SOMMAIRE DU Nº 89

Éditorial

Dans ce numéro 89 de la Gazette on trouvera le dernier texte du dossier conscré à Michel Herman ; il s'agit d'un remarquable survol de ses travaux fait par Albert Fathi.

Le numéro spécial consacré à l'approche de A. Carbone et M. Gromov de la biologie moléculaire (avec une introduction d' É. Westhof) a connu un vif succès ; nous nous en réjouissons d'autant plus qu'il s'agissait d'un travail d'édition inhabituel pour nous. En conséquence, comme il l'était annoncé dans le numéro 88, notre exploration de l'interaction entre les Mathématiques et la Biologie se poursuit par la contribution de Florence Muri-Majoube et Bernard Prum.

Enfin, hommage est rendu à la note de Henri Lebesgue qui fonde la théorie de l'intégration qui porte son nom ; il nous a semblé que le rôle de la Gazette était de mettre ce document à la portée du plus grand nombre.

Nous avons appris avec tristesse le décès de Jacques-Louis Lions survenu le jeudi 17 mai 2001. Jacques-Louis Lions a été président du C.N.E.S. et de l'Académie des Sciences. Nous lui rendrons hommage dans les numéros à venir.

— Gérard Besson

Mot du Président

Comme chaque année l'Assemblée Générale de notre société s'est réunie en juin 2001 [1], et le tiers du conseil a été renouvelé. Mireille Martin-Deschamps a alors achevé son troisième et dernier mandat comme Présidente de la SMF. Grâce à son remarquable dynamisme et aux nombreuses initiatives qu'elle a prises, notamment à l'occasion de l'année mondiale des mathématiques, la SMF joue un rôle de plus en plus visible dans le paysage mathématique français.

Les publications de la SMF représentent toujours une partie importante de notre activité. Un petit fascicule, intitulé « Explosion des Mathématiques », est en cours de réalisation. Il contribuera à montrer auprès d'un large public que la recherche mathématique est vivante et utile.

Le projet d'extension du CIRM et de la Maison de la SMF à Luminy évolue rapidement : on peut espérer que le chantier démarrera bientôt.

Pour que l'activité et l'influence de la SMF se développent, il convient d'augmenter le nombre d'adhérents : convaincre non seulement les chercheurs confirmés, mais aussi les doctorants ou les Professeurs de Spéciales, par exemple, de l'utilité de notre action.

Un des principaux enjeux de notre environnement scientifique est le formidable défi lancé aux scientifiques par les problèmes liés à l'enseignement. La SMF travaille sur ces questions depuis longtemps, nous ne sommes pas prêts de parvenir à une solution, mais nous poursuivons nos efforts pour progresser. Des actions communes avec d'autres sociétés savantes seront menées, non seulement pour les questions liées à l'enseignement scientifique, mais aussi dans d'autres domaines ; en particulier pour le développement des relations internationales. Le premier congrès SMF/AMS de Lyon en juillet 2001 sera certainement suivi d'autres événements analogues.

Vie de la Société

La journée annuelle de la SMF a eu lieu à Paris à l'IHP le 16 juin 2001, sur le thème : *« Mathématiques et mathématiciens au XX*^{e} *siècle »*. À l'issue de cette journée, les résultats des élections au conseil de la SMF ont été proclamés. Ces résultats, ainsi que la nouvelle composition du conseil et du bureau, se trouvent sur le serveur web de la SMF.

Le premier congrès franco-américain de mathématiques organisé conjointement par la SMF et l'AMS a lieu du 17 au 20 juillet 2001 à l'ENS de Lyon.

[1] Le rapport moral est disponible sur le serveur à l'adresse URL : http://smf.emath.fr/ .

TRIBUNE LIBRE

Naissance et postérité de l'intégrale de Lebesgue

Jean-Pierre Kahane (*Académie des Sciences*)

Il y a quelques mois, Gustave Choquet a proposé à la section de Mathématiques de célébrer le centenaire de l'intégrale de Lebesgue. L'intégrale de Lebesgue, en effet, a profondément marqué le développement des mathématiques au 20ème siècle, et elle est apparue il y a 100 ans, le 29 avril 1901, sous la forme d'une note aux Comptes rendus intitulée : « Sur une généralisation de l'intégrale définie ».

La première suite donnée à la proposition de Gustave Choquet est la note commémorative, signée de Jean-Michel Bony, Gustave Choquet et Gilles Lebeau, qui vient de paraître aux Comptes rendus.

La seconde est cette intervention devant l'Académie[1,2].

Mon programme est de vous parler de la naissance et de la postérité de l'intégrale de Lebesgue. La naissance, je vous l'ai dit, c'est une note aux Comptes rendus, dont nous allons voir reproduite sur l'écran la première page. C'est l'occasion de vous dire un mot des Comptes rendus à cette époque. Puis je vous parlerai un peu de Lebesgue et de ses contemporains, Émile Borel et René Baire. Puis nous ferons des mathématiques. J'essaierai de vous expliquer en quoi l'intégrale de Lebesgue diffère de celle de Riemann, et sa relation avec trois grandes questions qui ont intéressé Lebesgue : le calcul des primitives, la mesure des aires et les séries trigonométriques. Ce sera tout pour la naissance.

Pour la postérité, le choix est immense. L'intégrale et la mesure de Lebesgue ont joué un rôle déterminant dans deux secteurs essentiels des mathématiques du 20ème siècle, l'analyse fonctionnelle et la théorie des probabilités. Je me bornerai à l'évoquer sous deux angles : le théorème de Riesz-Fischer de 1907, et la théorie du mouvement brownien par Norbert Wiener dans les années 1920–1930. Et je conclurai par quelques remarques sur l'évolution des notions de mesure et d'intégrale à travers les âges, et jusqu'à présent.

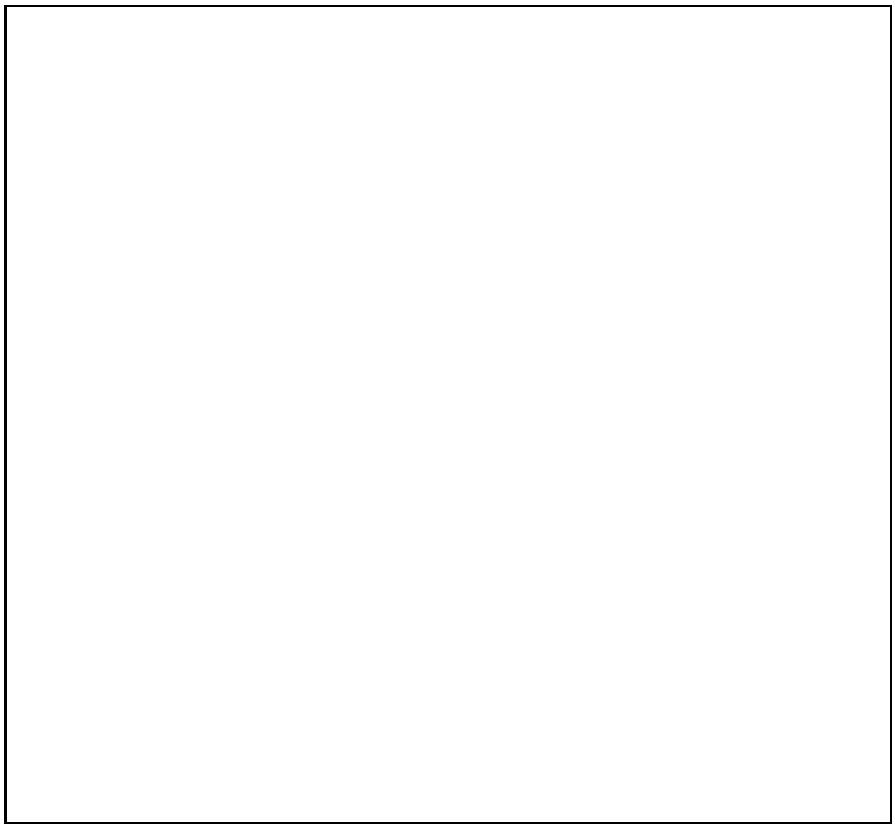
Les archives de l'Académie conservent le manuscrit de la note de Lebesgue du 29 avril 1901. Les quatre feuillets du manuscrit ont été découpés, de façon à permettre à trois typographes de travailler simultanément sur ce texte. Les délais d'impression étaient incroyablement rapides : à cette époque, et c'était

[1] Séance du 5 mars 2001.

