theory*, Cambridge University Press, 1995.

[23] Levin (L.), « Universal search problems » (en russe), *Probl. Peredachi Inf.*, n° 9, 1973, p. 265–266. Traduction en anglais dans Trakhtenbrot (B. A.), « A survey of Russian approaches to Perebor (brute-force search) algorithms », *Ann. Hist. Comput.*, n° 6, 1984, p. 384-400.

[24] Levin (L.), « Average case complete problems », *SIAM J. Comput.*, n° 15, 1986, p. 285–286.

[25] Papadimitriou (C.), *Computational Complexity*, Reading, Addison-Wesley, 1994.

[26] Razborov (A. A.), « Lower bounds on the monotone complexity of some boolean functions », *Soviet Math. Dokl.*, n° 31, 1985, p. 354–357.

[27] Razborov (A. A.) et Rudich (S.), « Natural proofs », *J. Comput. Syst. Sci.*, n° 55, 1997, p. 24–35.

[28] Rivest (R. L.), Shamir (A.) et Adleman (L.), « A method for obtaining digital signatures and public-key cryptosystems », *Comm. ACM*, n° 21, 1978, p. 120–126.

[29] Robertson (N.) et Seymour (P. D.), « Graph minors i-xiii », *J. Comb. Theory B*, 1983–1995.

[30] Shamir (A.), « IP = PSPACE », *J. ACM*, n° 39, 1992, p. 869–977.

[31] Shannon (C.), « The synthesis of two-terminal switching circuits », *Bell Syst. Tech. J.*, n° 28, 1949, p. 59–98.

[32] Shor (P.), « Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer », *SIAM J. Comput.*, n° 26, 1997, p. 1484–1509.

[33] Sipser (M.), *Introduction to the Theory of Computation*, Boston, PWS Publ., 1997.

[34] Smale (S.), « Mathematical problems for the next century », *Math. Intell.*, n° 20-2, 1998, p. 7–15.

[35] Smolensky (R.), « Algebraic methods in the theory of lower bounds for boolean circuit complexity », dans *ACM Symposium on Theory of Computing (STOC) 19*, New York, ACM, 1987, p. 77-82.

[36] Solovay (R.) et Strassen (V.), « A fast Monte-Carlo test for primality », *SIAM J. Comput.*, n° 6, 1977, p. 84–85.

[37] Turing (A.), « On computable numbers with an application to the Entscheidnungsproblem », *Proc. London Math. Soc.*, n° 42, 1936, p. 230–265.

[38] von Neumann (J.), « A certain zero-sum two-person game equivalent to the optimal assignment problem », dans Kahn (H. W.) et Tucker (A. W.), *Contributions to the Theory of Games II*, Princeton University Press, 1953, p. 5-12.

L'hypothèse de Riemann (par Enrico Bombieri)

Le problème

La fonction zêta de Riemann est la fonction de la variable complexe s, définie dans le demi-plan [1] $\Re(s) > 1$ par la série absolument convergente

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s},$$

et dans tout le plan complexe $\mathbb{C}$ par prolongement analytique. Comme l'a montré Riemann, $\zeta(s)$ se prolonge à $\mathbb{C}$ en une fonction méromorphe avec seulement un pôle simple en $s = 1$ avec un résidu 1, et vérifie l'équation fonctionnelle

$$\pi^{-s/2}\,\Gamma\left(\frac{s}{2}\right)\,\zeta(s) = \pi^{-(1-s)/2}\,\Gamma\left(\frac{1-s}{2}\right)\,\zeta(1-s). \qquad (1)$$

Dans un mémoire publié en 1859 qui a fait date, Riemann [18] a obtenu une formule analytique pour le nombre de nombres premiers jusqu'à une limite donnée. Cette formule s'exprime en termes de zéros de la fonction zêta, à savoir les solutions $\rho \in \mathbb{C}$ de l'équation $\zeta(\rho) = 0$.

Dans cet article, Riemann introduit la fonction de la variable complexe t définie par

$$\xi(t) = \tfrac{1}{2}\,s(s-1)\,\pi^{-s/2}\,\Gamma\left(\frac{s}{2}\right)\,\zeta(s)$$

avec $s = \frac{1}{2} + it$. Il montre que $\xi(t)$ est une fonction entière paire de t dont les zéros ont une partie imaginaire comprise entre $-i/2$ et $i/2$. Il affirme en outre, en esquissant une démonstration, que dans l'intervalle compris entre 0 et T, la fonction $\xi(t)$ a environ $(\mathrm{T}/2\pi)\log(\mathrm{T}/2\pi) - \mathrm{T}/2\pi$ zéros. Riemann poursuit :

> *Man findet nun in der That etwa so viel reelle Wurzeln innerhalb dieser Grenzen, und es ist sehr wahrscheinlich,*

1. On désigne par $\Re(s)$ et $\Im(s)$ les parties réelle et imaginaire de la variable complexe s. L'utilisation de la variable s se trouve déjà dans le célèbre travail de Dirichlet de 1837 sur les nombres premiers en progression arithmétique.

> *dass alle Wurzeln reell sind.*
> « On trouve, en effet, entre ces limites un nombre environ égal à celui-ci, de racines réelles, et il est très probable que toutes les racines sont réelles [2]. »

L'affirmation que tous les zéros de la fonction $\xi(t)$ sont réels est l'hypothèse de Riemann.

La fonction $\zeta(s)$ a des zéros aux entiers pairs négatifs -2, -4... On les appelle les « zéros triviaux ». Les autres zéros sont les nombres complexes $\frac{1}{2} + i\alpha$ où α est un zéro de $\xi(t)$. Ainsi, pour ce qui est de la fonction $\zeta(s)$, on peut énoncer ceci :

HYPOTHÈSE DE RIEMANN. *Les zéros non triviaux de $\zeta(s)$ ont une partie réelle égale à $\frac{1}{2}$.*

De l'avis de nombreux mathématiciens, l'hypothèse de Riemann, et son extension à des classes générales de fonctions L, est probablement aujourd'hui le problème ouvert le plus important en mathématiques pures.

Histoire et importance de l'hypothèse de Riemann

Pour les références relatives aux débuts de l'histoire des fonctions zêta et de la théorie des nombres premiers, on renvoie à Landau [13] et Edwards [6].

Le lien entre les nombres premiers et la fonction zêta au moyen du célèbre produit eulérien

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1},$$

valable pour $\Re(s) > 1$, apparaît pour la première fois dans le livre d'Euler, *Introductio in analysin infinitorum*, publié en 1748. Euler a également étudié les valeurs de $\zeta(s)$ pour les entiers pairs positifs et négatifs. Il a trouvé une équation fonctionnelle équivalente à l'équation fonctionnelle de Riemann pour la fonction très voisine $\sum(-1)^{n-1}/n^s$ (voir l'intéressant compte rendu des travaux d'Euler dans le livre de Hardy [8]) Le problème de la distribution des nombres premiers a été abordé pour la première fois par Gauss et Legendre à la fin du

2. N.D.T. Traduction de Léonce Laugel. *Œuvres mathématiques de Riemann*, Paris, Gauthier-Villars, 1898, p. 169.

XVIII^e siècle. Dans une lettre adressée à l'astronome Hencke en 1849, Gauss a déclaré qu'il avait découvert dans sa jeunesse que le nombre $\pi(x)$ de nombres premiers inférieurs ou égaux à x est bien approché par la fonction [3]

$$\mathrm{Li}(x) = \int_0^x \frac{\mathrm{d}t}{\log t}\,.$$

En 1837, Dirichlet a démontré son célèbre théorème d'existence d'un nombre infini de nombres premiers dans toute progression arithmétique $qn+a$ avec q et a qui sont des entiers strictement positifs premiers entre eux. Le 24 mai 1848, Tchebychev a lu à l'Académie de Saint-Pétersbourg son premier mémoire sur la distribution des nombres premiers, qui a été publié en 1850. Ce mémoire contient la première étude de la fonction $\pi(x)$ par des méthodes analytiques. Tchebychev commence par prendre le logarithme du produit eulérien, obtenant ainsi [4]

$$-\sum_p \log\left(1-\frac{1}{p^s}\right) + \log(s-1) = \log\big((s-1)\zeta(s)\big), \qquad (2)$$

qui est son point de départ.

Il démontre ensuite la formule intégrale

$$\zeta(s) - 1 - \frac{1}{s-1} = \frac{1}{\Gamma(s)} \int_0^\infty \left(\frac{1}{e^x-1} - \frac{1}{x}\right) e^{-x} x^{s-1}\,\mathrm{d}x, \qquad (3)$$

dont il déduit que $(s-1)\zeta(s)$ a pour limite 1 et a également des dérivées de tout ordre qui sont finies lorsque s tend vers 1 par valeurs supérieures. Il observe ensuite que les dérivées de tout ordre du premier membre de (2) peuvent être écrites sous la forme d'une fraction dont le numérateur est un polynôme en les dérivées de $(s-1)\zeta(s)$ et dont le dénominateur est une puissance entière de $(s-1)\zeta(s)$, d'où il découle que le premier membre de (2) a des dérivées de tout ordre qui sont finies lorsque s tend vers 1 par valeurs supérieures. À partir de là, il peut démontrer que s'il existe une formule asymptotique pour $\pi(x)$ sous la forme d'une somme finie $\sum a_k x/(\log x)^k$ jusqu'à un ordre $\mathrm{O}(x/(\log x)^{\mathrm{N}})$, alors $a_k = (k-1)!$ pour $k = 1, \ldots, \mathrm{N}-1$. Il s'agit bien du développement asymptotique de la fonction $\mathrm{Li}(x)$, ce qui confirme l'intuition de Gauss. Un deuxième article de Tchebychev donne des démonstrations rigoureuses de majorants et de minorants

3. L'intégrale est une valeur principale de Cauchy.

4. Tchebychev utilise $1+\rho$ à la place de notre s. On écrit ses formules en notation moderne.

explicites pour $\pi(x)$, d'un ordre de grandeur correct. Il introduit ici les fonctions de comptage

$$\vartheta(x) = \sum_{p \leqslant x} \log p\,, \quad \psi(x) = \vartheta(x) + \vartheta(\sqrt[2]{x}) + \vartheta(\sqrt[3]{x}) + \dots$$

et démontre l'identité [5]

$$\sum_{n \leqslant x} \psi\left(\frac{x}{n}\right) = \log [x]!\,.$$

À partir de cette identité, il obtient finalement des majorants et des minorants pour $\psi(x)$, $\vartheta(x)$ et $\pi(x)$.

Des variantes bien connues de la méthode de Tchebychev, basées sur le caractère entier de certains rapports de factorielles, sont apparues beaucoup plus tard et ne peuvent pas être attribuées à Tchebychev.

Le mémoire de Riemann sur $\pi(x)$ est vraiment étonnant pour la nouveauté des idées introduites. Il écrit d'abord $\zeta(s)$ en utilisant la formule intégrale, valable pour $\Re(s) > 1$,

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{e^{-x}}{1 - e^{-x}}\, x^{s-1}\, \mathrm{d}x, \tag{4}$$

et déforme ensuite le contour d'intégration dans le plan complexe pour obtenir une représentation valable pour tout s. Cela donne le prolongement analytique et l'équation fonctionnelle de $\zeta(s)$. Il donne ensuite une seconde démonstration de l'équation fonctionnelle sous la forme symétrique (1), introduit la fonction $\xi(t)$ et énonce certaines de ses propriétés en tant que fonction de la variable complexe t.

Riemann poursuit en écrivant le logarithme du produit eulérien sous la forme d'une transformée intégrale valable pour $\Re(s) > 1$

$$\frac{1}{s}\, \log \zeta(s) = \int_1^\infty \Pi\,(x) x^{-s-1}\, \mathrm{d}x, \tag{5}$$

où

$$\Pi\,(x) = \pi(x) + \tfrac{1}{2}\, \pi(\sqrt[2]{x}) + \tfrac{1}{3}\, \pi(\sqrt[3]{x}) + \cdots\,.$$

Par inversion de Fourier, il est capable d'exprimer $\Pi(x)$ comme une intégrale complexe et de la calculer en utilisant le calcul des résidus. Les résidus se trouvent aux singularités de $\log \zeta(s)$ en $s = 1$ et aux zéros

5. Ici $[x]$ désigne la partie entière de x.

de $\zeta(s)$. Enfin, une formule d'inversion exprimant $\pi(x)$ en fonction de $\Pi(x)$ donne la formule de Riemann. Il s'agit d'un résultat remarquable qui a immédiatement attiré l'attention.