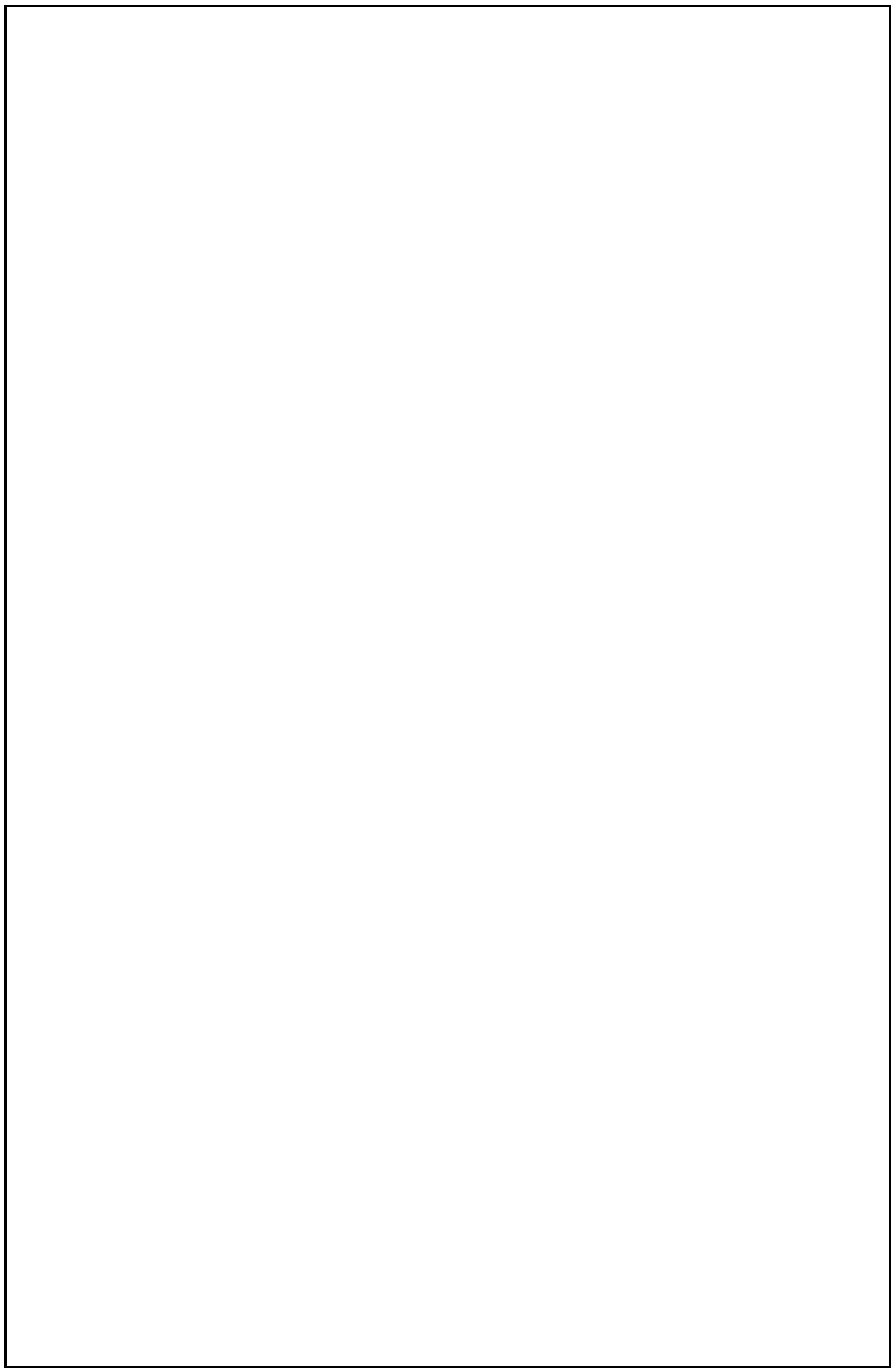
[2] La troisième sera un colloque, à l'École Normale Supérieure de Lyon, à l'initiative de notre confrère Étienne Ghys, les 27 et 28 avril, à l'intention des élèves de l'École et des professeurs de la région, sous la double responsabilité de l'ENS de Lyon et de la Société Mathématique de France.

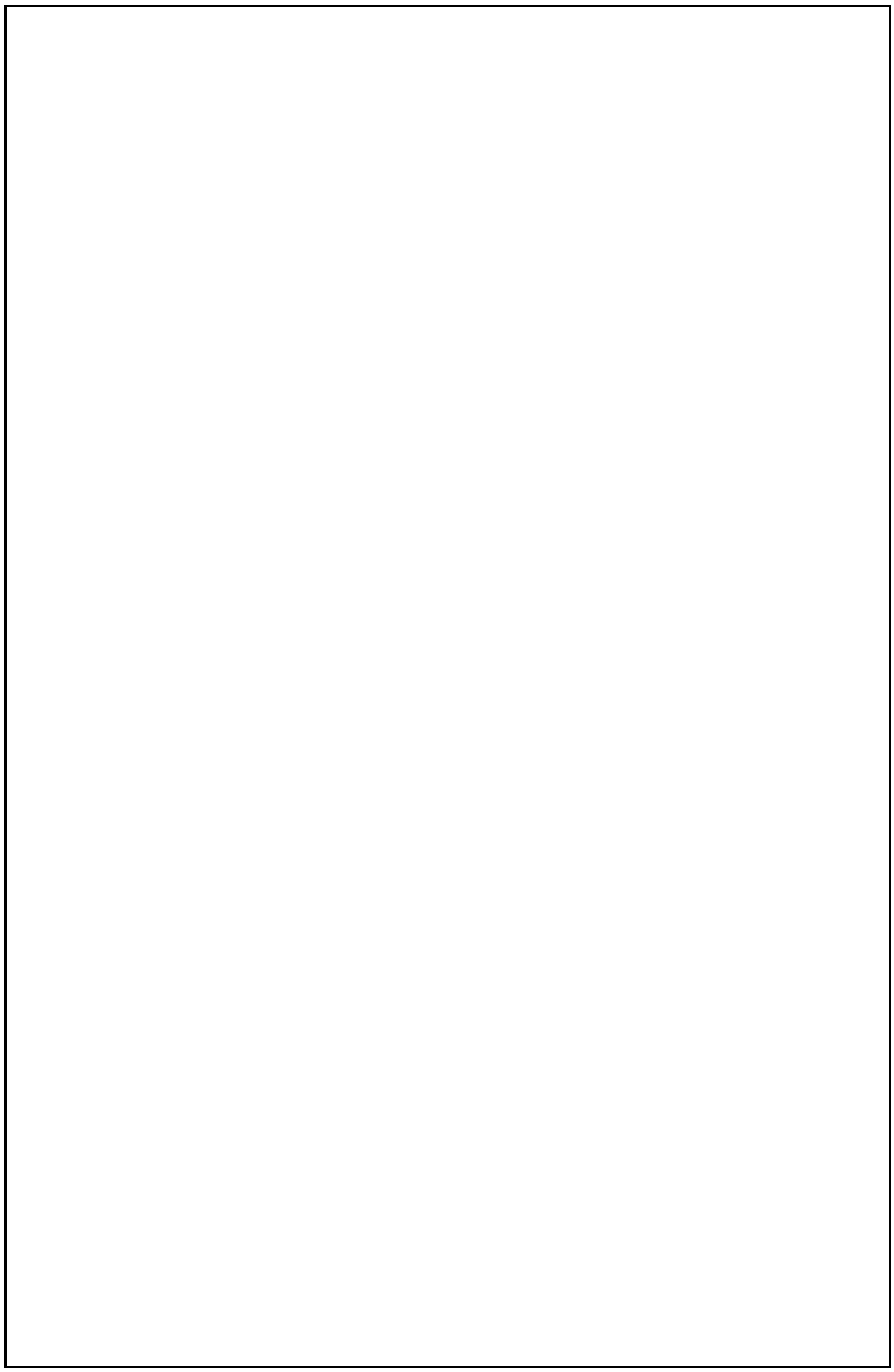
encore le cas au début des années 1950, on pouvait déposer une note à l'Académie le lundi avant 15h, corriger les épreuves le mercredi matin, et voir la note imprimée le lundi suivant.

Les Comptes rendus étaient un grand moyen de communication en mathématiques. Dans les années 1900 et 1901, on trouve les noms de Poincaré, Picard, Painlevé, Hadamard, Borel, Baire, Lebesgue parmi les Français, et parmi les étrangers de Stekloff, Liapounoff, Mittag-Leffler, Levi-Civita, Lindelöf, Fejér, Tzitzeica, Von Koch... C'est un grand journal international, très apprécié par les étrangers comme par les Français. La simple collection des notes en mathématiques de cette époque constituerait un bon document de synthèse sur une bonne partie de l'activité mathématique mondiale.



Note de Lebesgue aux CRAS
reproduite avec l'aimable autorisation de la bibliothèque de l'IHP.





Revenons aux mathématiciens français que j'ai mentionnés. En 1901, Poincaré a 47 ans, Picard 45, Painlevé 38, Hadamard 36, Borel a 30 ans, Baire 27 ans, Lebesgue va avoir 26 ans.

Émile Borel est le chef de file de la nouvelle génération. Il a été enfant prodige, son intelligence est extrêmement vive, il a des intuitions fulgurantes. En 1896, il pressent ce que devrait être une théorie moderne des probabilités, faisant intervenir une infinité dénombrable d'évènements. En 1898, il publie un livre sur la théorie des fonctions dont l'essentiel est consacré aux ensembles de points sur la droite que l'on peut obtenir, à partir des intervalles, par les opérations de réunion dénombrable et de passage au complémentaire ; c'est alors qu'il introduit une notion nouvelle de mesure des ensembles, en en dégageant les propriétés principales, mais sans en achever la théorie. Il va diriger une collection de monographies, chez Gauthier-Villars, où il accueillera en particulier les travaux de Baire et de Lebesgue.

René Baire a passé sa thèse sur les fonctions discontinues. Il les divise en classes, selon la manière dont elles se représentent comme sommes de séries de fonctions continues. Il distingue parmi les ensembles de points les ensembles de 1ère et les ensembles de 2ème catégorie. La théorie de Baire est de nature topologique ; c'est un éclairage différent de celui de Borel, avec une mise au point d'emblée parfaite.

Henri Lebesgue connait bien les travaux de Borel et de Baire, il admire beaucoup Baire et il le tutoie. Par contre, il vouvoie Borel et, dans ses premières lettres, il lui donne du « Cher Monsieur ». À l'École Normale Supérieure, où il a connu Baire, il a eu pour camarades de promotion Paul Montel et Paul Langevin, qu'il admire également et considère non seulement comme un physicien, mais comme un mathématicien, « une espèce à part dans le genre mathématicien », écrira-t-il plus tard. Il a le goût de la géométrie, il connait bien l'analyse classique, et aussi les travaux de Cantor et des Italiens, Dini, Peano, Volterra qui s'intéressent à ce que l'école française dominante, Hermite ou Poincaré, considère comme une tératologie. C'est l'époque où Hermite écrivait à Stieltjes avec quelque malice qu'il se détournait « avec effroi et horreur de cette plaie lamentable des fonctions continues qui n'ont pas de dérivées ». Et Poincaré exprimait encore sa méfiance de façon encore plus carrée : « Autrefois, quand on inventait une fonction nouvelle, c'était en vue de quelque but pratique ; aujourd'hui on les invente tout exprès pour mettre en défaut les raisonnements de nos pères, et on n'en tirera jamais que cela. » Or, dès ses années d'École, Lebesgue est attiré par ce qui est simple et qui sort des sentiers battus. Classiquement, on établit qu'une surface développable, c'est à dire applicable sur un plan, est formée de droites ; effectivement, c'est le cas quand on tord une feuille de papier. Mais, observe Lebesgue, ce n'est plus le cas quand on la froisse. Quelle est donc la théorie des surfaces, hors du cadre classique, qui permet de rendre compte de la situation générale ? Il s'était entretenu de cette question avec Paul Montel à l'École, donc avant 1897, il s'y attache pendant deux année, 1898 et 1899, où il est boursier après l'agrégation, et encore, à partir de 1900, quand il est professeur de classe préparatoire à l'École Centrale au lycée de Nancy. En 1899 et 1900, il publie 5 notes aux Comptes rendus qui concernent toutes les fonctions de plusieurs variables, les surfaces, et particulièrement l'aire des surfaces.

La note de 1901 s'appelle modestement « Sur une généralisation de l'intégrale définie ». Vue d'aujourd'hui, c'est la définition de l'intégrale de Lebesgue et de la mesure de Lebesgue, deux notions de portée immense. De plus, c'est un exposé admirablement clair et succinct de ces deux notions. Comme la plupart des grandes nouveautés, elle n'a pas été reconnue tout de suite. Le Jahrbuch über die Fortschritte der Mathematik, le livre allemand qui, chaque année, recense et analyse toutes les publications mathématiques, a consacré beaucoup de place, plusieurs pages au total, aux notes de Lebesgue sur les surfaces. Il consacre 3 lignes à la note de 1901.

Avant Lebesgue, la conception dominante de l'intégrale est celle de Riemann. Pour intégrer une fonction $y = f(x)$, Riemann découpe l'intervalle (a, b) où varie x et il considère les sommes :

$$\sum y_i(x_{i+1} - x_i) \ ,$$

où y_i désigne une valeur prise par y sur l'intervalle (x_i, x_{i+1}). Si ces sommes tendent vers une limite quand on raffine le découpage, cette limite est l'intégrale $\int_a^b f(x)dx$, et on dit que la fonction f est intégrable au sens de Riemann sur (a, b). Riemann indique une condition nécessaire et suffisante pour qu'il en soit ainsi, que Lebesgue traduira de façon très simple grâce à sa notion de mesure : il faut et il suffit que l'ensemble des points x où la fonction est discontinue soit de mesure nulle.

Lebesgue, lui, découpe l'intervalle où varie y, et il associe à chaque intervalle (y_j, y_{j+1}) de ce découpage la mesure de l'ensemble des x tels que

$$y_j \leqslant f(x) < y_{j+1}$$

Si cette mesure est m_j, une valeur approchée de l'intégrale sera

$$\sum m_j y_j$$

Si ces sommes tendent vers une limite quand on raffine le découpage, c'est l'intégrale au sens de Lebesgue, et Lebesgue dit que la fonction est sommable. Dans sa note de 1901, Lebesgue s'attache uniquement au cas où les deux intervalles où varient x et y sont bornés. Plus tard, il levera cette restriction, qui aurait interdit beaucoup de développements ultérieurs.

Voici comment Lebesgue explique son point de vue, dans une conférence à Copenhague en 1926.

« Avec le procédé de Riemann, on sommait les indivisibles dans l'ordre où ils étaient fournis par la variation de x. On opérait donc comme le ferait un commerçant sans méthode qui compterait pièces et billets au hasard de l'ordre où ils lui tomberaient sous la main ; tandis que nous opérons comme le commerçant méthodique qui dit :

j'ai $m(E_1)$ pièces de une couronne, valant $1.m(E_1)$
j'ai $m(E_2)$ pièces de deux couronnes, valant $2.m(E_2)$
j'ai $m(E_5)$ pièces de cinq couronnes, valant $5.m(E_5)$
etc. Donc j'ai en tout

$$S = 1m(E_1) + 2m(E_2) + 5m(E_5) + ... \ »$$

Vous voyez les notations qu'utilise Lebesgue dans cet exemple. Au lieu d'écrire simplement m_1, m_2, m_5, il met en évidence un ensemble, E_j, et la mesure de cet ensemble, $m(E_j)$. Dans le cas général d'une fonction à valeurs réelles, et non nécessairement entières, il apparait des ensembles E, définis par les inégalités du type $u \leqslant f(x) \leqslant v$, et il s'agit de donner un sens à $m(E)$.

Lebesgue avait en vue un théorème du type

$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$$

qu'il démontre, en effet, sous l'hypothèse que l'intégrale est prise sur un intervalle borné et que les fonctions f_n sont uniformément bornées, et, plus tard, sous l'hypothèse plus générale que les valeurs absolues $|f_n|$ sont toutes majorées par une fonction sommable. Le théorème de Lebesgue est la clé de voûte de la théorie. En même temps, il veut naturellement que l'intégrale soit additive, c'est-à-dire

$$\int (f+g) = \int f + \int g$$

Si l'on impose ces deux conditions, on est amené à une formule du type

$$\int \sum_1^\infty f_n = \sum_1^\infty \int f_n$$

qui doit être satisfaite, en particulier, si les f_n sont les fonctions indicatrices d'ensembles E_n disjoints. Cela veut dire

$$m(\sum_1^\infty E_n) = \sum_1^\infty m(E_n).$$

C'est l'additivité totale de la mesure telle que Borel l'a définie : la mesure d'une réunion dénombrable d'ensembles disjoints est la somme des mesures de ces ensembles. Borel part de la mesure des intervalles, qui est leur longueur, puis il passe aux réunions dénombrables d'intervalles et à leurs complémentaires, puis il répète les opérations de réunion dénombrable et de complémentation, autant de fois qu'il le veut, et il obtient ainsi les ensembles que nous appelons aujourd'hui boréliens. Il est convaincu de pouvoir construire une mesure totalement additive sur la famille en question, que nous appelons aujourd'hui la tribu borélienne, mais il ne le fait pas — ni en 1898 dans ses Leçons sur la théorie des fonctions, ni jamais plus tard.

Lebesgue a donc besoin de la mesure de Borel, mais la mesure de Borel n'existe encore qu'à l'état de projet, et sa construction pas à pas, en suivant les procédés de construction des ensembles boréliens, présente des difficultés. Au lieu de suivre cette voie, Lebesgue prend de la hauteur. Tout ensemble de points sur la droite peut être recouvert par une réunion d'intervalles. À chaque recouvrement on associe la somme des longueurs des intervalles. Puis on considère la borne inférieure de ces sommes pour tous les recouvrements possibles : c'est ce que Lebesgue appelle la mesure extérieure de l'ensemble. Très simplement, il définit une notion complémentaire, la mesure intérieure. Quand la mesure intérieure et la mesure extérieure coïncident, leur valeur commune s'appelle la mesure de l'ensemble, et l'ensemble est dit mesurable. C'est ce

qu'aujourd'hui nous appelons la mesure de Lebesgue, définie sur la tribu de Lebesgue. La tribu de Lebesgue est bien plus grosse que la tribu de Borel, mais la mesure de Lebesgue, sur la tribu de Borel, n'est autre que la mesure de Borel.

C'est donc Lebesgue, indiscutablement, qui a achevé la théorie de la mesure de Borel, tout en introduisant une nouvelle classe d'ensembles, parmi lesquels les ensembles de mesure nulle. Comme tout ensemble mesurable au sens de Lebesgue est la réunion d'un borélien et d'un ensemble de mesure nulle, il est arrivé à Borel de dire que l'apport de Lebesgue était d'avoir introduit les ensembles de mesure nulle. Lebesgue a fort mal pris la chose, et s'est brouillé avec Borel pour cette raison. Il s'était déjà brouillé avec Baire. Les lettres de Lebesgue dans le fonds Borel, qui est depuis un an déposé aux Archives de l'Académie, expliquent les circonstances de ces dissentiments. Au début des années 1900, Borel semblait choyé par la vie, époux d'une femme remarquable qui était la fille du mathématicien Paul Appell, Baire était sans cesse malade, et Lebesgue était pauvre, surchargé d'enseignement (21 heures par semaine au lycée de Nancy), accablé de problèmes familiaux et financiers.

C'est juste avant 1900 qu'avait été créée la fondation Peccot au Collège de France, pour permettre à un jeune mathématicien — moins de 30 ans — d'exposer ses travaux en lui accordant quelques subsides. Les cours Peccot existent toujours et ils sont prestigieux dans le monde mathématique. Les premiers titulaires en ont été Borel pendant trois ans, puis Lebesgue à la place de Baire, qui était malade, puis Baire, puis de nouveau Lebesgue. Les querelles autour du cours Peccot, comme autour des nominations universitaires, n'ont qu'une valeur anecdotique, mais ces cours Peccot ont eu un impact considérable.