Même si la méthode initiale de Riemann a pu être influencée par Tchebychev (on trouve plusieurs références explicites à Tchebychev dans le *Nachlass* non publié de Riemann [6]), sa grande contribution a été de montrer comment la distribution des nombres premiers est déterminée par les zéros complexes de la fonction zêta. À première vue, l'hypothèse de Riemann semble n'être qu'une propriété intéressante et plausible de la fonction spéciale $\zeta(s)$, et Riemann lui-même semble partager ce point de vue. Il écrit :

> *Hiervon wäre allerdings ein strenger Beweis zu wünschen ; ich habe indess die Aufsuchung desselben nach einigen flüchtigen vergeblichen Versuchen vorläufig bei Seite gelassen, da er für den nächsten Zweck meiner Untersuchung entbehrlich schien.*
> « Il serait à désirer, sans doute, que l'on eût une démonstration rigoureuse de cette proposition ; néanmoins j'ai laissé cette recherche de côté pour le moment après quelques rapides essais infructueux, car elle paraît superflue pour le but immédiat de mon étude [7]. »

D'autre part, il ne faut pas tirer de ce commentaire la conclusion que l'hypothèse de Riemann n'était pour Riemann qu'une remarque occasionnelle d'un intérêt mineur. La validité de l'hypothèse de Riemann équivaut à dire que l'écart du nombre de nombres premiers par rapport à la moyenne $\mathrm{Li}(x)$ est

$$\pi(x) = \mathrm{Li}(x) + \mathrm{O}\big(\sqrt{x}\log x\big).$$

Le terme d'erreur ne peut pas être amélioré de beaucoup, puisqu'on sait qu'il oscille dans les deux sens à un ordre d'au moins $\mathrm{Li}(\sqrt{x})\log\log\log x$ (Littlewood). Au vu des commentaires de Riemann à la fin de son mémoire sur l'approximation de $\pi(x)$ par $\mathrm{Li}(x)$, il est tout à fait probable qu'il ait compris que son hypothèse était au cœur de la question de la qualité de l'approximation de $\pi(x)$ que l'on peut obtenir à partir de sa formule.

6. Le *Nachlass*, constitué des notes non publiées de Riemann, est conservé à la bibliothèque mathématique de l'université de Göttingen. La partie concernant la fonction zêta a été analysée en profondeur par C. L. Siegel [22].

7. N.D.T. Traduction de Léonce Laugel, *ibid.*, p. 169.

La non validité de l'hypothèse de Riemann ferait de la distribution des nombres premiers un vrai capharnaüm. Ce seul fait distingue l'hypothèse de Riemann comme la principale question ouverte de la théorie des nombres premiers.

L'hypothèse de Riemann est devenue un problème central des mathématiques pures pas seulement en raison de ses conséquences fondamentales pour la loi de distribution des nombres premiers. L'une des raisons est que la fonction zêta de Riemann n'est pas un objet isolé, mais plutôt le prototype d'une classe générale de fonctions appelées « fonctions L » associées à des objets algébriques (représentations automorphes) ou arithmétiques (variétés arithmétiques). On les appellera « fonctions L globales ». Il s'agit de séries de Dirichlet avec un certain produit eulérien, qui doivent vérifier une équation fonctionnelle et une hypothèse de Riemann. Les facteurs du produit eulérien peuvent également être considérés comme des sortes de fonctions zêta de nature locale, qui devraient également vérifier une hypothèse de Riemann appropriée (la propriété dite de Ramanujan). Les propriétés les plus importantes des objets algébriques ou arithmétiques sous-jacents à une fonction L peuvent ou devraient être décrites par la localisation de ses zéros et de ses pôles ainsi que par ses valeurs en des points particuliers.

Les conséquences d'une hypothèse de Riemann pour les fonctions L globales sont importantes et variées. On mentionne ici, pour indiquer la variété des situations auxquelles elle peut être appliquée, une forme effective extrêmement forte du théorème de densité de Tchebotariov pour les corps de nombres, la représentabilité non triviale de 0 par une forme cubique non singulière en sept variables ou plus (à condition qu'elle satisfasse aux conditions de congruence nécessaires pour la résolubilité, Hooley [9]), et le test de primalité déterministe en temps polynomial de Miller. D'autre part, de nombreux résultats profonds en théorie des nombres qui sont des conséquences d'une hypothèse de Riemann généralisée peuvent être démontrés indépendamment de celle-ci, ce qui rend la conjecture encore plus vraisemblable.

Il n'entre pas dans le cadre de cet article d'esquisser la définition des fonctions L globales, renvoyant plutôt à Iwaniec et Sarnak [10] pour une étude des propriétés attendues de ces fonctions. Il suffit ici de dire que l'étude des propriétés analytiques de ces fonctions présente d'extraordinaires difficultés.

On ne connaît le prolongement analytique des fonctions L en fonctions méromorphes ou entières que dans des cas particuliers. Par exemple, l'équation fonctionnelle pour la fonction L d'une courbe ellip-

tique sur $\mathbb{Q}$ et pour ses torsions par des caractères de Dirichlet est une conséquence facile de (et équivalente à) l'existence d'une paramétrisation de la courbe au moyen de fonctions modulaires pour un groupe de Hecke $\Gamma_0(N)$. La vraie difficulté réside dans la preuve de cette modularité. Personne ne sait démontrer cette équation fonctionnelle par des méthodes analytiques. Cependant, la modularité des courbes elliptiques sur $\mathbb{Q}$ a été établie directement, d'abord dans le cas semi-stable dans les travaux spectaculaires de Wiles [28] et de Taylor et Wiles [24] qui ont conduit à la solution du dernier théorème de Fermat, puis dans le cas général dans une prépublication récente de Breuil, Conrad, Diamond et Taylor.

Certaines fonctions L ne sont pas directement associées à des objets arithmétiques ou géométriques. L'exemple le plus simple de fonctions L qui ne sont pas de nature arithmétique ou géométrique est celui des formes d'onde de Maass pour une surface de Riemann X uniformisée par un sous-groupe arithmétique Γ de $\mathrm{PGL}(2,\mathbb{R})$. Ce sont des images réciproques $f(z)$ vers le revêtement universel $\Im(z) > 0$ de X de fonctions propres simultanées pour l'action du laplacien hyperbolique et des opérateurs de Hecke sur X.

Le cas le plus important est encore une fois le groupe $\Gamma_0(N)$. Dans ce cas, on introduit une notion de forme d'onde « primitive », analogue à la notion de caractère primitif de Dirichlet, ce qui signifie que la forme d'onde n'est pas induite par une autre forme d'onde pour un $\Gamma_0(N')$ avec N' qui est un diviseur strict de N. Pour une forme d'onde primitive, l'action des opérateurs de Hecke T_n est définie pour tout n et la fonction L peut être définie par $\sum \lambda_f(n) n^{-s}$, où $\lambda_f(n)$ est la valeur propre de T_n agissant sur la forme d'onde $f(z)$. Une telle fonction L a un produit eulérien et vérifie une équation fonctionnelle analogue à celle de $\zeta(s)$. On s'attend aussi à ce qu'elle satisfasse une hypothèse de Riemann.

On ne connaît à ce jour pas un seul exemple de validité ou d'invalidité d'une hypothèse de Riemann pour une fonction L. L'hypothèse de Riemann pour $\zeta(s)$ ne semble pas être plus facile que pour les fonctions L de Dirichlet (sauf peut-être pour les zéros réels non triviaux), ce qui amène à penser que sa solution peut nécessiter d'attaquer des problèmes plus généraux au moyen d'idées entièrement nouvelles.

Arguments en faveur de l'hypothèse de Riemann

Malgré un peu de scepticisme exprimé dans le passé dû peut-être aux nombreuses tentatives infructueuses pour remplacer des raisonnements heuristiques par une démonstration rigoureuse, il est juste de dire qu'il y a aujourd'hui de nombreux arguments en faveur de l'hypothèse de Riemann. On a déjà insisté sur le fait que l'hypothèse de Riemann généralisée était cohérente avec nos connaissances actuelles en théorie des nombres. Il y a aussi un argument particulier d'une nature plus directe, que l'on va examiner maintenant.

Premièrement, voyons un argument numérique très fort.

D'une manière intéressante, les premiers calculs numériques des premiers zéros de la fonction zêta apparaissent déjà dans le *Nachlass* de Riemann. Une vérification numérique de l'hypothèse de Riemann dans un intervalle donné peut être effectuée de la façon suivante. Le nombre $\mathrm{N(T)}$ de zéros de $\zeta(s)$ dans le rectangle $\mathcal{R}$ de sommets en $-1-i\mathrm{T}$, $2-i\mathrm{T}$, $2+i\mathrm{T}$, $-1+i\mathrm{T}$ est donné par l'intégrale de Cauchy

$$\mathrm{N(T)}-1=\frac{1}{2\pi i}\int_{\partial\mathcal{R}}-\frac{\zeta'(s)}{\zeta(s)}\,ds,$$

pourvu que T ne soit pas la partie imaginaire d'un zéro (le -1 du premier membre de cette formule est dû au pôle simple de $\zeta(s)$ en $s=1$). La fonction zêta et ses dérivées peuvent être calculées avec une précision arbitraire en utilisant la formule sommatoire de Maclaurin ou la formule de Riemann-Siegel [22]. Le nombre $\mathrm{N(T)}-1$, qui est un entier, se calcule alors en divisant par $2\pi i$ l'évaluation numérique de l'intégrale et en arrondissant sa partie réelle à l'entier le plus proche (ceci est seulement d'un intérêt théorique, car de bien meilleures méthodes existent en pratique pour calculer $\mathrm{N(T)}$ exactement). D'un autre côté, puisque la fonction $\xi(t)$ est continue et réelle lorsque la variable t est réelle, il y aura un zéro d'ordre impair entre deux points quelconques entre lesquels $\xi(t)$ change de signe. En choisissant judicieusement les points en question, on peut détecter les changements de signe de $\xi(t)$ dans l'intervalle $[-\mathrm{T},\mathrm{T}]$. Si le nombre de changements de signe est égal a $\mathrm{N(T)}$, on en conclut que tous les zéros de $\zeta(s)$ dans $\mathcal{R}$ sont simples et vérifient l'hypothèse de Riemann. De cette manière, Van de Lune, Te Riele et Winter [15] ont démontré que les 1,5 milliard premiers zéros de $\zeta(s)$, ordonnés selon leurs parties imaginaires positives croissantes, sont simples et vérifient l'hypothèse de Riemann.

L'hypothèse de Riemann est équivalente à l'assertion que tous les maxima locaux de $\xi(t)$ sont strictement positifs et que tous les minima locaux sont strictement négatifs. Si un contre-exemple existait, il devrait se trouver au voisinage des pics inhabituellement élevés de $\left|\zeta(\frac{1}{2}+it)\right|$. L'intervalle des valeurs de T ci-dessus est $\mathrm{T} \approx 5 \times 10^8$ et n'est pas assez grand pour que $\left|\zeta(\frac{1}{2}+it)\right|$ montre de tels pics, dont on sait qu'ils peuvent exister en fin de compte. De plus, les calculs effectués par Odlyzko [17] dans des intervalles bien choisis montrent que l'hypothèse de Riemann est vérifiée pour plus de 3×10^8 zéros à des hauteurs qui vont [8] jusqu'à 2×10^{20}. Ces calculs corroborent également fortement des conjectures indépendantes de Dyson et Montgomery [16] sur la distribution des intervalles entre les zéros.

Le calcul des zéros de fonctions L est plus difficile, mais cela a été fait dans plusieurs cas, notamment pour des exemples de fonctions L de Dirichlet, de fonctions L de courbes elliptiques, de fonctions L de Maass et de fonctions L d'Artin non abéliennes qui proviennent des corps de nombres de bas degré. On n'a trouvé de cette manière aucune exception à l'hypothèse de Riemann généralisée.

Deuxièmement, on sait que d'éventuelles exceptions à l'hypothèse de Riemann doivent être rares si l'on s'éloigne de la droite $\Re(s) = \frac{1}{2}$.

Soit $\mathrm{N}(\alpha, \mathrm{T})$ le nombre de zéros de $\zeta(s)$ dans le rectangle $\alpha \leqslant \Re(s) \leqslant 2$, $0 \leqslant \Im(s) \leqslant \mathrm{T}$. Le résultat prototype remonte à Bohr et Landau en 1914, à savoir $\mathrm{N}(\alpha, \mathrm{T}) = \mathrm{O}(\mathrm{T})$ pour tout α fixé tel que $\frac{1}{2} < \alpha < 1$. Une amélioration significative du résultat de Bohr et Landau a été obtenue par Carlson en 1920, qui a démontré le « théorème de densité » $\mathrm{N}(\alpha, \mathrm{T}) = \mathrm{O}(\mathrm{T}^{4\alpha(1-\alpha)+\varepsilon})$ pour tout $\varepsilon > 0$ fixé. Le fait que l'exposant ici soit strictement plus petit que 1 est important pour les applications arithmétiques, par exemple dans l'étude des nombres premiers dans de petits intervalles. L'exposant dans le théorème de Carlson a été plusieurs fois précisé pour différents intervalles de α, en particulier pour l'intervalle $\frac{3}{4} < \alpha < 1$. Curieusement, le meilleur exposant connu jusqu'à aujourd'hui dans l'intervalle $\frac{1}{2} < \alpha \leqslant \frac{3}{4}$ reste l'exposant d'Ingham $3(1-\alpha)/(2-\alpha)$, obtenu en 1940. Pour des références à ces résultats, le lecteur peut consulter la révision récente par Heath-Brown de la monographie classique de Titchmarsh [23] ainsi que le livre d'Ivić [11].