Après la note de 1901, Lebesgue a developpé la théorie de l'intégrale dans trois œuvres majeures : sa thèse, passée en 1902, qui s'appelle « Intégrale, longueur, aire » — et le titre indique bien le lien entre l'intégrale et les questions géométriques qui avaient fait l'objet de ses premières notes — puis, suite au premier cours Peccot, les « Leçons sur l'intégration et la recherche des fonctions primitives » et, suite au second cours Peccot, les « Leçons sur les séries trigonométriques ». Je vais tenter de donner un aperçu des matières traitées dans ces travaux. Mais auparavant, je veux souligner le rôle qu'ont eu les cours Peccot pour la continuation des travaux de Baire et de Lebesgue. Les auditeurs de ces trois cours, de Lebesgue, Baire et Lebesgue, étaient peu nombreux, mais il y avait parmi eux Arnaud Denjoy, qui a rédigé le cours de Baire et qui a été l'auteur d'une nouvelle théorie de l'intégration, la « totalisation », fondée à la fois sur les travaux de Cantor, de Baire et de Lebesgue, et Pierre Fatou, dont la thèse en 1906, à la suite du second cours Peccot de Lebesgue, sur « Séries trigonométriques et séries de Taylor », a montré tout le parti que l'analyse classique pouvait tirer du nouvel outil qu'était l'intégrale de Lebesgue.

Ainsi Fatou et Denjoy ont-ils été, en France et à cette époque, les principaux continuateurs de Lebesgue. À l'étranger, en Angleterre, en Belgique, en Autriche et en Hongrie, en Russie et en Pologne, l'intégrale de Lebesgue a été immédiatement adoptée, exposée et utilisée. Mais en France, elle n'était pas enseignée. Lebesgue lui-même ne l'a exposée qu'à l'occasion des cours Peccot. Après qu'il ait été nommé professeur au Collège de France, en 1921, ses cours, sauf rares exceptions, ont porté sur de tout autres sujets. En 1950, alors que dans tous les pays du monde l'intégrale de Lebesgue était un sujet classique,

on pouvait être en France agrégé de mathématiques sans en avoir entendu parler. Szolem Mandelbrojt a évoqué dans ses souvenirs la déception qu'il avait éprouvée, en arrivant en France, de ne voir enseignée nulle part l'intégrale de Lebesgue et les théories qui en étaient issues.

Revenons donc aux trois écrits principaux de Lebesgue, de 1902, 1904 et 1906. Le premier, la thèse, est d'inspiration géométrique. Quoiqu'il contienne et complète la matière de la note de 1901 sur l'intégrale, il est inspiré du début à la fin par les questions qui l'avaient intéressé en 1899 et en 1900 : la longueur d'une courbe, et surtout l'aire d'une surface. Je vais me borner ici aux aires planes et à leur relation avec la théorie de la mesure. Camille Jordan avait exposé une théorie des aires planes qui a fortement influencé Lebesgue. Étant donné un domaine plan D, Jordan considère tous les polygones qui contiennent D, et tous les polygones qui sont contenus dans D ; la borne inférieure des aires des polygones contenant D est « l'étendue extérieure » de D, la borne supérieure des polygones contenus dans D est « l'étendue intérieure ». Lorsque les deux coïncident, on dit que le domaine est quarrable et que son aire est la valeur commune de l'étendue extérieure et de l'étendue intérieure. On voit que Lebesgue a tiré l'idée des mesures extérieure et intérieure de l'approche de Jordan — mais les définitions devaient être modifiées de façon radicale.

Pour faire saisir la différence entre l'étendue de Jordan et l'aire selon Lebesgue, je vais me contenter d'une figure, analogue à celle qu'expose Lebesgue dans une note de sa thèse. Partons d'un triangle ABC, dont la surface est S. Choisissons sur le côté BC deux points B' et C', et appelons S_1 la surface du triangle $AC'B'$. Lorsque l'on ôte du triangle ABC le $AC'B'$, on obtient deux nouveaux triangles, ABC' et $AB'C$, dont la surface totale est $S - S_1$. Puis on évide le triangle ABC' et le triangle $AB'C$ de la même manière, en leur ôtant des triangles de sommets C' et B' respectivement, dont les surfaces sont S_2 et S_3. On obtient maintenant quatre triangles dont la somme des surfaces est $S - S_1 - S_2 - S_3$. En continuant indéfiniment de cette manière, on a une suite de domaines emboités, formés de triangles, contenus dans le triangle ABC, et dont le complément relativement à ABC est constitué de triangles de surfaces S_1, S_2, S_3, S_4 etc. À la limite, on obtient une courbe Γ, contenue dans le triangle ABC, dont le complément est constitué d'une infinité de triangles dont la somme des surfaces est $\sum_1^\infty S_n$. Si l'on complète cette courbe Γ par un arc de cercle joignant B et C, à l'extérieur du triangle ABC, on obtient une courbe fermée simple qui délimite un domaine D. Lorsque $\sum_1^\infty S_n = S$, le domaine D est quarrable au sens de Jordan, et il ne l'est pas quand $\sum_1^\infty S_n < S$. Par contre, il est toujours mesurable au sens de Lebesgue. Au sens de Lebesgue, la courbe Γ a toujours pour aire $S - \sum_1^\infty S_n$, que cette quantité soit positive ou nulle. Aujourd'hui, nous appellerions une telle courbe une fractale.

La note de 1901 met en évidence une propriété importante de l'intégrale de Lebesgue : elle résout le problème des primitives pour des fonctions bornées. Cela veut dire que si une fonction f est dérivable sur un intervalle et que sa dérivée, f', est bornée, la différence

$$f(x) - \int_0^x f'(t)dt$$

est une constante. Si f' est intégrable au sens de Riemann, il en est ainsi. Mais il existe des cas où f' existe et est bornée, et n'est pas intégrable au sens de Riemann. Volterra en a donné un exemple, qui est reproduit dans la thèse de Lebesgue. Les « Leçons sur l'intégration » de 1904 contiennent la description et l'analyse des notions d'intégrale qui se sont succédées depuis Cauchy, en particulier avec Dirichlet et Riemann, la caractérisation des fonctions intégrables au sens de Cauchy-Dirichlet et au sens de Riemann, et leur relation à la théorie des fonctions dérivées. C'est Denjoy, en 1912, qui a généralisé l'intégrale de Lebesgue de façon à calculer les primitives des fonctions dérivées les plus générales, c'est à dire à s'affranchir de l'hypothèse que f' est bornée.

À la suite des « Leçons sur l'intégration » d'excellents livres ont exposé l'intégrale de Lebesgue et ses applications, en particulier par Charles de la Vallée Poussin en 1916 (« Fonctions d'ensembles, intégrales de Lebesgue, classes de Baire ») et par Felix Hausdorff en 1927 (« Mengenlehre »). Mais la bible sur le sujet est l'ouvrage de Stanislas Saks paru en 1933 dans la collection polonaise des Monographies mathématiques, il est écrit en français et a pour titre « Théorie de l'intégrale ». Voici son appréciation sur le lien réalisé par Lebesgue entre intégration et dérivation :

« Le mérite de M. Lebesgue ne se réduit pas seulement à la création d'une notion nouvelle et plus générale de l'intégrale, ni même à la mise de cette notion en relation étroite avec la théorie de la mesure, la valeur de son œuvre consiste en premier lieu dans sa théorie de la dérivabilité, qu'il a construite parallèlement à celle de l'intégrale. C'est grâce à elle que la découverte de M. Lebesgue a trouvé tant d'applications dans les branches les plus diverses de l'analyse... »

La première de ces branches est la théorie des séries trigonométriques. Historiquement, il y a un rapport étroit entre la notion d'intégrale et les séries trigonométriques. Cela s'explique par le fait que les coefficients d'une série trigonométrique se calculent au moyen d'intégrales, par les formules de Fourier : si

$$f(t) = \sum c_n e^{int}$$

on a

$$c_n = \int_0^{2\pi} f(t) e^{-int} dt/2\pi.$$

La notation de l'intégrale définie, $\int_a^b$, a été introduite par Fourier à cette occasion, et Fourier attribuait à ces formules une valeur uiverselle. En fait, une bonne partie de la théorie des séries, de celle de l'intégrale, et plus tard de l'analyse fonctionnelle allait consister à leur donner un sens. Quelques étapes marquantes sont l'article de Dirichlet de 1830, la thèse de Riemann sur les séries trigonométriques de 1854, la thèse de Cantor en 1870, le théorème de Fejér de 1900 (encore une note aux Comptes rendus !), les leçons de Lebesgue publiées en 1906, le théorème de Riesz-Fischer en 1907, la totalisation de Denjoy sous sa seconde forme en 1921, la théorie des distributions de Schwartz en 1949, et la liste n'est pas close.

Pour Dirichlet, il s'agissait d'intégrer des fonctions continues, en raccordant les intégrales aux points de discontinuité. Comme exemple de fonction non intégrable, il donnait la fonction définie sur $(0, 1)$, qui vaut 1 sur les irrationnels

et 0 sur les rationnels. Pour Riemann, nous l'avons vu, la notion d'intégrale est plus générale ; la fonction de Dirichlet est toujours non intégrale. Pour Lebesgue au contraire, c'est une fonction intégrable (pour éviter la confusion, Lebesgue dit « sommable ») et son intégrale vaut 1.

Riemann dans sa thèse avait mené l'étude des fonctions qui sont sommes en tout point d'une série trigonométrique. Il les avait caractérisées sans avoir utilisé le calcul des coefficients, puisqu'en général ces fonctions ne sont pas intégrables au sens de Riemann. Cantor avait montré que les coefficients sont bien déterminés par la somme de la série. Un très beau résultat de Lebesgue est que, si une fonction somme d'une série trigonométrique est bornée, elle est intégrable en son sens, et les coefficients s'obtiennent par les formules de Fourier. Pour lever la restriction que la fonction est bornée, il a fallu attendre Denjoy et sa seconde totalisation.

Les leçons de Lebesgue sur les séries trigonométriques contiennent une foule de résultats qui, joints à la thèse de Fatou, ont immédiatement conquis le public mathématique en dehors de la France. Elles ont aussi deux particularités. C'est la première fois, que dès le départ, Lebesgue donne la définition générale de l'intégrale, des fonctions non-bornées comme des fonctions bornées. Et c'est la dernière fois que Lebesgue présente un exposé de son intégrale.

En évoquant la naissance de l'intégrale de Lebesgue j'ai été amené, déjà, à parler de sa descendance. S'agissant de la postérité de l'intégrale de Lebesgue, je me bornerai maintenant à trois aspects : l'analyse fonctionnelle, la théorie moderne des probabilités, et quelques prolongements directs de la mesure et de l'intégrale de Lebesgue.

Pendant l'année 1907, le Hongrois Frédéric Riesz et l'Autrichien Ernst Fischer ont publié une série de 5 notes aux Comptes rendus qui ont donné un nouvel éclairage à l'intégrale de Lebesgue. Par des méthodes légèrement différentes, ils donnent une caractérisation des coefficients de Fourier des fonctions dont le carré est intégrable sur l'intervalle $(0, 2\pi)$: les carrés de leurs valeurs absolues forment une série convergente. Ainsi, non seulement on a l'égalité à laquelle Fatou a donné le nom de Parseval,

$$\sum_{-\infty}^{\infty} |c_n|^2 = \int_0^{2\pi} |f(t)|^2 \frac{dt}{2\pi},$$

lorsque les c_n sont les coefficients de Fourier-Lebesgue (je veux dire, obtenus par les formules de Fourier en utilisant l'intégrale de Lebesgue), mais encore, comme l'explique Frédéric Riesz de manière plaisante dans un article de 1949, les formules de Fourier constituent « un billet d'aller et retour permanent » entre deux espaces à une infinité de dimensions, l'espace ℓ^2 des suites de carré sommable et l'espace L^2 des fonctions de carré intégrable. En termes plus savants, la transformation de Fourier est un isomorphisme entre l'espace L^2 et l'espace ℓ^2, déjà introduit par Hilbert.

La clé de la démonstration, chez Fischer comme chez Riesz, tient en trois mots : « L^2 est complet ». Plus tard, Frédéric Riesz explicite une version plus générale de ce résultat : « L^p est complet », en désignant par L^p l'espace des fonctions dont la p-ième puissance de la valeur absolue est intégrable, p étant un nombre réel quelconque $\geqslant 1$.

L'expression « L^2 est complet » signifie que l'on peut traiter L^2 à la manière d'un espace euclidien, comme Fischer et Riesz en avaient clairement l'intuition en parlant l'un et l'autre d'une sorte de « géométrie des fonctions ». L'idée de traiter les fonctions comme des points dans un espace abstrait n'était pas nouvelle : on la trouve chez Volterra, Hadamard et surtout dans la thèse de Maurice Fréchet de 1906, mais elle prenait là une forme frappante. Pour les fondateurs de la mécanique quantique, l'espace L^2 a été un outil immédiatement disponible. D'autres classes de fonctions ou de distributions peuvent être construits, à partir de L^2, en conservant cette structure géométrique d'espaces de Hilbert : ce sont les espaces de Sobolev, essentiels dans la théorie des équations aux dérivées partielles.

Les espaces L^p avec $p \neq 2$ ont une géométrie différente. Ce sont des espaces de Banach. C'est d'ailleurs Banach, dans sa « Théorie des opérations linéaires », le premier volume des *Monographies mathématiques polonaises*, qui a donné la forme actuelle aux théorèmes « L^2 est complet », « L^p est complet » : dans les écrits de Fischer et de Riesz, il fallait de longues phrases pour exprimer cela. En 1907, les espaces « L^p » n'étaient pas encore nommés, et l'adjectif « complet » avait un autre sens. Banach, en définissant « complet » avec le sens actuel et en nommant les espaces L^p (d'après l'initiale de Lebesgue), a fait passer dans les définitions l'essence de ces propositions, et c'est pourquoi, maintenant, elles s'énoncent de façon si simple.

Ainsi, au départ, il y a un problème sur les coefficients de Fourier, et une solution élégante, c'est le théorème de Riesz-Fischer. Puis une proposition intermédiaire, utilisée comme lemme, émerge comme résultat important, et prend le nom de théorème. Elle est encore longue et compliquée à énoncer. Enfin, deux définitions viennent permettre d'énoncer ce théorème en trois mots. L'élixir de la théorie est passée dans les définitions.

L'analyse fonctionnelle contemporaine utilise bien d'autres espaces que les espaces L^p, mais elle ne peut pas se passer des espaces L^p. Il n'est pas abusif de placer Lebesgue à la base de l'analyse fonctionnelle, tant par l'importance de l'intégrale comme forme linéaire que par les espaces fonctionnels qu'elle a permis de considérer.

L'influence de Lebesgue sur la théorie des probabilités a suivi deux voies principales, auxquelles on peut attacher les noms du Polonais Hugo Steinhaus et de l'Américain Norbert Wiener, au cours de la même décennie, 1920–1930.

L'idée de Steinhaus est de fonder les probabilités sur la considération de l'intervalle $(0, 1)$ de la droite réelle, que je désignerai par I. Il constitue ainsi un dictionnaire : un évènement est une partie de I mesurable au sens de Lebesgue, sa probabilité est la mesure de Lebesgue de cette partie ; une variable aléatoire est une fonction définie sur I et mesurable au sens de Lebesgue, sa distribution est l'image de la mesure de Lebesgue par cette fonction, et son espérance, quand elle existe, est l'intégrale de Lebesgue de cette fonction, et ainsi de suite. Il montre comment construire sur I une suite de variables aléatoires indépendantes ayant toutes pour distribution la mesure de Lebesgue sur I. À partir de là, il a un modèle mathématique pour toutes les suites ou séries de variables aléatoires indépendantes. La première application qu'il donne est celle-ci : si, dans une série de Taylor admettant un cercle C comme cercle de

convergence, on change les phases des coefficients de manière aléatoire suivant la probabilité naturelle, on obtient presque sûrement (c'est-à-dire avec probabilité égale à 1) une fonction qui ne se prolonge pas analytiquement hors de C. C'était là une idée de Borel, qui datait de 1896, mais que ni Borel ni personne d'autre n'avait jusqu'alors converti en énoncé mathématique.

L'idée de Wiener était tout autre. Wiener avait lu Einstein et Jean Perrin, qui avaient établi, théoriquement et expérimentalement, une théorie physique du mouvement brownien. Jean Perrin avait observé que les trajectoires du mouvement brownien rappelaient les fonctions continues sans dérivées des mathématiciens, que l'on tenait jusque là pour des curiosités tout à fait étrangères au monde physique. La réalité physique du mouvement brownien suggérait qu'il y en ait un modèle mathématique. Le programme de Wiener était de munir l'ensemble des fonctions continues du temps t, partant d'un point donné au temps $t = 0$, d'une mesure de probabilité telle que les équations d'Einstein soient satisfaites et que, presque sûrement, ces fonctions jouissent de propriétés convenables, dont celle de n'être dérivables nulle part. Construire une mesure sur un espace plus gros que l'intervalle I n'était pas chose nouvelle. Mais la construction de la mesure de Wiener était un tour de force. Il s'agissait, à partir de conditions données a priori sur le processus — c'est un processus gaussien et ses accroissements sont indépendants — de construire la mesure de façon que ces propriétés soient vérifiées. Une fois la mesure de probabilité construite, on peut traiter la fonction comme un objet aléatoire. Wiener l'appelait « the fundamental random function », puis, suivant Paul Lévy, on l'a appelé tout simplement « le mouvement brownien ». C'est un objet fascinant, d'usage constant dans plusieurs domaines des mathématiques, et sur lequel physiciens et mathématiciens fournissent en permanence de merveilleuses conjectures et de superbes théorèmes.

Ainsi Wiener n'utilisait pas la mesure de Lebesgue, mais il en construisait une variante sur un espace de fonctions. Au contraire Steinhaus prenait comme espace de probabilité universel l'intervalle $(0, 1)$ de la droite réelle, muni de la mesure de Lebesgue. On peut parfaitement construire le mouvement brownien à partir du point de vue de Steinhaus, au moyen de la série de Fourier-Wiener, qui est une série trigonométrique dont les coefficients sont des variables aléatoires gaussiennes indépendantes, et c'est ce que Wiener a fait dans le dernier exposé qu'il a présenté de sa « fundamental random function » en 1934.

Or, au moment précis ou Wiener se ralliait au point de vue de Steinhaus, Kolmogorov reprenait l'idée de Wiener, de construire un espace de probabilité adapté à un processus donné par une loi. Les Grundbegriffe der Wahrscheinlichkeitsrechnung de Kolmogorov définissent un espace de probabilité comme un ensemble Ω, muni d'une tribu de parties qu'on appelle les évènements, et d'une mesure totalement additive sur cette tribu qu'on appelle la probabilité. C'est le triplet $(\Omega, \mathcal{A}, P)$ qui, depuis lors, est le signe de ralliement des probabilistes. Le point de vue de Kolmogorov est plus flexible que celui de Steinhaus parce qu'en cours de route, dans le développement d'une théorie, on peut changer d'espace Ω ou de tribu $\mathcal{A}$. Mais les espaces de probabilité les plus importants sont isomorphes à l'espace de Lebesgue, c'est à dire à l'intervalle $(0, 1)$ de la droite réelle, muni de la tribu et de la mesure de Lebesgue.

La postérité de l'intégrale de Lebesgue serait impressionnante au seul titre des espaces fonctionnels et des espaces de probabilité. Mais elle comporte bien d'autres descendants, que je veux évoquer rapidement.