8. Les calculs les plus récents d'Odlyzko, qui sont en voie d'achèvement, permettront d'explorer complètement l'intervalle $[10^{22}, 10^{22} + 10^{10}]$.

Troisièmement, on sait que plus de 40 % des zéros non triviaux de $\zeta(s)$ sont simples et vérifient l'hypothèse de Riemann (Selberg [20], Levinson [14], Conrey [2]). La plupart de ces résultats ont été généralisés à d'autres fonctions L, notamment toutes les fonctions L de Dirichlet et les fonctions L associées à des formes modulaires ou à des formes d'onde de Maass.

Un argument supplémentaire : les variétés sur des corps finis

On peut dire que le meilleur argument en faveur de l'hypothèse de Riemann découle de la théorie correspondante développée dans le contexte des variétés algébriques sur des corps. La situation la plus simple est la suivante.

Soit C une courbe projective non singulière sur un corps fini $\mathbb{F}_q$ de caractéristique p avec $q = p^a$ éléments. Soit Div(C) le groupe additif des diviseurs de la courbe C définie sur $\mathbb{F}_q$, c'est-à-dire des sommes finies formelles $\mathfrak{a} = \sum a_i \mathrm{P}_i$ avec des $a_i \in \mathbb{Z}$ et des P_i qui sont des points de C définis sur une extension finie de $\mathbb{F}_q$, tel que $\phi(\mathfrak{a}) = \mathfrak{a}$ où ϕ est l'endomorphisme de Frobenius sur C qui élève les coordonnées à la puissance q. Le nombre $\deg(\mathfrak{a}) = \sum a_i$ est le degré du diviseur $\mathfrak{a}$. On dit que le diviseur $\mathfrak{a}$ est effectif si tous les a_i sont des entiers strictement positifs ; dans ce cas, on écrit $\mathfrak{a} > 0$. Enfin, un diviseur premier $\mathfrak{p}$ est un diviseur positif qui ne peut pas s'exprimer comme une somme de deux diviseurs positifs. Par définition, la norme d'un diviseur $\mathfrak{a}$ est $\mathrm{N}\mathfrak{a} = q^{\deg(\mathfrak{a})}$.

La fonction zêta de la courbe C, telle que définie par E. Artin, H. Hasse et F. K Schmidt, est

$$\zeta(s, \mathrm{C}) = \sum_{\mathfrak{a}>0} \frac{1}{\mathrm{N}\mathfrak{a}^s}.$$

Cette fonction a un produit eulérien

$$\zeta(s, \mathrm{C}) = \prod_{\mathfrak{p}} (1 - \mathrm{N}\mathfrak{p}^{-s})^{-1}$$

et une équation fonctionnelle

$$q^{(g-1)s}\zeta(s, \mathrm{C}) = q^{(g-1)(1-s)}\zeta(1-s, \mathrm{C}),$$

où g est le genre de la courbe C′ ; ceci est une conséquence du théorème de Riemann-Roch. La fonction $\zeta(s, \mathrm{C})$ est une fonction rationnelle de la

variable $t = q^{-s}$. Elle est donc périodique [9] de période $2\pi i/\log q$ et elle a des pôles simples en les points $s = 2\pi im/\log q$ et $s = 1 + 2\pi im/\log q$ pour $m \in \mathbb{Z}$. Exprimée en fonction de la variable t, la fonction zêta devient une fonction rationnelle $\mathrm{Z}(t, \mathrm{C})$ de t, avec des pôles simples en $t = 1$ et $t = q^{-1}$. L'utilisation de la variable t au lieu de q^{-s} est plus naturelle dans le cas géométrique. On appellera fonctions Zêta (avec un Z majuscule) les objets correspondants.

L'hypothèse de Riemann pour $\zeta(s, \mathrm{C})$ est l'assertion que tous ses zéros ont pour partie réelle $\frac{1}{2}$; cela correspond pour la fonction Zêta $\mathrm{Z}(t, \mathrm{C})$, qui a un numérateur de degré $2g$, au fait que ses zéros ont pour valeur absolue $q^{-1/2}$.

Ceci est très facile à vérifier si $g = 0$ parce que le numérateur est 1. Pour $g = 1$, Hasse a trouvé une démonstration en 1934. Le cas général pour le genre arbitraire a finalement été établi par Weil au début des années quarante (voir sa lettre à E. Artin du 10 juillet 1942, où il donne une esquisse complète de la théorie des correspondances sur une courbe [25]). Ses résultats furent finalement publiés sous forme de livre en 1948 [26].

À travers ses recherches, Weil a été amené à la formulation de conjectures générales à propos des fonctions Zêta de variétés algébriques générales sur des corps finis, reliant leurs propriétés à la structure topologique de la variété algébrique qui les sous-tend. Ici, l'hypothèse de Riemann, sous une forme simplifiée, est l'assertion que les réciproques des zéros et des pôles de la fonction Zêta (appelées racines caractéristiques) ont pour valeur absolue $q^{-d/2}$ avec d un entier positif ou nul, et s'interprètent comme les valeurs propres de l'automorphisme de Frobenius agissant sur la cohomologie de la variété. Après que M. Artin, A. Grothendieck et J.-L.Verdier ont développé l'outil fondamental de la cohomologie étale, la démonstration de l'hypothèse de Riemann correspondant aux fonctions Zêta de variétés arbitraires sur des corps finis a finalement été fournie par Deligne [3, 4]. Le théorème de Deligne peut certainement être considéré comme le couronnement des plus profondes réalisations des mathématiques du XX$^{\mathrm{e}}$ siècle. Ses applications nombreuses à la résolution de problèmes anciens en théorie des nombres, en géométrie algébrique et en mathématiques discrètes témoignent de l'importance de ces hypothèses de Riemann généralisées.

À notre avis, on ne peut pas considérer ces résultats dans le domaine géométrique comme ne relevant pas de la compréhension de l'hypothèse

9. De façon similaire, $\zeta(s)$ est presque périodique dans le demi-plan $\Re(s) \geqslant 1 + \delta$, $\delta > 0$.

de Riemann classique, ces analogies étant trop convaincantes pour être purement et simplement rejetées.

Un argument supplémentaire : la formule explicite

Une généralisation importante d'un point de vue conceptuel de la formule explicite de Riemann pour $\pi(x)$, obtenue par Weil [27] en 1952, offre un indice de ce qui peut continuer de se cacher derrière le problème.

Considérons la classe W des fonctions $f(x)$ de la demi-droite $\mathbb{R}_+$ à valeurs complexes, continues et continûment différentiables sauf pour un nombre fini de points pour lesquels à la fois $f(x)$ et $f'(x)$ présentent au plus une discontinuité de première espèce, et auxquels les valeurs de $f(x)$ et de $f'(x)$ sont définies comme moyennes de leurs limites à droite et à gauche. Supposons également qu'il existe $\delta > 0$ tel que $f(x) = \mathrm{O}(x^{\delta})$ lorsque $x \to 0^+$ et $f(x) = \mathrm{O}(x^{-1-\delta})$ lorsque $x \to +\infty$.

Soit $\tilde{f}(s)$ la transformée de Mellin

$$\tilde{f}(s) = \int_0^{\infty} f(x)\, x^s\, \frac{dx}{x},$$

qui est une fonction analytique de s pour $-\delta < \Re(s) < 1 + \delta$.

Pour la fonction zêta de Riemann, la formule de Weil peut s'exprimer comme suit. Soit $\Lambda(n) = \log p$ si $n = p^a$ est une puissance d'un nombre premier p, et 0 sinon. On a :

Formule explicite. *Pour tout* $f \in \mathcal{W}$,

$$\begin{aligned}\widetilde{f}(0) - \sum_{\rho} \widetilde{f}(\rho) + \widetilde{f}(1) = &\sum_{n=1}^{\infty} \Lambda(n) \left\{ f(n) + \frac{1}{n} f\left(\frac{1}{n}\right) \right\} \\ &+ (\log 4\pi + \gamma) f(1) \\ &+ \int_1^{\infty} \left\{ f(x) + \frac{1}{x} f\left(\frac{1}{x}\right) - \frac{2}{x} f(1) \right\} \frac{dx}{x - x^{-1}}.\end{aligned}$$

La première somme porte sur tous les zéros non triviaux de $\zeta(s)$ et doit se comprendre comme

$$\lim_{\mathrm{T} \to +\infty} \sum_{|\Im(\rho)| < \mathrm{T}} \widetilde{f}(\rho).$$

Dans son article, Weil a montré qu'il y a une formule correspondante pour les fonctions zêta et les fonctions L des corps de nombres aussi

bien que pour les fonctions Zêta des courbes sur les corps finis. Les termes dans le second membre de l'équation peuvent s'écrire comme une somme de termes de nature locale, associés aux valeurs absolues sur le corps de nombres sous-jacent ou sur le corps de fonctions sous-jacent dans le cas d'un corps de caractéristique strictement positive. De plus, dans ce dernier cas, la formule explicite peut se déduire de la formule du point fixe de Lefschetz appliquée à l'endomorphisme de Frobenius sur la courbe C. Les trois termes du premier membre de l'équation, à savoir $\widetilde{f}(0)$, $\sum \widetilde{f}(\rho)$ et $\widetilde{f}(1)$, correspondent à la trace de l'automorphisme de Frobenius de la cohomologie l-adique de C (le terme intéressant $\sum \widetilde{f}(\rho)$ correspond à la trace sur H^1), alors que le second membre correspond au nombre de points fixes de l'endomorphisme de Frobenius, c'est-a-dire aux diviseurs premiers de degré 1 sur C.

Weil a aussi démontré que l'hypothèse de Riemann est équivalente à la stricte négativité du second membre pour toutes les fonctions $f(x)$ du type

$$f(x) = \int_0^\infty g(xy)\,\overline{g(y)}\,dy,$$

quand $g \in \mathcal{W}$ vérifie les conditions supplémentaires

$$\int_0^\infty g(x)\,\frac{dx}{x} = \int_0^\infty g(x)\,dx = 0.$$

Dans le cas géométrique des courbes sur un corps fini, la stricte négativité est une conséquence assez facile du théorème de l'indice algébrique pour les surfaces :

Théorème de l'indice algébrique. *Soit* X *une surface projective non singulière définie sur un corps algébriquement clos. Alors la forme quadratique « auto-intersection »* $(\mathrm{D}\cdot\mathrm{D})$, *restreinte au groupe des diviseurs* D *sur* X *de degré 0 dans le plongement projectif de X, est semi-définie négative.*

Le théorème de l'indice algébrique pour les surfaces est essentiellement dû à Severi [10] en 1906 [21, §2, théorème I]. La démonstration utilise le théorème de Riemann-Roch sur X et la finitude des familles de

10. Severi a montré qu'un diviseur D sur X est algébriquement équivalent à 0 modulo torsion s'il est de degré 0 et si $(\mathrm{D}\cdot\mathrm{D}) = 0$. Sa démonstration reste valable sans modification sous l'hypothèse plus faible $(\mathrm{D}\cdot\mathrm{D}) \geqslant 0$, ce qui donne le théorème de l'indice.

courbes sur X d'un degré donné ; on ne connaît aucune autre démonstration par des méthodes algébriques à ce jour, même si quelques auteurs ont par la suite redécouvert indépendamment le raisonnement de Severi.

Le théorème de l'indice algébrique pour les variétés projectives non singulières de dimension paire sur les nombres complexes a d'abord été formulé et démontré par Hodge comme conséquence de sa théorie des formes harmoniques. On ne connaît aucune démonstration algébrique du théorème de Hodge. Cela reste un problème ouvert fondamental que de le généraliser aux variétés sur des corps de caractéristique strictement positive.

Les travaux de Montgomery [16], Odlyzko [17] et Rudnick et Sarnak [19] sur les corrélations pour les espacements entre zéros de $\xi(t)$ suggèrent que les fonctions L peuvent être groupées en quelques familles, dans chacune desquelles la corrélation est universelle ; la corrélation conjecturée pour les espacements est la même que celle pour la distribution asymptotique des valeurs propres de matrices aléatoires orthogonales, unitaires ou symplectiques dans des familles universelles adéquates quand la dimension tend vers ∞. Tout ceci est compatible avec la vision exprimée par Hilbert et Pólya que les zéros de $\xi(t)$ pourraient être les valeurs propres d'un opérateur linéaire auto-adjoint sur un espace de Hilbert adéquat. Il faut aussi rappeler qu'une théorie inconditionnelle correspondante pour les corrélations des espacements des racines caractéristiques des fonctions Zêta de familles de variétés algébriques sur un corps fini a été développée par Katz et Sarnak [12] en utilisant les méthodes introduites par Deligne dans sa démonstration de l'hypothèse de Riemann pour les variétés sur les corps finis. Ainsi le problème des corrélations des espacements pour les zéros des fonctions L semble être un problème très profond.