Je commencerai par la théorie géométrique de la mesure, où l'article fondateur est celui de Felix Hausdorff, en 1919. Hausdorff transpose à une situation plus générale la construction de la mesure extérieure de Lebesgue. Au lieu de la droite, il considère un espace sans spécification particulière. Au lieu d'intervalles, il dispose d'un stock de petits corps pesants, susceptibles de recouvrir un objet quelconque situé dans l'espace. Étant donné un tel objet, il attache à chaque recouvrement une masse, qui est la masse totale des corps qui participent au recouvrement — en général, une infinité de corps est nécessaire. Puis, il prend la borne inférieure de ces masses, pour tous les recouvrements astreints à une certaine condition de finesse — par exemple que le diamètre des corps ne dépasse pas un ε donné — puis la limite de cette borne inférieure lorsque l'on affine cette condition — par exemple, lorsque l'on fait tendre ε vers zéro. Il obtient alors pour l'objet en question une mesure, qui a toutes les propriétés de la mesure extérieure de Lebesgue, et qui peut être finie ou infinie. C'est la mesure de Hausdorff. Si l'espace est muni d'une distance, on peut choisir comme stock de petits corps l'ensemble de toutes les boules, et pour masse d'une boule une puissance donnée du diamètre, disons, d^α. On obtient alors la mesure de Hausdorff en dimension α. Elle est en général infinie lorsque α est petit, et nulle lorsque α est grand. La dimension de Hausdorff est la valeur de α qui délimite ces deux situations.

Avant même la grande vogue de la fractalité, mesure et dimension de Hausdorff ont joué un rôle important en analyse harmonique et en théorie du potentiel. En théorie du potentiel la notion naturelle n'est plus la mesure, mais la capacité d'un ensemble. Choquet, dans sa théorie des capacités, est un héritier lointain de Lebesgue.

Une autre généralisation de la mesure de Lebesgue concerne les espaces dans lesquels on peut définir des objets congruents. Selon certaines conditions, on peut construire une mesure, unique à une normalisation près, telle que deux objets congruents aient la même mesure. C'est là une découverte du Hongrois Alfred Haar, simplifiée ensuite par Banach, et exposée dans la « Théorie de l'intégrale » de Saks sous forme définitive. La mesure de Haar est un outil essentiel dans l'étude des groupes topologiques.

Sans que le terme de tribu soit utilisé à l'époque, il semblait à Borel qu'aucune opération de l'analyse ne ferait jamais sortir de la tribu borélienne. C'était aussi l'avis de Lebesgue, et il avait cru le démontrer en 1905. Rarement erreur a été plus fructueuse. Au début de l'année 1917, les Comptes rendus publient deux notes des Russes Nicolas Lusin et M. Ya. Souslin qui établissent le contraire : on peut trouver une suite de polynômes $P_n(x)$, qui convergent vers une valeur $y(x)$ quand x appartient à un certain ensemble borélien, mais tels que l'ensemble des valeurs prises par $y(x)$ ne soit pas borélien. Ainsi l'image borélienne d'un borélien n'est pas nécessairement un borélien. Il y a mieux : l'ensemble des couples $(x, y(x))$ est un borélien, mais sa projection sur l'axe des y, qui est l'ensemble des $y(x)$, ne l'est pas. La projection d'un borélien n'est pas nécessairement un borélien. L'analyse classique force donc à sortir de la tribu borélienne. Entre la tribu de Borel et celle de Lebesgue se trouve la tribu de Lusin, constituée par

les ensembles que Lusin appelle analytiques et qui sont des images continues de boréliens.

Au cours des années 1920 s'est développée à Moscou une école mathématique extrêment brillante, dont Lusin a été le fondateur. Ainsi Lebesgue et son œuvre ont été beaucoup mieux connus à Moscou qu'ils ne l'étaient en France. Hongrie, Pologne et Russie ont été les foyers de rayonnement de la pensée de Lebesgue et de son héritage.

Je terminerai par quelques mots sur l'intégrale, autrefois et aujourd'hui. À l'époque de Cauchy se sont dégagées des notions qui sont restées intangibles, telles que la convergence des suites ou des séries, la continuité, la dérivabilité, l'analyticité. La notion générale de fonction s'est aussi établie à cette époque. Il était naturel de chercher à préciser ce qu'est une fonction intégrable. Après les tentatives de Cauchy et de Dirichlet, de rattacher l'intégrabilité à la continuité des fonctions sur certains intervalles, l'intégrale de Riemann a fait l'unanimité pendant plusieurs décennies : une fonction était intégrable si elle admettait une intégrale au sens de Riemann. C'est pourquoi Lebesgue utilise un autre adjectif, sommable, pour qualifier les fonctions susceptibles d'intégration selon sa methode. C'est avec Hardy et Littlewood que l'on a commencé à parler de fonctions intégrables au sens de Lebesgue. L'intégrale qui s'est imposée au 20ème siècle est l'intégrale de Lebesgue, mais elle a subi bien des extensions et modifications : les intégrales de Denjoy, de Perron, de Henstock ; celle de Radon, de Saks, de Haar ; celles de Wiener, d'Itô, de Feynman. Ces différentes intégrales interviennent dans l'étude des fonctions d'une variable réelle, en analyse fonctionnelle, en probabilités et en physique théorique. L'intégrale de Feynman a donné du fil à retordre aux mathématiciens. Une approche rappelle celle de Lebesgue, et elle est due à Pierre Cartier et Cécile de Witt : elle consiste à imposer des conditions à l'intégrale en question qui en assurent l'unicité, et alors seulement à donner un procédé de construction.

On a beaucoup de choix pour enseigner l'intégrale, à tous les niveaux, parce que c'est une notion polysémique. On peut dire, avec Youri Manin, que l'intégrale, c'est la quantité de quelque chose dans un domaine. Mais d'autre part, intégrer une équation différentielle, c'est partir d'une donnée sur la variation instantanée pour décrire l'évolution d'un phénomène. Ainsi au lycée, on peut privilégier la solution de $y' = f(x)$, comme on l'a fait pendant quelques années, ou privilégier le calcul des aires, comme on le fait maintenant : en tout cas, il faut tenir les deux bouts. Au niveau supérieur, on peut rattacher l'intégrale aux fonctions de variables réelles, à l'analyse fonctionnelle ou aux probabilités. Bourbaki avait choisi de présenter l'intégrale comme forme linéaire sur un espace de fonctions continues, donc le substrat était un espace topologique. Denjoy, fidèle à la distinction entre le point de vue de Baire, topologique, et celui de Lebesgue, métrique, avait violemment attaqué Bourbaki, et aujourd'hui le point de vue probabiliste donne raison à Denjoy. Cependant, à lire l'autobiographie de Laurent Schwartz, on voit que cette erreur de Bourbaki, si erreur il y a, a joué un rôle décisif pour découvrir la bonne définition des distributions, comme formes linéaires sur certains espaces de fonctions.

La conclusion est que, quelle que soit la perfection d'un exposé mathématique, et l'exposé de Lebesgue en 1901, d'emblée, atteint une sorte de perfection,

les mathématiques sont beaucoup plus riches que tout exposé qui peut en être fait. L'efflorescence mathématique qui a suivi l'intégrale de Lebesgue m'a paru justifier ce long compte rendu.

Michel Herman : souvenirs de ses travaux

Albert Fathi (*ENS Lyon*)

Au revoir Michel,
Ce que je connais de tes travaux est si peu,
Ce que je voudrais connaître est immense.
Bien à toi,

Albert

Ce que l'on trouvera ici n'est pas du tout une discussion objective des travaux de Michel Herman. Il s'agit d'un point de vue personnel et subjectif sur Michel comme mathématicien et ami. Tout ce que je rapporte de personnel est basé sur ma mémoire et elle seule. Comme tel, tout a certainement été modifié et embelli au cours des années. Je m'en excuse d'avance auprès du lecteur. Ma seule vraie excuse est d'avoir été présent avec Michel dès le début de ses 25 Glorieuses (années) et d'avoir assisté en direct à la gestation de ses mathématiques.

Dans la suite, je ne parlerai que de ses articles effectivement publiés et parus. Ceci ne représente que la partie émergée de l'iceberg, surtout au cours de ces dix dernières années où Michel n'a que très peu publié et beaucoup produit.

Michel Herman est un ancien élève de l'École polytechnique. Il est de la même promotion qu'Alain Chenciner et François Laudenbach. On trouvera par ailleurs le texte d'Alain sur Michel sur cette amitié de plus de 35 ans [41]. À l'École polytechnique, Laurent Schwartz a eu une grande influence sur Michel Herman. Il l'a très souvent mentionnée et il a toujours été très reconnaissant envers Laurent Schwartz pour la confiance qu'il lui a accordé dès le début.

Dans les premières années de la carrière de mathématicien de Michel Herman, sa rencontre avec Harold Rosenberg a certainement été déterminante. J'ai toujours pensé que c'est Harold qui lui a fait découvrir les systèmes dynamiques à travers le théorème de Denjoy et ses généralisations aux feuilletages, sujet en pleine effervescence dans la fin des années 60, voir [42]. C'est Harold qui a été le directeur de thèse de Michel, même s'il a partout clamé qu'il a beaucoup appris et peu dirigé dans ce cas là, voir [43].

La plus ancienne publication mathématique portant le nom de Michel Herman que je connaisse est [44]. C'est la rédaction conjointe avec Alain Chenciner et François Laudenbach du cours sur le fibré tangent de Larry Siebenmann donné au premier semestre 66 à Orsay. Ceux qui connaissent les exigences de Larry en matière de rédaction ne seront pas surpris que les notes aient mis 3 années pour être publiées. Apprendre à rédiger des mathématiques avec Larry Siebenmann est une expérience que nous avons partagée à quelques années d'intervalle et nous nous en flattions mutuellement au cours des années.