Tout ceci amène plusieurs questions fondamentales.

Y a-t-il une théorie dans le cas global qui jouerait le même rôle que celui de la cohomologie pour les fonctions Zêta des variétés sur un corps de caractéristique strictement positive ? Y a-t-il un analogue de l'automorphisme de Frobenius dans le cas classique ? Y a-t-il un théorème de l'indice généralisé avec lequel on pourrait démontrer l'hypothèse de Riemann classique ? On est ici dans le domaine des conjectures et de la spéculation. Dans le cadre adélique proposé par Tate et Weil, les articles [1, 5, 7] fournissent des idées sur ces problèmes fondamentaux.

D'un autre côté, il y a les fonctions L telles que celles associées à des formes d'onde de Maass qui ne semblent pas trouver leur origine dans la géométrie et pour lesquelles on s'attend toujours à ce qu'elles

vérifient une certaine hypothèse de Riemann. Pour ces fonctions, on n'a pas de modèles algébrique ou géométrique pour guider notre pensée. Il est possible que des idées entièrement nouvelles soient nécessaires pour étudier ces objets si énigmatiques.

Bibliographie

[1] Connes (A.), « Trace formula in noncommutative geometry and the zeros of the Riemann zeta function », *Selecta Math. (NS)*, n° 5, 1999, article 29.

[2] Conrey (J. B.), « More than two fifths of the zeros of the Riemann zeta function are on the critical line », *J. reine angew. Math.*, n° 399, 1989, p. 1-26.

[3] Deligne (P.), « La conjecture de Weil I », *Publ. math. IHES*, n° 43, 1974, p. 273-308.

[4] Deligne (P.), « La conjecture de Weil II », *Publ. math. IHES*, n° 52, 1980, p. 137-252.

[5] Deninger (C.), « Some analogies between number theory and dynamical systems on foliated spaces », dans *Proc. Int. Congress Math. (Berlin, 1998)*, vol. I, Bielefeld, Doc. Math., p. 163-186.

[6] Edwards (H. M.), *Riemann's Zeta Function*, New York, Academic Press, 1974.

[7] Haran (S.), « Index theory, potential theory, and the Riemann hypothesis », dans *L-functions and Arithmetic (Durham, 1990)*, London Mathematical Society, 1991, p. 257-270.

[8] Hardy (G. H.), *Divergent Series*, Oxford University Press, 1949, p. 23-26.

[9] Hooley (C.), « On Waring's problem », *Acta Math.*, n° 157, 1986, p. 49-97.

[10] Iwaniec (H.) et Sarnak (P.), « Perspectives on the analytic theory of L-functions », dans Alon (N.), Bourgain (J.), Connes (A.), Gromov (M.) et Milman (V.), *Visions in Mathematics*, Bâle, Birkhäuser, 2000, p. 705-741.

[11] Ivić (A.), *The Riemann Zeta-Function – The Theory of the Riemann Zeta-Function with Applications*, New York, John Wiley & Sons, 1985.

[12] Katz (N. M.) et Sarnak (P.), *Random Matrices, Frobenius Eigenvalues, and Monodromy*, Providence, American Mathematical Society, 1999.

[13] Landau (E.), *Primzahlen*, vol. II, 2e éd., New York, Chelsea, 1953.

[14] Levinson (N.), « More than one-third of the zeros of the Riemann zeta-function are on $\sigma = 1/2$ », *Adv. Math.*, n° 13, 1974, p. 383-436.

[15] Van de Lune (J.), Te Riele (J. J.) et Winter (D. T.), « On the zeros of the Riemann zeta function in the critical strip, IV », *Math. Comp.*, n° 46, 1986, p. 667-681.

[16] Montgomery (H. L.), « Distribution of the zeros of the Riemann zeta function », dans *Proc. Int. Cong. Math. (Vancouver 1974)*, vol. I, p. 379-381.

[17] Odlyzko (A. M.), « Supercomputers and the Riemann zeta function », dans Kartashev (L. P.) et Kartashev (S. I.), *Supercomputing 89 : Supercomputing Structures Computations*, St-Petersburg, International Supercomputing Institute, 1989, p. 348-352.

[18] Riemann (B.), « Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse », *Monatsber. der Königl. Preuss. Akad. der Wissen. zu Berlin*, 1859-1860, p. 671-680 ; voir aussi *Gesammelte math. Werke und wissensch. Nachlass*, 2. Aufl., 1892, p. 145-155.

[19] Rudnick (Z.) et Sarnak (P.), « Zeros of principal L-functions and random matrix theory », *Duke Math. J.*, n° 82, 1996, p. 269-322.

[20] Selberg (A.), « On the zeros of the zeta-function of Riemann », *Norske Vid. Selsk. Forhdl.*, n° 15, 1942, p. 59-62 ; voir aussi *Collected Papers*, vol. I, Berlin, Springer, 1989, p. 156-159.

[21] Severi (F.), « Sulla totalita delle curve algebriche tracciate sopra una superficie algebrica », *Math. Ann.*, n° 62, 1906, p. 194-225.

[22] Siegel (C. L.), « Über Riemanns Nachlaß zur analytischen Zahlentheorie », *Quell. Stud. Gesch. Math.*, n° 2, 1932, p. 45-80 ; voir aussi *Gesammelte Abhandlungen*, Band I, Berlin, Springer, 1966, p. 275-310.

[23] Titchmarsh (E. C.), *The Theory of the Riemann Zeta Function*, 2e éd. revue par R. D. Heath-Brown, Oxford University Press, 1986.

[24] Taylor (R.) et Wiles (A.), « Ring theoretic properties of certain Hecke algebras », *Ann. Math.*, n° 141, 1995, p. 553-572.

[25] Weil (A.), *Œuvres scientifiques – Collected Papers*, vol. I, New York, Springer, 1980, p. 280-298.

[26] Weil (A.), *Sur les courbes algébriques et les variétés qui s'en déduisent*, Paris, Hermann & Cie, 1948.

[27] Weil (A.), « Sur les « formules explicites » de la théorie des nombres premiers », *Meddel. Lunds Univ. Mat. Sem.*, 1952, p. 252-265 ; voir aussi *Œuvres scientifiques – Collected Papers*, vol. II, New York, Springer, p. 48-61.

[28] Wiles (A.), « Modular elliptic curves and Fermat's Last Theorem », *Ann. Math.*, n° 141, 1995, p. 443-551.

La théorie de Yang-Mills quantique (par Arthur Jaffe et Edward Witten)

La physique des théories de jauge

Depuis le début du XX^e siècle, on sait que la description de la nature à l'échelle subatomique passe par la mécanique quantique. En mécanique quantique, la position et la vitesse d'une particule sont des opérateurs non commutatifs dans un espace de Hilbert. Les notions classiques telles que « la trajectoire d'une particule » ne s'appliquent pas.

Mais la mécanique quantique des particules n'est qu'une partie de l'histoire. Dans la physique du XIX^e siècle et du début du XX^e siècle, de nombreux aspects de la nature étaient décrits en termes de champs : les champs électriques et magnétiques qui entrent dans les équations de Maxwell, et le champ gravitationnel régi par les équations d'Einstein. Étant donné que les champs interagissent avec les particules, il est devenu évident à la fin des années 1920 qu'une description cohérente de la nature devait intégrer des concepts quantiques pour les champs comme pour les particules.

Ce faisant, des quantités telles que les composantes du champ électrique en différents points de l'espace-temps deviennent des opérateurs non commutatifs. Lorsque l'on tente de construire un espace de Hilbert sur lequel ces opérateurs agissent, les surprises sont nombreuses. La distinction entre champs et particules s'effondre, puisque l'espace de Hilbert d'un champ quantique est construit en termes d'excitations semblables à des particules. Les particules conventionnelles telles que les électrons sont réinterprétées comme des états du champ quantifié. On redécouvre ainsi la prédiction de l'« antimatière » : pour chaque particule, il doit y avoir une antiparticule correspondante avec la même masse et une charge électrique opposée. Peu après la prédiction de P. A. M. Dirac sur la base de la théorie quantique des champs, on a découvert l'antiparticule de l'électron (le « positron ») dans les rayons cosmiques.

Les théories quantiques des champs les plus importantes pour décrire la physique des particules élémentaires sont les théories de jauge. L'exemple classique d'une théorie de jauge est la théorie de l'électromagnétisme de Maxwell. Pour l'électromagnétisme, le groupe de symé-

trie de jauge est le groupe abélien U(1). Si A représente la forme de connexion de la jauge U(1), localement une 1-forme sur l'espace-temps, alors la forme de courbure (le tenseur électromagnétique) est la 2-forme $\mathrm{F} = d\mathrm{A}$ et les équations de Maxwell en l'absence de charges et de courants ont la forme $0 = d\mathrm{F} = d * \mathrm{F}$. Ici, $*$ désigne l'opérateur de dualité de Hodge. En effet, Hodge a introduit sa célèbre théorie des formes harmoniques comme une généralisation des solutions des équations de Maxwell. Les équations de Maxwell décrivent les champs électriques et magnétiques à grande échelle et aussi, comme Maxwell l'a découvert, la propagation des ondes lumineuses à une vitesse caractéristique, la vitesse de la lumière.

L'idée d'une théorie de jauge est née des travaux de Hermann Weyl. On trouvera dans [34] une discussion intéressante sur l'histoire de la symétrie de jauge et de la découverte de la théorie de Yang-Mills [50], également connue sous le nom de « théorie de jauge non abélienne ». Au niveau classique, on remplace le groupe de jauge U(1) de l'électromagnétisme par un groupe de jauge compact G. La définition de la courbure résultant de la connexion doit être modifiée en $\mathrm{F} = d\mathrm{A} + \mathrm{A} \wedge \mathrm{A}$, et les équations de Maxwell sont remplacées par les équations de Yang-Mills, $0 = d_{\mathrm{A}}\mathrm{F} = d_{\mathrm{A}} * \mathrm{F}$, où d_{A} est l'extension covariante de jauge de la dérivée extérieure.

Ces équations classiques peuvent être déduites comme des équations variationnelles à partir du lagrangien de Yang-Mills

$$\mathrm{L} = \frac{1}{4g^2} \int \mathrm{Tr}\, \mathrm{F} \wedge *\mathrm{F}, \tag{1}$$

où Tr désigne une forme quadratique invariante sur l'algèbre de Lie de G. Les équations de Yang-Mills sont non linéaires, contrairement aux équations de Maxwell. Comme pour les équations d'Einstein du champ gravitationnel, on ne connaît que quelques solutions exactes de l'équation classique. Mais les équations de Yang-Mills ont certaines propriétés en commun avec les équations de Maxwell : elles fournissent notamment une description classique des ondes sans masse qui se déplacent à la vitesse de la lumière.

Dans les années cinquante, lorsque la théorie de Yang-Mills a été découverte, on savait déjà que la version quantique de la théorie de Maxwell, connue sous le nom d'électrodynamique quantique, rendait compte de manière extrêmement précise des champs et des forces électromagnétiques. En fait, l'électrodynamique quantique a amélioré de plusieurs ordres de grandeur la précision de certaines prédictions

antérieures de la théorie quantique et a prédit de nouvelles séparations des niveaux d'énergie.

Il était donc naturel de se demander si une théorie de jauge non abélienne pouvait décrire d'autres forces dans la nature, notamment la force faible (responsable entre autres de certaines formes de radioactivité) et la force forte ou nucléaire (responsable entre autres de la liaison des protons et des neutrons dans les noyaux). La nature sans masse des ondes classiques de Yang-Mills constituait un obstacle sérieux à l'application de la théorie de Yang-Mills aux autres forces, car les forces faible et nucléaire sont à courte portée et de nombreuses particules sont massives. Ces phénomènes ne semblaient donc pas être associés à des champs à longue portée décrivant des particules sans masse.

Dans les années soixante et soixante-dix, les physiciens ont surmonté ces obstacles à l'interprétation physique des théories de jauge non abéliennes. Dans le cas de la force faible, cela a été accompli par la théorie électrofaible de Glashow, Salam et Weinberg [40, 47] avec le groupe de jauge $H = SU(2) \times U(1)$. En élaborant la théorie avec un « champ de Higgs » supplémentaire, on évite la nature sans masse des ondes classiques de Yang-Mills. Le champ de Higgs se transforme en une représentation bidimensionnelle de H. Sa valeur non nulle et approximativement constante dans l'état de vide réduit le groupe de structure de H à un sous-groupe U(1) (diagonalement inclus dans $SU(2) \times U(1)$). Cette théorie décrit à la fois les forces électromagnétiques et les forces faibles, d'une manière plus ou moins unifiée. En raison de la réduction du groupe de structure à U(1), les champs à longue portée sont ceux de l'électromagnétisme uniquement, conformément à ce que l'on observe dans la nature.

La solution au problème des champs de Yang-Mills sans masse pour les interactions fortes est d'une nature totalement différente. Cette solution n'est pas venue de l'ajout de champs à la théorie de Yang-Mills, mais de la découverte d'une propriété remarquable de la théorie de Yang-Mills quantique elle-même, c'est-à-dire de la théorie quantique dont le lagrangien classique est donné en (1). Cette propriété est appelée « liberté asymptotique » [21, 38]. En gros, cela signifie que le champ présente à courte distance un comportement quantique très similaire à son comportement classique. Mais à longue distance, la théorie classique n'est plus un bon guide pour le comportement quantique du champ.

La liberté asymptotique ainsi que d'autres découvertes expérimentales et théoriques faites dans les années soixante et soixante-dix ont permis de décrire la force nucléaire par une théorie de jauge non abé-

lienne dans laquelle le groupe de jauge est G = SU(3). Les champs supplémentaires décrivent au niveau classique les « quarks », qui sont des objets de spin 1/2 quelque peu analogues à l'électron, mais se transformant dans la représentation fondamentale de SU(3). La théorie de jauge non abélienne de la force forte est appelée « chromodynamique quantique ».

L'utilisation de la chromodynamique quantique pour décrire la force forte a été motivée par toute une série de découvertes expérimentales et théoriques faites dans les années soixante et soixante-dix concernant les symétries et le comportement à haute énergie des interactions fortes. Mais la théorie de jauge classique non abélienne est très différente du monde observé des interactions fortes. Pour que la chromodynamique quantique décrive avec succès la force forte, elle doit présenter au niveau quantique les trois propriétés suivantes, chacune d'entre elles étant radicalement différente du comportement de la théorie classique :

1. Elle doit avoir un « écart de masse », c'est-à-dire qu'il doit exister une constante $\Delta > 0$ telle que toute excitation du vide ait une énergie au moins égale à Δ.
2. Elle doit avoir un « confinement des quarks », c'est-à-dire que même si la théorie est décrite en termes de champs élémentaires, comme les champs de quarks qui se transforment de manière non triviale par SU(3), les états des particules physiques telles que le proton, le neutron et le pion sont invariants par rapport à SU(3).
3. Elle doit présenter une « brisure de symétrie chirale », ce qui signifie que le vide n'est potentiellement invariant (dans la limite où les masses des quarks s'annulent) que sous l'effet d'un certain sous-groupe du groupe de symétrie complet qui agit sur les champs de quarks.

Le premier point est nécessaire pour expliquer pourquoi la force nucléaire est forte mais courte. Le deuxième est nécessaire pour expliquer pourquoi l'on ne voit jamais de quarks individuels. Le troisième est nécessaire pour expliquer la théorie de l'« algèbre des courants » des pions mous qui a été développée dans les années soixante.

L'expérience — la chromodynamique quantique a connu de nombreux succès dans la confrontation avec l'expérience — et les simulations informatiques (voir par exemple [8]) réalisées depuis la fin des années soixante-dix ont fortement suggéré que la chromodynamique quantique possédait effectivement les propriétés citées ci-dessus. Ces propriétés peuvent être observées dans une certaine mesure dans les cal-

culs théoriques effectués dans plusieurs modèles très simplifiés (comme la théorie de jauge sur réseau fortement couplée, voir par exemple [48]). Mais elles ne sont pas entièrement comprises sur le plan théorique. Il n'existe pas de calcul théorique convaincant, qu'il soit ou non mathématiquement complet, démontrant l'une des trois propriétés dans la chromodynamique quantique, contrairement à une troncature très simplifiée de celle-ci.

En quête de compréhension mathématique

En examinant la physique des théories de jauge dans la dernière section, on a considéré à la fois les propriétés classiques telles que le mécanisme de Higgs pour la théorie électrofaible et les propriétés quantiques qui n'ont pas d'analogues classiques telles que l'écart de masse et le confinement pour la chromodynamique quantique. Les propriétés classiques des théories de jauge sont à la portée des méthodes mathématiques connues et les théories de jauge classiques non abéliennes ont de fait joué un rôle très important en mathématiques au cours des vingt dernières années, en particulier dans l'étude des variétés à trois ou quatre dimensions. En revanche, on ne dispose pas encore d'exemple mathématiquement complet de théorie quantique de jauge dans un espace-temps à quatre dimensions, ni même d'une définition précise de la théorie quantique de jauge à quatre dimensions. Cela changera-t-il au cours du XXI^{e} siècle ? On l'espère !

Certaines structures mathématiques importantes sont apparues pour la première fois en physique avant que leur importance mathématique ne soit pleinement reconnue. C'est ce qui s'est passé avec la découverte du calcul infinitésimal, nécessaire au développement de la mécanique newtonienne, avec l'analyse fonctionnelle et la théorie des représentations de groupes, sujets dont l'importance est devenue plus claire avec la mécanique quantique, et même avec l'étude de la géométrie riemannienne, dont le développement s'est considérablement intensifié lorsqu'il est devenu évident, grâce à l'invention de la relativité générale par Einstein pour décrire la gravité, que ce sujet jouait un rôle dans la description de la nature. Ces domaines des mathématiques ne sont devenus largement accessibles qu'après un temps considérable au cours duquel les idées ont été digérées, simplifiées et intégrées dans la culture mathématique générale.

La théorie quantique des champs a pris une place de plus en plus importante dans la physique tout au long du XX^{e} siècle. Il y a des rai-

sons de penser qu'elle pourrait être importante pour les mathématiques du XXIe siècle. En effet, de nombreux sujets mathématiques qui ont été activement étudiés au cours des dernières décennies semblent avoir des formulations naturelles au moins à un niveau heuristique en termes de théorie quantique des champs. De nouvelles structures reliant les probabilités, l'analyse, l'algèbre et la géométrie sont apparues, pour lesquelles un cadre mathématique général n'en est encore qu'à ses balbutiements.

Du côté analytique, un sous-produit des démonstrations d'existence et de la construction mathématique de certaines théories quantiques des champs a été la construction de nouvelles sortes de mesures, en particulier des mesures non gaussiennes avec invariance euclidienne sur des espaces de fonctionnelles généralisées. Les champs de Dirac et les champs de jauge nécessitent des mesures sur des espaces de fonctions qui prennent des valeurs dans une algèbre de Grassmann et sur des espaces de fonctions à valeurs dans d'autres géométries.

La théorie de la renormalisation provient de la physique de la théorie quantique des champs et fournit une base pour l'étude mathématique des singularités locales (régularité ultraviolette) et de la décroissance globale (régularité infrarouge) dans les théories quantiques des champs. La liberté asymptotique assure une régularité décisive dans le cas où les inégalités de Sobolev classiques sont à la limite. De manière surprenante, les idées de la théorie de la renormalisation s'appliquent également à d'autres domaines des mathématiques, y compris aux travaux classiques sur la convergence des séries de Fourier et aux progrès récents sur les systèmes dynamiques classiques.

Du côté algébrique, les recherches sur les modèles résolubles de la théorie quantique des champs et de la mécanique statistique ont conduit à de nombreuses découvertes sur des sujets tels que l'équation de Yang-Baxter, les groupes quantiques, l'équivalence de Bose-Fermi en dimension 2 et la théorie conforme des champs rationnelle.

La géométrie regorge de nouvelles structures mathématiques enracinées dans la théorie quantique des champs, dont beaucoup ont été activement étudiées au cours des vingt dernières années. Parmi les exemples, on peut citer la théorie de Donaldson des variétés de dimension 4, le polynôme de Jones des nœuds et ses généralisations, la symétrie miroir de variétés complexes, la cohomologie elliptique et la symétrie $SL(2, \mathbb{Z})$ dans la théorie des algèbres affines de Kac-Moody.

La théorie quantique des champs a suggéré dans certains cas de nouvelles perspectives sur des problèmes mathématiques, tandis que son intérêt mathématique dans d'autres cas est d'être jusqu'à aujourd'hui

une source de motivation. Dans le cas des exemples géométriques cités ci-dessus, on ne dispose pas encore d'une définition mathématique des théories quantiques des champs pertinentes (ou d'une définition qui permette de justifier les techniques physiques pertinentes). Les théorèmes d'existence qui placent les théories quantiques des champs sur des bases mathématiques solides sont nécessaires pour que les applications géométriques des théories quantiques des champs deviennent une partie intégrante des mathématiques. Quel que soit le rôle futur de la théorie quantique des champs dans les mathématiques pures, c'est un grand défi pour les mathématiciens de comprendre les principes physiques qui ont été si importants et productifs tout au long du XX^{e} siècle.

Enfin, la théorie quantique des champs est le point de départ d'une quête qui pourrait s'avérer centrale dans la physique du XXI^{e} siècle : l'effort d'unification de la gravité et de la mécanique quantique, peut-être dans le cadre de la théorie des cordes. Pour que les mathématiciens puissent participer à cette quête ou même comprendre les résultats possibles, la théorie quantique des champs doit être développée davantage en tant que branche des mathématiques. Il est important non seulement de comprendre la solution de problèmes particuliers issus de la physique, mais aussi de placer ces résultats dans un nouveau cadre mathématique. On espère que ce cadre permettra un développement unifié de plusieurs domaines des mathématiques et de la physique, et qu'il fournira également un terrain pour le développement de nouvelles mathématiques et d'une nouvelle physique.

C'est pourquoi le conseil scientifique de l'Institut de mathématiques Clay a choisi un problème du millénaire sur les théories de jauge quantiques. Pour résoudre ce problème, il faut à la fois comprendre l'un des grands mystères non résolus de la physique, à savoir l'existence d'un écart de masse, et produire un exemple mathématiquement complet de théorie quantique des champs de jauge dans un espace-temps à quatre dimensions.

Les champs quantiques

Un champ quantique, ou opérateur local de champ quantique, est une fonction généralisée sur l'espace-temps à valeurs dans un espace d'opérateurs qui obéit à certains axiomes. Les propriétés requises pour les champs quantiques sont décrites à un niveau de précision physique dans de nombreux manuels (voir par exemple [27]). Gårding et Wightman [45] ont proposé des axiomes mathématiquement précis pour les

théories quantiques des champs sur $\mathbb{R}^4$ avec une signature de Minkowski, tandis que Haag et Kastler ont introduit un schéma apparenté pour des fonctions locales du champ [24].

Fondamentalement, il faut que l'espace de Hilbert $\mathcal{H}$ du champ quantique porte une représentation du groupe de Poincaré, également appelé « groupe de Lorentz inhomogène ». Le hamiltonien H et la quantité de mouvement $\vec{\mathrm{P}}$ sont les éléments auto-adjoints de l'algèbre de Lie du groupe qui engendrent des translations dans le temps et l'espace. Un vecteur du vide est un élément de $\mathcal{H}$ qui est invariant par la représentation du groupe de Poincaré. On suppose que la représentation a une énergie positive, $0 \leqslant \mathrm{H}$, et un vecteur du vide $\Omega \in \mathcal{H}$ qui est unique à une phase près. Les fonctions invariantes de jauge des champs quantiques agissent également comme des transformations linéaires sur $\mathcal{H}$ et se transforment de manière covariante par le groupe de Poincaré. Les champs quantiques situés dans des régions de l'espace-temps qui ne peuvent pas être jointes par un signal lumineux doivent être indépendants. Gårding et Wightman ont formulé l'indépendance comme la commutativité des opérateurs de champ (anti-commutativité pour deux champs fermioniques).

Un des succès de la théorie quantique des champs axiomatique du XX$^{\mathrm{e}}$ siècle a été la découverte de la manière de convertir une théorie des champs avec invariance euclidienne sur un espace-temps euclidien en une théorie des champs avec invariance lorentzienne sur l'espace-temps de Minkowski, et *vice versa*. Wightman a utilisé l'énergie positive pour établir le prolongement analytique des espérances des théories des champs de Minkowski à l'espace euclidien. Kurt Symanzik a interprété les espérances euclidiennes comme un ensemble statistique de champs de Markov classiques [46], avec une densité de probabilité proportionnelle à $\exp(-\mathrm{S})$, où S désigne la fonctionnelle d'action euclidienne. E. Nelson a reformulé l'image de Symanzik et a montré que l'on pouvait également construire un espace de Hilbert et un champ quantique à partir d'un champ de Markov [33]. Osterwalder et Schrader ont ensuite découvert la condition élémentaire de « réflexion-positivité » pour remplacer la propriété de Markov. Cela a donné lieu à une théorie générale qui établit l'équivalence entre les schémas d'axiomes lorentziens et euclidiens [35]. Voir également [13].