Est-ce ce premier exemple qui a rendu Michel si exigeant sur la qualité des manuscrits que l'on donne à publier ? Toujours est-il que ses archives contiennent un nombre impressionnant de manuscrits contenant des résultats remarquables et dont il n'était pas content de la rédaction au point de ne pas les publier, bien que l'état dans lequel ils sont soit bien supérieur à beaucoup d'articles publiés même dans des revues très prestigieuses. De cette rédaction du cours de Siebenmann, Michel a gardé toute sa vie un très vif intérêt pour la topologie géométrique et ses avancées, nous en discutions très souvent.

Le premier travail publié [1] de Michel Herman date de 1971. Il y montre que le groupe linéaire $GL(\mathbf{H})$ et le groupe unitaire $U(\mathbf{H})$ d'un espace de Hilbert $\mathbf{H}$ de dimension infinie sont parfaits, c'est-à dire qu'ils sont égaux à leurs sous-groupes engendrés par les commutateurs ou encore que leurs abélianisés sont triviaux. C'est Nicole Desolneux-Moulis, qui a analysé ce travail pour les Mathematical Reviews (**44 #7586**), elle y remarque que la démonstration est élégante mais que le résultat était déjà connu. Nicole Desolneux-Moulis a été jusqu'à la fin de sa vie (elle est décédée dans l'année précédant la mort de Michel) un des mathématiciens très proches de Michel. Elle était aussi un des participants les plus assidus au séminaire de systèmes dynamiques de Michel Herman à l'École polytechnique et à l'Université Paris 7 par la suite. Parmi les souvenirs forts que j'ai se trouvent deux discussions bien précises, au café après le séminaire, avec Michel et Nicole qui m'ont amené en l'espace de quelques minutes à comprendre un problème et à publier un article.

Le second travail publié dans la foulée [2] concerne le groupe $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$ des difféomorphismes C^∞ isotopes à l'identité du tore $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ de dimension n. On cherchait à comprendre la structure algébrique des groupes de difféomorphismes, car ils intervenaient dans le calcul de l'homologie des divers espaces classifiants des feuilletages et donc dans la théorie de leurs classes caractéristiques. On conjecturait que le groupe $\mathrm{Diff}^\infty_+(M)$ des difféomorphismes C^∞ isotopes à l'identité de la variété compacte sans bord M était simple. David Epstein venait de montrer que le sous-groupe des commutateurs de $\mathrm{Diff}^\infty_+(M)$ était simple, [45]. Il restait à voir qu'il était parfait. C'est ce que Michel a fait dans le cas $\mathbb{T}^n$. Un très bel argument géométrique de Bill Thurston permet d'en déduire le théorème pour toute autre variété compacte sans bord. Il est à noter que, dans le cas de la dimension 1, le résultat de [2] a été aussi annoncé par John Mather [46]. Dans la suite, les travaux de Michel et John ont bien sûr connu bien d'autres croisements et moments d'influence réciproque. On peut se référer à l'article [47] de John Mather pour un exposé succinct sur la simplicité du groupe $\mathrm{Diff}^\infty_+(M)$. Il est toutefois intéressant d'esquisser la démonstration contenue dans [2] et développée dans [5], car on voit déjà là apparaître des objets et des théorèmes qui sont des fils conducteurs dans l'œuvre de Herman : l'action de $\mathrm{PSL}(2,\mathbb{R})$ sur le cercle ainsi que le théorème local de conjugaison à des rotations. Une remarque fondamentale est que l'action naturelle du groupe $\mathrm{PSL}(2,\mathbb{R})$ sur le cercle $\mathbb{S}$ comme bord du disque de Poincaré contient les rotations. On identifie $\mathbb{S}$ naturellement à $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. On voit donc que l'on peut trouver une copie de $\mathrm{PSL}(2,\mathbb{R})^n$ dans $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$ en prenant le produit des actions sur chacun des facteurs de $\mathbb{T}^n$. Cette copie de $\mathrm{PSL}(2,\mathbb{R})^n$ contient les translations $R_\alpha : \mathbb{T}^n \to \mathbb{T}^n, x \mapsto x + \alpha$ où $\alpha \in \mathbb{T}^n$.

Comme $\mathrm{PSL}(2,\mathbb{R})$ est parfait, on en conclut que $W(\mathbb{T}^n)$, le sous-groupe distingué de $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$ engendré par les R_α, $\alpha \in \mathbb{T}^n$, est contenu dans le sous-groupe des commutateurs de $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$. Comme ce dernier groupe est connexe pour la topologie C^∞, il suffit de voir que l'intérieur du sous-groupe $W(\mathbb{T}^n)$ est non vide. C'est là qu'interviennent les translations diophantiennes. Rappelons que R_γ est une translation diophantienne, si $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{T}^n = \mathbb{T} \times \cdots \times \mathbb{T}$ satisfait une condition diophantienne, c'est-à-dire qu'il existe deux constantes C, d strictement positives et telles que :

$$\forall (k_1 \ldots, k_n) \in \mathbb{Z}^n \setminus \{0\}, \Big\| \sum_{i=1}^n k_i \gamma_i \Big\| \geqslant \frac{C}{\max_{i=1}^n |k_i|^d},$$

où $\|\theta\| = \inf\{|t| \mid t \in \mathbb{R}, t = \theta \mod 1\}$, pour $\theta \in \mathbb{T}$. Si $\gamma \in \mathbb{T}^n$ satisfait à une condition diophantienne, il existe un voisinage V de R_γ dans $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$ pour la topologie C^∞, tel que pour tout $f \in V$, il existe $\lambda(f) \in \mathbb{T}^n$ et $g(f) \in \mathrm{Diff}^\infty_+(\mathbb{T}^n)$, tels que :

$$f = R_{\lambda(f)} \circ g(f)^{-1} \circ R_\gamma \circ g(f).$$

C'est ce que Michel Herman appellera sous diverses formes le théorème de conjugaison locale. Il en résulte que l'intérieur de $W(\mathbb{T}^n)$ contient toutes les translations diophantiennes, ce qui finit de montrer que $\mathrm{Diff}^\infty_+(\mathbb{T}^n)$ est parfait. Le théorème de conjugaison locale est, sous ses premières formes, essentiellement dû à Arnold et Moser et appartient à ce que l'on appelle aujourd'hui la théorie KAM (pour Kolmogorov, Arnold et Moser). Toutefois les versions d'Arnold et Moser à l'époque ne couvraient pas le cas C^∞ sous cette forme et donc Michel Herman, avec la collaboration de Francis Sergeraert, démontre la version C^∞ dans [3], en utilisant un théorème de fonctions implicites dans les espaces de Fréchet qui venait d'être démontré par Sergeraert.

Tout au long de ses travaux Michel Herman retournera encore et encore à ce théorème de conjugaison locale lui en donnant plusieurs démonstrations et en obtenant de nouvelles conséquences. L'action de $\mathrm{PSL}(2,\mathbb{R})$ sur le cercle sera aussi un autre thème récurrent dans ses travaux. On peut dire que l'on commence à voir une première forme de sa méthodologie pour comprendre les phénomènes de la théorie des Systèmes Dynamiques : trouver des exemples explicites dont on peut comprendre toutes les perturbations (éventuellement contraintes dans un sous-espace ad hoc de tous les systèmes). Michel de manière presque obsessionnelle insistait pour comprendre d'abord certains phénomènes dans le cas des tores car disait-il « on peut tout écrire et calculer dans des coordonnées explicites », alors, que la tendance naturelle de beaucoup de dynamiciens est au contraire de trouver des formulations générales dans le cadre des variétés, afin d'éviter de faire des calculs explicites.

Il est peut-être bon de souligner ici, pour le non-spécialiste, le lien étroit des premiers travaux de Herman avec la conjecture d'Arnold sous forme locale. Pour les notations et les concepts utilisés, nous renvoyons à la partie de la notice de John Mather [48] sur Michel où il décrit la conjecture d'Arnold. Dans le cas du cercle $\mathbb{T}$, si $\gamma \in \mathbb{T}$ satisfait à une condition diophantienne, le théorème de conjugaison locale est équivalent au fait que tout difféomorphisme f proche de R_γ et de même nombre de rotation est C^∞ conjugué à R_γ. Ceci résulte du fait que pour $g \in \mathrm{Diff}^\infty_+(\mathbb{T}^1)$, la fonction $\lambda \to \rho(R_\lambda \circ g)$ est continue

localement monotone croissante et même strictement croissante en tout point λ avec $\rho(R_\lambda \circ g) \notin \mathbb{Q}/\mathbb{Z}$.

C'est au début des années 70, que j'ai dû croiser Michel. Mon premier souvenir de lui date d'une séance du séminaire de topologie d'Orsay. Je pense que c'était un exposé de Claude Roger, qui servait de mentor aux topologues débutants qu'étaient Daniel Luçon, Alexis Marin, Alain Prouté, Yves-Marie Visetti et moi-même, nous avions droit à ses cours particuliers. C'était certainement un des premiers exposés auxquels nous assistions sans vraiment comprendre tout ce qui se passait pendant cette nouvelle expérience. L'exposé concernait, il me semble, les classifiants de feuilletages et Michel est intervenu vigoureusement du fond de la salle. À la sortie du séminaire, Lucien Guillou, qui jouait le rôle de vieil initié pour les nouveaux élèves de Larry Siebenmann que nous étions, nous transmit le jugement de notre maître commun Larry Siebenmann : « Herman est follement amoureux des mathématiques », ce qui devait être pris comme le plus grand des compliments.