On espère que la poursuite de l'exploration mathématique de la théorie quantique des champs conduira à des raffinements des ensembles d'axiomes utilisés jusqu'à présent, peut-être pour incorporer des propriétés considérées comme importantes par les physiciens telles que

l'existence d'un développement en produit d'opérateurs ou d'un tenseur énergie-impulsion local.

Le problème

Pour démontrer l'existence d'une théorie quantique de jauge en dimension 4 avec le groupe de jauge G, il faut définir une théorie quantique des champs (au sens ci-dessus) avec des opérateurs quantiques locaux en correspondance avec les polynômes locaux invariants de jauge de la courbure F et de ses dérivées covariantes, tels que $\operatorname{Tr} \mathrm{F}_{ij}\mathrm{F}_{kl}(x)$ [1]. Les fonctions de corrélation des opérateurs de champ quantique doivent s'accorder à courte distance avec les prédictions de la liberté asymptotique et de la théorie perturbative de renormalisation telles qu'elles sont décrites dans les manuels. Ces prédictions incluent entre autres l'existence d'un tenseur des contraintes et d'un développement en produit des opérateurs avec des singularités locales prescrites qui sont prédites par la liberté asymptotique.

Le vecteur du vide Ω ayant une invariance de Poincaré, il s'agit d'un état propre d'énergie nulle : $\mathrm{H}\Omega = 0$. L'axiome de l'énergie positive affirme que dans toute théorie quantique des champs, le support du spectre de H est dans l'intervalle $[0, \infty[$. Une théorie quantique des champs a un écart de masse Δ si H n'a pas de spectre dans l'intervalle $]0, \Delta[$ pour un certain $\Delta > 0$. La borne supérieure de ces Δ est la masse m, et l'on exige que $m < \infty$.

Existence d'une théorie de Yang-Mills et écart de masse. *Démontrer que pour tout groupe de jauge simple et compact* G*, une théorie de Yang-Mills quantique non triviale existe sur* $\mathbb{R}^4$ *et a un écart de masse* $\Delta > 0$*. L'existence inclut l'établissement de propriétés axiomatiques au moins aussi fortes que celles citées dans [35, 45].*

Commentaires

Une conséquence importante de l'existence d'un écart de masse est la formation d'amas. Soit $\vec{x} \in \mathbb{R}^3$ un point dans l'espace. Soient H

1. Il n'existe pas de bijection naturelle entre ces « polynômes différentiels » classiques et les opérateurs quantifiés, car la correspondance comporte des subtilités standard impliquant une renormalisation [27]. On s'attend à ce que l'espace des polynômes différentiels classiques de dimension $\leqslant d$ corresponde à l'espace des opérateurs quantiques locaux de dimension $\leqslant d$.

et $\vec{\mathrm{P}}$ l'énergie et la quantité de mouvement, les générateurs associés à la translation en temps et en espace. Pour toute constante strictement positive $\mathrm{C} < \Delta$ et pour tout opérateur de champ quantique local $\mathcal{O}(\vec{x}) = e^{-i\vec{\mathrm{P}}\cdot\vec{x}}\mathcal{O}e^{i\vec{\mathrm{P}}\cdot\vec{x}}$ tel que $\langle\Omega, \mathcal{O}\Omega\rangle = 0$, on a

$$|\langle\Omega, \mathcal{O}(\vec{x})\mathcal{O}(\vec{y})\Omega\rangle| \leqslant \exp(-\mathrm{C}|\vec{x} - \vec{y}|) \tag{2}$$

tant que $|\vec{x} - \vec{y}|$ est suffisamment grand. La formation d'amas est une propriété de localité qui en gros peut permettre d'appliquer des résultats mathématiques établis sur $\mathbb{R}^4$ à n'importe quelle variété de dimension 4, comme cela a été expliqué à un niveau heuristique (pour une extension supersymétrique de la théorie de jauge quadridimensionnelle) dans [49]. Ainsi, l'écart de masse n'a pas seulement une signification physique (comme expliqué dans l'introduction), mais il peut aussi être important dans les applications mathématiques des théories quantiques de jauge quadridimensionnelles à la géométrie. En outre, l'existence d'un écart uniforme pour les approximations à volume fini peut jouer un rôle fondamental dans la démonstration de l'existence de la limite à volume infini.

Il existe de nombreuses extensions naturelles du problème du millénaire. Entre autres, on aimerait démontrer l'existence d'un état isolé à une particule (un écart supérieur, en plus de l'écart de masse), démontrer le confinement, démontrer l'existence d'autres théories de jauge quadridimensionnelles (incorporant des champs supplémentaires qui préservent la liberté asymptotique), comprendre les questions dynamiques (telles que l'écart de masse possible, le confinement et la brisure de symétrie chirale) dans ces théories plus générales, et étendre les théorèmes d'existence de $\mathbb{R}^4$ à une variété de dimension 4 arbitraire.

Mais une solution au problème de l'existence et de l'écart de masse, comme indiqué ci-dessus, constituerait un tournant dans la compréhension mathématique de la théorie quantique des champs, avec la possibilité d'ouvrir de nouveaux horizons pour ses applications.

Une perspective mathématique

Depuis une cinquantaine d'années, Wightman et d'autres se demandent s'il existe des exemples mathématiquement bien définis de théories quantiques des champs relativistes et non linéaires. On a maintenant une réponse partielle : de nombreux résultats sur l'existence et les propriétés physiques de théories quantiques des champs non linéaires

ont été démontrés grâce à l'émergence d'un ensemble de travaux connus sous le nom de « théorie quantique des champs constructive ».

Les réponses sont partielles, car dans la plupart de ces théories des champs, on remplace l'espace-temps de Minkowski $\mathbb{M}^4$ par un espace-temps de dimension inférieure $\mathbb{M}^2$ ou $\mathbb{M}^3$, ou par une approximation compacte comme un tore. De manière équivalente, dans la formulation euclidienne, on remplace l'espace-temps euclidien $\mathbb{R}^4$ par $\mathbb{R}^2$ ou $\mathbb{R}^3$. Certains résultats sont connus pour la théorie de Yang-Mills sur un 4-tore $\mathbb{T}^4$ approximant $\mathbb{R}^4$. Bien que la construction ne soit pas complète, il y a de nombreuses indications que les méthodes connues pourraient être étendues pour construire une théorie de Yang-Mills sur $\mathbb{T}^4$.

En fait, à l'heure actuelle, on ne connaît aucune théorie des champs relativiste non triviale qui satisfasse aux axiomes de Wightman (ou à tout autre système d'axiomes raisonnable) en dimension 4. Le simple fait de disposer d'une construction mathématique détaillée de la théorie de Yang-Mills sur un espace compact représenterait donc une avancée majeure. Pourtant, même si l'on y parvenait, aucune idée actuelle ne permet d'établir l'existence d'un écart de masse uniforme dans le volume. Les méthodes actuelles ne suggèrent pas non plus comment obtenir l'existence de la limite en volume infini $\mathbb{T}^4 \to \mathbb{R}^4$.

Les méthodes

Depuis le début de la théorie quantique des champs, deux méthodes principales sont apparues pour démontrer l'existence de champs quantiques dans un espace de configuration non compact (tel que l'espace de Minkowski), à savoir

— trouver une solution analytique exacte ;
— résoudre une suite de problèmes approchés et établir la convergence de ces solutions vers la limite souhaitée.

Des solutions exactes peuvent se présenter pour les champs non linéaires pour des valeurs spéciales de la constante de couplage qui conduisent à des symétries supplémentaires ou à des modèles intégrables. On peut les obtenir avec d'astucieux changements de variables. Dans le cas de la théorie de Yang-Mills en dimension 4, une solution exacte semble improbable, bien qu'il puisse y avoir un jour une solution asymptotique dans la limite où N est grand.

La seconde méthode consiste à utiliser des approximations mathématiques pour montrer la convergence des solutions approchées vers les solutions exactes des problèmes non linéaires, ce que l'on appelle la

« théorie quantique des champs constructive ». Deux approches principales, l'étude de la théorie quantique sur l'espace de Hilbert et l'étude des intégrales fonctionnelles classiques, sont liées par la construction d'Osterwalder-Schrader. L'établissement d'estimations uniformes *a priori* est un élément central de la théorie quantique des champs constructive. Il y a trois méthodes pour obtenir des estimations :

(a) les inégalités de corrélation,

(b) les symétries de l'interaction,

(c) les développements convergents.

Les méthodes d'inégalité de corrélation ont l'avantage d'être générales. Mais les inégalités de corrélation reposent sur des propriétés spéciales de l'interaction qui ne s'appliquent souvent qu'aux bosons scalaires ou aux théories de jauge abéliennes. L'utilisation de la symétrie ne s'applique également qu'à des valeurs particulières des constantes de couplage ; elle est généralement combinée à une autre méthode pour obtenir un contrôle analytique. Dans la plupart des exemples connus, les séries de perturbations, c'est-à-dire les séries de puissances de la constante de couplage, sont des développements divergents ; même la sommation de Borel et les autres méthodes de sommation ont une applicabilité limitée.

Cela a conduit à des méthodes de développement généralement connues sous le nom de « développements en amas ». Chaque terme d'un développement en amas dépend des constantes de couplage d'une manière compliquée. Les termes apparaissent souvent comme des intégrales fonctionnelles. Il faut une connaissance quantitative suffisante des propriétés de chaque terme du développement pour garantir la convergence de la somme et établir ses propriétés qualitatives. Des estimations raffinées donnent le taux de décroissance exponentielle des fonctions de Green, les masses, l'existence d'une brisure de symétrie (ou sa préservation), etc.

Au cours des trente dernières années, une panoplie de méthodes de développement a émergé comme un outil central pour établir des résultats mathématiques dans la théorie quantique des champs constructive. Dans leurs diverses incarnations, ces développements englobent des idées sur la nature asymptotique de la théorie des perturbations, de la localisation dans l'espace-temps, de la localisation dans l'espace des phases, de la théorie de la renormalisation, des approximations semi-classiques (y compris les effets « non perturbatifs ») et de la brisure de symétrie. On trouvera une introduction à bon nombre de ces

méthodes et des références dans [18], ainsi qu'un examen plus récent des résultats dans [28]. Ces méthodes de développement peuvent être compliquées et la littérature est immense. On ne peut donc qu'espérer présenter au lecteur quelques idées. On présente d'avance nos excuses pour les omissions importantes.

Premiers exemples : les champs scalaires

Depuis les années quarante, le champ quantique de Klein-Gordon φ fournit un exemple de théorie des champs linéaire, scalaire, de masse m (découlant d'un potentiel quadratique). Ce champ, ainsi que le champ spinoriel libre de Dirac qui lui est associé, ont servi de modèles pour la formulation des premiers schémas d'axiomes dans les années cinquante [45].

Les moments d'une mesure de Borel $d\mu$ avec invariance euclidienne, réflexion-positive et ergodique sur l'espace $\mathcal{S}'\left(\mathbb{R}^d\right)$ des distributions tempérées peuvent vérifier les axiomes d'Osterwalder-Schrader. La mesure gaussienne $d\mu$ de moyenne nulle et de covariance $\mathrm{C} = \left(-\Delta + m_0^2\right)^{-1}$ donne le champ libre de masse m_0. Mais on exige que la mesure non gaussienne $d\mu$ donne des champs non linéaires. Pour la mesure gaussienne, la positivité par réflexion est équivalente à la positivité de la transformation $\Theta\mathrm{C}$, restreinte à $\mathrm{L}^2\left(\mathbb{R}_+^d\right) \subset \mathrm{L}^2\left(\mathbb{R}^d\right)$. Ici, $\Theta : t \to -t$ désigne l'opérateur de réflexion temporelle et $\mathbb{R}_+^d = \{(t, \vec{x}) : t \geqslant 0\}$ est le sous-espace des temps positifs.

La première démonstration que des champs non linéaires vérifient les axiomes de Wightman et la première construction de telles mesures non gaussiennes ne sont apparues que dans les années soixante-dix. Les premiers exemples concernaient des champs avec de petites nonlinéarités polynomiales sur $\mathbb{R}^2$: d'abord en volume fini, puis dans la limite du volume infini [15, 19, 22]. Ces théories des champs obéissent aux axiomes de Wightman (ainsi qu'à toutes les autres formulations axiomatiques), les champs décrivent des particules avec une masse bien définie, et les champs produisent des états à plusieurs particules avec une diffusion non triviale [19]. La matrice de diffusion peut être développée comme une série asymptotique par rapport aux constantes de couplage, et les résultats concordent terme à terme avec la description standard de la diffusion dans la théorie des perturbations que l'on trouve dans les textes de physique [37].

Il existe aussi une théorie quantique des champs de Wightman quartique sur $\mathbb{R}^3$. Elle s'obtient en construisant une mesure non gaussienne

remarquable $d\mu$ sur $\mathcal{S}'\left(\mathbb{R}^3\right)$ [10,16]. Cela mérite une étude plus approfondie.

On se concentre maintenant sur certaines propriétés de la perturbation la plus simple de la densité d'action du champ libre, à savoir le polynôme quartique pair

$$\lambda\varphi^4 + \frac{1}{2}\sigma\varphi^2 + c \tag{3}$$

Les coefficients $0 < \lambda$, σ et $c \in \mathbb{R}$ sont des paramètres réels qui sont tous nuls pour le champ libre. Pour $0 < \lambda \ll 1$, on peut choisir $\sigma(\lambda)$ et $c(\lambda)$ de sorte que la théorie des champs décrite par (3) existe, soit unique et ait une masse égale à m telle que $|m - m_0|$ soit petit.

En raison de la singularité locale du champ non linéaire, il faut d'abord éliminer l'interaction. La méthode la plus simple consiste à tronquer le développement de Fourier du champ φ dans (3) par $\varphi_\kappa(x) = \int_{|k|\leqslant\kappa} \widetilde{\varphi}(k)e^{-ikx}dk$ et à compactifier le volume spatial de la perturbation en $\mathcal{V}$. On obtient la théorie quantique des champs souhaitée comme limite des approximations tronquées. Les constantes σ et c ont la forme $\sigma = \alpha\lambda + \beta\lambda^2$ et $c = \gamma\lambda + \delta\lambda^2 + \epsilon\lambda^3$. On souhaite que les espérances des produits de champs aient une limite quand $\kappa \to \infty$. On choisit α et γ (en dimension 2) et l'on choisit tous les coefficients $\alpha, \beta, \gamma, \delta$ et ϵ (en dimension 3) pour qu'ils dépendent de κ de la manière suggérée par la théorie des perturbations. On démontre alors que les espérances convergent quand $\kappa \to \infty$, même si les constantes spécifiées α, ..., divergent. Ces constantes fournissent la renormalisation nécessaire de l'interaction. Dans le cas tridimensionnel, il faut également normaliser les vecteurs dans l'espace de Fock avec une constante qui diverge avec κ. On appelle cette constante une constante de renormalisation de la fonction d'onde.

L'opérateur « masse » en unités naturelles est $\mathrm{M} = \sqrt{\mathrm{H}^2 - \vec{\mathrm{P}}^2} \geqslant 0$. Le vecteur du vide Ω est tel que $\mathrm{M}\Omega = 0$. Les états massifs à une seule particule sont des vecteurs propres associés à une valeur propre $m > 0$. Si m est une valeur propre isolée de M, on déduit des axiomes de Wightman et de la théorie de la diffusion de Haag-Ruelle qu'il existe des états de diffusion asymptotiques avec un nombre arbitraire de particules (voir [18,24]).

Le problème fondamental de la « complétude asymptotique » est la question de savoir si ces états asymptotiques (y compris les états liés possibles) engendrent $\mathcal{H}$. Au cours des trente dernières années, plusieurs nouvelles méthodes sont apparues qui ont permis de démontrer la com-

plétude asymptotique dans la théorie de la diffusion pour la mécanique quantique non relativiste. Cela permet d'espérer que l'on puisse maintenant s'attaquer au problème ouvert de la complétude asymptotique pour tout exemple connu de théorie quantique des champs non linéaire.

Contrairement à l'existence de champs quantiques avec une non-linéarité φ^4 en dimensions 2 et 3, la question de la généralisation de ces résultats à quatre dimensions est problématique. On sait que les champs scalaires auto-interagissant avec une non-linéarité quartique n'existent pas en dimension 5 ou plus [1, 12] (les démonstrations s'appliquent aux théories de champ avec un seul champ scalaire). L'analyse de la dimension 4, qui marque la limite entre l'existence et la non-existence, est plus subtile. Si l'on fait quelques hypothèses raisonnables (mais pas entièrement démontrées), on peut également conclure à la trivialité du couplage quartique en dimension 4. Non seulement il s'agit d'une démonstration convaincante, mais en plus, le couplage quartique n'a pas la propriété de liberté asymptotique en dimension 4. Ainsi, tous les enseignements tirés des marches aléatoires, de la théorie des perturbations et de l'analyse de renormalisation pointent vers la trivialité de l'interaction quartique en dimension 4.

D'autres interactions polynomiales en dimension 4 sont encore plus problématiques : l'énergie classique de l'interaction cubique n'est pas bornée inférieurement, ce qui en fait un candidat improbable pour une théorie quantique où la positivité de l'énergie est un axiome. De plus, les interactions polynomiales de degré supérieur à l'interaction quartique sont plus singulières dans la théorie des perturbations que l'interaction quartique.

Toutes ces raisons complètent l'importance physique et géométrique de la théorie de Yang-Mills et en font un candidat naturel pour la théorie quantique des champs constructive en dimension 4.

Une constante de couplage grande

En deux dimensions, la théorie des champs avec la densité d'énergie (3) existe pour tout λ strictement positif. Pour $0 \leqslant \lambda \ll 1$, la solution est unique sous diverses conditions. Mais pour $\lambda \gg 1$, il existe deux solutions différentes, chacune caractérisée par son état fondamental ou sa « phase ». La solution dans chaque phase vérifie les axiomes d'Osterwalder-Schrader et de Wightman avec un écart de masse non nul et un état du vide unique avec invariance de Poincaré. Les solutions distinctes apparaissent comme une bifurcation d'une solution approchée

unique avec un volume fini $\mathcal{V}$ quand $\mathcal{V} \to \infty$.

On peut observer ce comportement en réordonnant et en mettant à l'échelle l'interaction $\lambda\varphi^4$ (3) avec $\lambda \gg 1$ pour obtenir un double puits de potentiel équivalent de la forme

$$\lambda\left(\varphi^2 - \frac{1}{\lambda}\right)^2 + \frac{1}{2}\sigma\varphi^2 + c \tag{4}$$

Ici, $\lambda \ll 1$ est une nouvelle constante de couplage et les constantes de renormalisation σ et c sont quelque peu différentes de celles mentionnées ci-dessus. Les deux solutions pour un λ donné sont liées par la brisure de symétrie $\varphi \to -\varphi$ de l'interaction (4). La démonstration de ces faits repose sur un développement en amas convergent autour de chaque minimum du potentiel provenant de (4) et sur la démonstration que la probabilité d'un effet tunnel entre les deux solutions est faible [20].

Une valeur critique λ_c de λ dans (3) sépare l'unicité de la solution (pour $\lambda < \lambda_c$) de l'existence d'une transition de phase (pour $\lambda > \lambda_c$). Lorsque λ croît jusqu'à λ_c, l'écart de masse $m = m(\lambda)$ décroît de façon monotone et continue jusqu'à zéro [17, 23, 32].

Le comportement détaillé de la théorie des champs (ou de la masse) au voisinage de $\lambda = \lambda_c$ est extraordinairement difficile à analyser. Pratiquement rien n'a été démontré. Les physiciens ont une image qualitative basée sur l'hypothèse d'un comportement en loi de puissance fractionnaire $m(\lambda) \sim |\lambda_c - \lambda|^\nu$ au-dessus ou en dessous du point critique, où l'exposant ν dépend de la dimension. On s'attend également à ce que la constante de couplage critique λ_c corresponde à la plus grande force physique entre les particules, et que ces théories critiques soient proches des limites d'échelle des modèles de type Ising en physique statistique. On s'attend à ce qu'une meilleure compréhension de ces idées débouche sur de nouveaux outils de calcul pour les champs quantiques et la physique statistique.

On comprend partiellement un cas multiphase plus général. On peut trouver un nombre arbitraire n de phases en faisant un bon choix pour la densité d'énergie polynomiale $\mathcal{P}_n(\varphi)$ avec n minima. Il est intéressant d'étudier la perturbation d'un tel polynôme fixé $\mathcal{P}_n$ par des polynômes $\mathcal{Q}$ de degré inférieur et à petits coefficients. Parmi ces perturbations, on peut trouver des familles de polynômes $\mathcal{Q}(\varphi)$ qui produisent des théories des champs avec exactement $n' \leqslant n$ phases [26].

Interactions de Yukawa et théorie de jauge abélienne

L'existence d'interactions boson-fermion est également connue en deux dimensions, et des résultats partiels existent en trois dimensions. En deux dimensions, les interactions de Yukawa de la forme $\bar{\psi}\psi\varphi$ existent avec une renormalisation appropriée, ainsi que leurs généralisations de la forme $\mathcal{P}(\varphi) + \bar{\psi}\psi\mathcal{Q}''(\varphi)$, voir [18, 42]. Le cas supersymétrique $\mathcal{P} = |\mathcal{Q}'|^2$ nécessite une attention particulière pour traiter les éliminations de divergences (voir [28] pour les références).

Un modèle de Higgs bidimensionnel et continu décrit un champ de jauge abélien interagissant avec un champ scalaire chargé. Brydges, Fröhlich et Seiler ont construit cette théorie et montré qu'elle vérifie les axiomes d'Osterwalder-Schrader [7], ce qui fournit le seul exemple complet d'une théorie de jauge en interaction qui vérifie les axiomes. Il existe un écart de masse dans ce modèle [4]. L'extension de toutes ces conclusions à un modèle de Higgs non abélien, même en deux dimensions, représenterait une avancée significative.

Il y a des résultats partiels sur l'interaction tridimensionnelle $\bar{\psi}\psi\varphi$ (voir [30]) ainsi que pour d'autres interactions plus singulières [14].

La théorie de Yang-Mills

Une grande partie des progrès mathématiques examinés ci-dessus résulte de la compréhension de l'intégration fonctionnelle et de l'utilisation de ces méthodes pour construire des théories des champs euclidiens. L'intégration fonctionnelle pour les théories de jauge soulève de nouveaux problèmes techniques qui tournent autour du riche groupe de symétries, en particulier la symétrie de jauge. Le choix de la jauge et la transformation entre différents choix compliquent la structure mathématique. La symétrie de jauge offre cependant la possibilité d'une liberté asymptotique. Certaines idées et propositions dans la littérature physique [5, 9] ont conduit à un cadre étendu. Mais les implications de ces idées pour une construction mathématique de la théorie de Yang-Mills doivent être mieux comprises.

Wilson a proposé une approche différente basée sur l'approximation de l'espace-temps continu par un réseau, sur lequel il a défini une action avec invariance de jauge [48]. Avec un groupe de jauge compact et un espace-temps compactifié, l'approximation par le réseau réduit l'intégration fonctionnelle à une intégrale en dimension finie. Il faut alors vérifier l'existence de limites des espérances appropriées d'observables

avec invariance de jauge lorsque la maille du réseau tend vers zéro et lorsque le volume tend vers l'infini.

La positivité par réflexion est valable pour l'approximation de Wilson [36], ce qui constitue un avantage majeur. Il existe peu de méthodes pour retrouver la réflexion-positivité au cas où elle serait perdue par la régularisation, comme c'est le cas avec la régularisation dimensionnelle, la régularisation de Pauli-Villars et bien d'autres méthodes. La mise en place d'un espace de Hilbert quantique fait partie de la solution à ce problème du millénaire.

Balaban a étudié ce programme dans un réseau tridimensionnel avec des conditions aux limites périodiques approximant un tore spatio-temporel [2]. Il a étudié les transformations de renormalisation (intégration des degrés de liberté avec une grande impulsion suivie d'une mise à l'échelle) et a établi de nombreuses propriétés intéressantes de l'action effective qu'elles produisent. Ces estimations sont uniformes par rapport à la maille du réseau lorsque sa largeur tend vers zéro. Le choix des jauges est central dans ce travail, ainsi que l'utilisation des normes de l'espace de Sobolev pour intégrer une analyse des effets géométriques.

On définit ces jauges dans des cellules de phase : les choix varient localement dans l'espace-temps ainsi qu'à différentes échelles de longueur. Les choix évoluent par récurrence au fur et à mesure des transformations de renormalisation, des jauges adaptées à la régularité locale (jauges ultra-violettes) à celles adaptées aux distances macroscopiques (jauges infrarouges). Il s'agit d'une étape importante vers la démonstration de l'existence de la limite continue dans un espace-temps compactifié. Ces résultats doivent être généralisés à l'étude des espérances des fonctions des champs avec invariance de jauge.

Bien que ces travaux en dimension 3 soient importants en soi, une percée qualitative a été réalisée avec la généralisation de cette analyse en dimension 4 par Balaban [3]. Cela inclut une analyse de la liberté asymptotique pour contrôler le flot du groupe de renormalisation ainsi que l'obtention d'estimations quantitatives sur les effets découlant des grandes valeurs du champ de jauge.

Des travaux approfondis ont également été réalisés sur la régularisation continue de l'interaction de Yang-Mills. Cette régularisation pourrait permettre une meilleure compréhension [29, 39].

Ces étapes vers la compréhension de la théorie de Yang-Mills quantique conduisent à une perspective de généralisation des méthodes actuelles pour établir une construction complète de la théorie quan-

tique des champs de Yang-Mills sur un espace-temps compact à quatre dimensions. Il est vraisemblablement nécessaire de revoir les résultats connus à un niveau profond, de simplifier les méthodes et de les généraliser.

De nouvelles idées sont nécessaires pour démontrer l'existence d'un écart de masse uniforme dans le volume de l'espace-temps. Un tel résultat permettrait vraisemblablement d'étudier la limite quand $\mathbb{T}^4 \to \mathbb{R}^4$ †.

Remarques complémentaires

Comme la théorie de jauge en dimension 4 est une théorie dans laquelle l'écart de masse n'est pas classiquement visible, sa mise en évidence peut nécessiter un changement non classique de variables ou une « transformation par dualité ». Par exemple, la dualité a été utilisée pour établir un écart de masse dans le problème du gaz de Coulomb en mécanique statistique, où le phénomène est connu sous le nom d'écrantage de Debye : les charges d'essai macroscopiques dans un gaz de Coulomb neutre subissent une force mutuelle qui décroît exponentiellement avec la distance. La démonstration mathématique de ce phénomène d'écrantage passe par l'identité de la fonction de partition du gaz de Coulomb avec celle d'une théorie des champs $\cos(\lambda\varphi)$ (sinus-Gordon), et par le potentiel parabolique d'approximation au voisinage d'un minimum de ce potentiel (voir [6]).

Un point de vue sur l'écart de masse dans la théorie de Yang-Mills suggère qu'il pourrait provenir du potentiel quartique $(\mathrm{A} \wedge \mathrm{A})^2$ dans l'action, où $\mathrm{F} = d\mathrm{A} + g\mathrm{A} \wedge \mathrm{A}$ (voir [11]) et pourrait être lié à la courbure dans l'espace des connexions (voir [44]). Bien que l'action de Yang-Mills ait des directions plates, certains problèmes de mécanique quantique avec des potentiels impliquant des directions plates (directions pour lesquelles le potentiel reste borné quand $|x| \to \infty$) conduisent à des états bornés [43].

Une idée importante sur la dualité qui pourrait éclairer les propriétés dynamiques de la théorie de jauge en dimension 4 implique le développement en 1/N [25]. On pense que la théorie quantique de jauge en dimension 4 avec le groupe de jauge SU(N) (ou SO(N), ou Sp(N)) pourrait être équivalente à une théorie des cordes avec 1/N comme

†. On exclut en particulier l'existence faible (compacité) comme solution à la partie existence du problème du millénaire, à moins que l'on n'utilise également d'autres techniques pour établir des propriétés de la limite (telles que l'existence d'un écart de masse et les axiomes).

constante de couplage de la corde. Une telle description pourrait donner une explication claire de l'écart de masse et du confinement et peut-être un bon point de départ pour une démonstration rigoureuse (pour N suffisamment grand). Des progrès surprenants ont été réalisés dans ce sens pour certains systèmes de jauge en dimension 4 fortement couplés avec de la matière [31], mais il n'existe pas encore d'approche efficace de la théorie de jauge sans fermions. Les recherches sur les théories supersymétriques et les théories des cordes ont permis de découvrir diverses autres approches pour comprendre l'écart de masse dans certaines théories de jauge en dimension 4 avec des champs de matière (voir par exemple [41]).

Bibliographie

[1] Aizenman (M.), « Geometric analysis of φ^4 fields and Ising models », *Comm. Math. Phys.*, n° 86, 1982, p. 1-48.

[2] Balaban (T.), « Ultraviolet stability of three-dimensional lattice pure gauge field theories », *Comm. Math. Phys.*, n° 102, 1985, p. 255-275.

[3] Balaban (T.), « Renormalization group approach to lattice gauge field theories. I : generation of effective actions in a small field approximation and a coupling constant renormalization in 4D », *Comm. Math. Phys.*, n° 109, 1987, p. 249-301.

[4] Balaban (T.), Brydges (D.), Imbrie (J.) et Jaffe (A.), « The mass gap for Higgs models on a unit lattice », *Ann. Phys.*, n° 158, 1984, p. 281-319.

[5] Becchi (C.), Rouet (A.) et Stora (R.), « Renormalization of gauge theories », *Ann. Phys.*, n° 98, 1976, p. 287-321.

[6] Brydges (D.)et Federbush (P.), « Debye screening », *Comm. Math. Phys.*, n° 73, 1980, p. 197-246.

[7] Brydges (D.), Fröhlich (J.) et Seiler (E.), « On the construction of quantized gauge fields, I », *Ann. Phys.*, n° 121, 1979, p. 227-284, II : *Comm. Math. Phys.*, n° 71, 1980, p. 159-205, III : *Comm. Math. Phys.*,n° 79, 1981, p. 353-399.

[8] Creutz (M.), « Monte Carlo study of quantized SU(2) gauge theory », *Phys. Rev. D*, n° 21, 1980, p. 2308-2315.

[9] Faddeev (L. D.) et Popov (V. N.), « Feynman diagrams for the Yang-Mills fields », *Phys. Lett. B*, n° 25, 1967, p. 29-30.

[10] Feldman (J.) et Osterwalder (K.), « The Wightman axioms and the mass gap for weakly coupled φ^4_3 quantum field theories », *Ann. Phys.*, n° 97, 1976, p. 80-135.

[11] Feynman (R.), « The quantitative behavior of Yang-Mills theory in $2+1$ dimensions », *Nucl. Phys. B*, n° 188, 1981, p. 479-512.

[12] Fröhlich (J.), « On the triviality of $\lambda\varphi_d^4$ theories and the approach to the critical point », *Nucl. Phys.*, n° 200, 1982, p. 281-296.

[13] Fröhlich (J.), Osterwalder (K.) et Seiler (E.), « On virtual representations of symmetric spaces and theory analytic continuation », *Ann. Math.*, n° 118, 1983, p. 461-489.

[14] Gawedzki (K.), « Renormalization of a non-renormalizable quantum field theory », *Nucl. Phys. B*, n° 262, 1985, p. 33-48.

[15] Glimm (J.) et Jaffe (A.), « The $\lambda\varphi_2^4$ quantum field theory without cut-offs », I : *Phys. Rev.*, n° 176, 1968, p. 1945-1951, II : *Ann. Math.*, n° 91, 1970, p. 362-401, III : *Acta. Math.*, n° 125, 1970, p. 203-267, IV : *J. Math. Phys.*, n° 13, 1972, p. 1568-1584.

[16] Glimm (J.) et Jaffe (A.), « Positivity of the φ_3^4 Hamiltonian », *Fortschr. Phys.*, n° 21, 1973, p. 327-376.

[17] Glimm (J.) et Jaffe (A.), « φ_2^4 quantum field model in the single phase region : differentiability of the mass and bounds on critical exponents », *Phys. Rev.*, n° 10, 1974, p. 536-539.

[18] Glimm (J.) et Jaffe (A.), *Quantum Physics*, 2e édition, Springer, 1987, et *Selected Papers*, vol. I et II, Boston, Birkhäuser, 1985.
Le volume II inclut des réimpressions de [15, 16, 19, 20].

[19] Glimm (J.), Jaffe (A.) et Spencer (T.), « The Wightman axioms and particle structure in the $P(\varphi)_2$ quantum field model », *Ann. Math.*, n° 100, 1974, p. 585-632.

[20] Glimm (J.), Jaffe (A.) et Spencer (T.), « A convergent expansion about mean field theory », I : *Ann. Phys.*, n° 101, 1976, p. 610-630, II : *Ann. Phys.*, n° 101, 1976, p. 631-669.

[21] Gross (D. J.) et Wilczek (F.), « Ultraviolet behavior of non-abelian gauge theories », *Phys. Rev. Lett.*, n° 30, 1973, p. 1343-1346.

[22] Guerra (F.), Rosen (L.) et Simon (B.), « The $P(\varphi)_2$ Euclidean quantum field theory as classical statistical mechanics », *Ann. Math.*, n° 101, 1975, p. 111-259.

[23] Guerra (F.), Rosen (L.) et Simon (B.), « Correlation inequalities and the mass gap in $P(\varphi)_2$, III. Mass gap for a class of strongly coupled theories with non-zero external field », *Comm. Math. Phys.*, n° 41, 1975, p. 19-32.

[24] Haag (R.), *Local Quantum Physics*, Berlin, Springer-Verlag, 1992.

[25] Hooft (G. 'T), « A planar diagram theory for strong interactions », *Nucl. Phys. B*, n° 72, 1974, p. 461-473.

[26] Imbrie (J.), « Phase diagrams and cluster expansions for low temperature $\mathcal{P}(\varphi)_2$ models », *Comm. Math. Phys.*, n° 82, 1981, p. 261-304 et p. 305-343.

[27] Itzykson (C.) et Zuber (J.-B.), *Quantum Field Theory*, New York, McGraw-Hill, 1980.

[28] Jaffe (A.), « Constructive quantum field theory », dans Kibble (T.), *Mathematical Physics*, Singapour, World Scientific, 2000.

[29] Magnen (J.), Rivasseau (V.) et Sénéor (R.), « Construction of YM_4 with an infrared cutoff », *Comm. Math. Phys.*, n° 155, 1993, p. 325-383.

[30] Magnen (J.) et Sénéor (R.), « Yukawa quantum field theory in three dimensions », dans Lebowitz (J.) et coll., *Third International Conference on Collective Phenomena*, New York Academy of Sciences, 1980.

[31] Maldacena (J.), « The large N limit of superconformal field theories and supergravity », *Adv. Theor. Math. Phys.*, n° 2, 1998, p. 231-252.

[32] McBryan (O.) et Rosen (J.), « Existence of the critical point in φ^4 field theory », *Comm. Math. Phys.*, n° 51, 1976, p. 97-105.

[33] Nelson (E.), « Quantum fields and Markoff fields », dans *Proc. Sympos. Pure Math. XXIII 1971*, Providence, American Mathematical Society, 1973, p. 413-420.

[34] O'Raifeartaigh (L.), *The Dawning of Gauge Theory*, Princeton University Press, 1997.

[35] Osterwalder (K.) et Schrader (R.), « Axioms for Euclidean Green's functions », *Comm. Math. Phys.*, n° 31, 1973, p. 83-112, et *Comm. Math. Phys.*, n° 42, 1975, p. 281-305.

[36] Osterwalder (K.) et Seiler (E.), « Gauge theories on the lattice », *Ann. Phys.*, n° 110, 1978, p. 440-471.

[37] Osterwalder (K.) et Sénéor (R.), « A nontrivial scattering matrix for weakly coupled $P(\varphi)_2$ models », *Helv. Phys. Acta*, n° 49, 1976, p. 525-535.

[38] Politzer (H. D.), « Reliable perturbative results for strong interactions ? », *Phys. Rev. Lett.*, n° 30, 1973, p. 1346-1349.

[39] Rivasseau (V.), *From Perturbative to Constructive Renormalization*, Princeton University Press, 1991.

[40] Salam (A.), « Weak and electromagnetic interactions », dans Svartholm (N.), *Elementary Particle Theory*, Stockholm, 1968, p. 366-377.

[41] Seiberg (N.) et Witten (E.), « Electric-magnetic duality, monopole condensation, and confinement in N = 2 supersymmetric Yang-Mills theory », *Nucl. Phys. B*, n° 426, 1994, p. 19-52.

[42] Seiler (E.), « Schwinger functions for the Yukawa model in two dimensions », *Comm. Math. Phys.*, n° 42, 1975, p. 163-182.

[43] Simon (B.), « Some quantum operators with discrete spectrum but classically continuous spectrum », *Ann. Phys.*, n° 146, 1983, p. 209-220.

[44] Singer (I. M.), « The geometry of the orbit space for nonabelian gauge theories », *Phys. Scripta*, n° 24, 1981, p. 817-820.

[45] Streater (R.) et Wightman (A.), *PCT, Spin and Statistics and all That*, New York, W. A. Benjamin, 1964.

[46] Symanzik (K.), « Euclidean quantum field theory », dans Jost (R.), *Local Quantum Theory*, New York, Academic Press, 1969, p. 152-226.

[47] Weinberg (S.), « A model of leptons », *Phys. Rev. Lett.*, n° 19, 1967, p. 1264-1266.

[48] Wilson (K. G.), « Quarks and strings on a lattice », dans Zichichi (A.), *New Phenomena In Subnuclear Physics, Proceedings of the 1975 Erice School*, New York, Plenum Press, 1977.

[49] Witten (E.), « Supersymmetric Yang-Mills theory on a four-manifold », *J. Math. Phys.*, n° 35, 1994, p. 5101-5135.

[50] Yang (C. N.) et Mills (R. L.), « Conservation of isotopic spin and isotopic gauge invariance », *Phys. Rev.*, n° 96, 1954, p. 191-195.

Index

Les sept problèmes mathématiques du prix du millénaire

Ce recueil regroupe les traductions en français des textes officiels de présentation des sept problèmes mathématiques du millénaire. Il donne donc un petit aperçu de ce à quoi ressemblent les mathématiques contemporaines.