Michel Herman a obtenu sa thèse de 3e cycle en mars 1972. Elle représente une singularité parmi ses publications. Elle concerne la théorie de Nash sur les variétés (algébriques). Il y donne une caractérisation des sous-variétés d'une variété qui sont des intersections complètes. En particulier, une variété compacte est une intersection complète, algébrique et non-singulière dans un espace euclidien si et seulement si son fibré tangent est stablement trivial. Je n'ai découvert cette thèse de 3e cycle qu'en écrivant cet article. Alain Chenciner, à qui elle est dédiée, m'en a fait parvenir une copie. Bien que nous ayons, surtout à nos premières rencontres, parlé de cette théorie de Nash que je trouve fascinante, Michel ne m'a jamais donné une copie de cette thèse.

Dans la suite de ses travaux, Michel Herman s'est intéressé à la structure de groupe des difféomorphismes analytiques. Dans [6], il montre, par un argument très astucieux de densité des difféomorphismes analytiques dans les difféomorphismes C^∞, que le groupe des difféomorphismes analytiques de $\mathbb{T}^n$ isotopes à l'identité est simple. Encore une fois, la démonstration s'appuie sur des exemples explicites, obtenus à partir de $\mathrm{PSL}(2,\mathbb{R})$, ainsi que sur une version spécifique du théorème de conjugaison locale. À la fin de cet article, Michel Herman donne comme problème ouvert la simplicité du groupe des difféomorphismes analytiques isotopes à l'identité d'une variété compacte quelconque. Ce problème est à ma connaissance toujours ouvert. La principale difficulté par rapport au cas C^∞ est évidemment l'absence de difféomorphismes analytiques à support dans des boules. Pratiquement chaque année, Michel me rappelait ce problème. Comprendre la frontière entre analytique et C^∞ (et plus généralement C^r et $C^{r+\varepsilon}$, voir par exemple [30] pour le cas $r = 3$, ainsi que la notice de Mather [48]) le fascinait.

Les articles [7] et [8] sont dans la même veine. Dans [7], il démontre la simplicité de l'algèbre de Lie des champs de vecteurs analytiques sur une variété compacte, c'est-à-dire la version infinitésimale du problème ouvert mentionné plus haut. Dans [8], il montre que le groupe des difféomorphismes analytiques de $\mathbb{R}$ est simple. Dans le cas C^∞, c'est le groupe des difféomorphismes à support compact qui est simple.

L'article suivant [9] porte comme titre « Sur la conjugaison des difféomorphismes du cercle à des rotations ». Il s'agit d'un article présenté aux « Journées

de géométrie de dimension infinie » organisées par Nicole Desolneux-Moulis à Lyon au printemps 1975. J'y assistai parce que je travaillais à cette époque sur les variétés sur le cube de Hilbert. Je ne me souviens pas d'une conférence faite par Michel (son domaine de recherches était très loin de mes préoccupations mathématiques). J'ai quand même un souvenir d'une discussion avec lui sur les restaurants, ce qui augurait d'une autre passion commune qui nourrirait notre amitié future. Il est intéressant de noter que cet article est présenté à une conférence au printemps 75 avant que Michel n'ait obtenu ses premiers résultats positifs sur la conjecture d'Arnold, et qu'il est paru effectivement en 76, donc après ces premiers résultats. Pourtant, la conjecture d'Arnold n'y est pas mentionnée ! Toutefois, il est clair aujourd'hui qu'il n'est question que de cela. Il reprend l'exemple célèbre :

$$f_b(x) = x + b + a \sin 2\pi x,$$

introduit par Arnold dans [49] et qui a conduit Arnold à poser sa conjecture. On voit déjà apparaître là des éléments qui seront importants dans la résolution de la conjecture d'Arnold. Le critère de conjugaison C^∞ est énoncé : un difféomorphisme f est C^∞ conjugué à une rotation si et seulement si pour tout entier $r \geqslant 1$, les dérivées d'ordre r des itérées f^n, $n \in \mathbb{Z}$, sont équibornées. On y voit introduite une autre « arme » utilisée très efficacement par Michel tout au long de ses travaux : comprendre les conjugués d'une famille de transformations dans leur adhérence, c'est la seconde forme de sa méthodologie. Dans ce cas précis, il annonce que pour $\gamma \in \mathbb{T}$, un nombre de Liouville (c'est-à-dire non rationnel et ne satisfaisant pas à une condition diophantienne), dans la fermeture pour la topologie C^∞ des conjugués C^∞ de R_γ, les difféomorphismes C^1 conjugués à une rotation forment un ensemble maigre au sens de Baire. Une excellente référence sur ce que Michel savait à ce moment-là est l'exposé de Harold Rosenberg [50] au Séminaire Bourbaki en novembre 1975, c'est peut-être par là qu'il faut commencer.

Dans [10], Michel Herman annonce l'existence d'homéomorphismes PL de $\mathbb{T}$ topologiquement conjugués à une rotation irrationnelle, mais dont la mesure invariante est étrangère à la mesure de Lebesgue. Ceci force la conjugaison à être singulière, sa dérivée est Lebesgue presque partout nulle. Il s'agit encore une fois d'une communication présentée à un congrès qui a lieu en 75 et dont les actes paraîtront en 76. La conjecture d'Arnold est encore absente, tout en pointant son nez sous la surface. Cet article est le premier article de Michel, où la théorie ergodique apparaît. Cette théorie est une des composantes essentielles de sa démonstration de la conjecture d'Arnold. Bien que la théorie ergodique apparaisse ici pour la première fois dans ses travaux, Michel en avait déjà développé une connaissance presque encyclopédique au cours de conversations quasi-hebdomadaires avec François Ledrappier, Jean-Marie Strelcyn et Jean-Paul Thouvenot, voir la notice de ce dernier [51]. Durant plusieurs années, il a été un fidèle participant du séminaire de théorie ergodique de Paris VI.

Je peux dater ma relation amicale et mathématique avec Michel d'un jour très précis de septembre 75. Nous nous sommes rencontrés ce jour-là à Offilib, la meilleure librairie scientifique parisienne pendant longtemps. Nous l'écumions tous les deux indépendamment : les livres, surtout de mathématiques, ont été un autre grand ciment de notre amitié. Nous sommes allés prendre un pot,

avec Lucien Guillou que j'avais encore une fois traîné à Offilib, au café en face, celui qu'on appelait la plage à cause des ses chaises en osier. J'étais à la recherche d'un nouveau sujet de recherches, Bob Edwards avait « liquidé » celui sur lequel Larry Siebenmann nous avait aiguillé. Michel nous a parlé des transformations préservant la mesure et des merveilles de la théorie ergodique, que nous ignorions totalement. En une petite heure, il a totalement réorienté ma recherche et lui a donné la direction qu'elle a encore aujourd'hui. À partir de ce jour-là, j'ai vécu les mathématiques de Michel aux premières loges.

Quelques semaines après cette conversation Michel me fit savoir qu'il avait déjà obtenu une réponse positive à la conjecture d'Arnold, bien que l'ensemble des nombres de rotation pour lesquels il pouvait démontrer l'existence de la conjugaison C^∞ soit de mesure nulle. Si ma mémoire est bonne, il m'avait dit qu'il pensait que la solution était correcte, car il l'avait « testée » sur Adrien Douady. En fait les textes de Harold Rosenberg [43] et de Dennis Sullivan [52] sont certainement des comptes rendus plus conformes à ce qui s'est passé (voir aussi la notice de Raphaël Douady [53], qui deviendra par la suite un élève de Michel). Michel fit ensuite un exposé au séminaire de Dennis Sullivan à l'IHÉS. J'assistai évidemment à cet exposé qui eut lieu dans les derniers jours d'octobre 75. Je ne me rappelle plus qui était là, mais, c'était un dialogue entre trois titans : Michel, Dennis Sullivan et Pierre Deligne. C'est certainement un des grands moments mathématiques que j'ai connus. Au bout de 2 heures, Michel n'ayant qu'effleuré la question, Sullivan a proposé de se réunir le 1er novembre, vu que tout le monde était certainement libre, ce jour-là. Nous sommes donc revenu dans un IHÉS désert et difficile d'accès. Michel a tenu la rampe toute la journée, avec une courte interruption pour le déjeuner. Mon souvenir le plus marquant est le suivant : Pierre Deligne n'arrêtait pas de poser des questions ; à un moment, exaspéré, Dennis Sullivan lui a reproché de poser une question dont il lui était clair que Deligne connaissait la réponse. Ce dernier lui a répondu du tac au tac qu'il posait des questions dont il connaissait la réponse afin de se donner le temps de réfléchir aux points qu'il ne comprenait pas. Ce fut un très grand jour pour Michel. Cette participation de Pierre Deligne a certainement aidé Michel pour la suite. Pierre Deligne a amené des améliorations importantes à la preuve du théorème fondamental. Ces résultats ont été annoncés dans une note aux CRAS [11]. L'exposé [54] de Pierre Deligne au séminaire Bourbaki en février 1976 est le meilleur texte introductif aux premiers travaux de Herman sur la conjecture d'Arnold.

Il est amusant de constater que l'exposé [50] de Rosenberg a été fait mi-novembre 75, quelques semaines après que Michel donne une solution partielle de la conjecture d'Arnold. Dans ce texte, ce fait n'est pas mentionné. Je ne me souvenais plus si Harold a mentionné dans son exposé oral les derniers résultats de Michel, mais, il m'a confirmé qu'il l'a fait. L'exposé de Deligne [54] a été fait à la séance suivante du séminaire Bourbaki.

La thèse d'État [12] est soutenue le 23 avril 1976. On y trouve une démonstration de la conjecture d'Arnold pour les nombres de rotations de type borné qui forment un ensemble de mesure de Lebesgue nulle mais de dimension de Hausdorff 1. On y trouve aussi, les versions complètes de [9] et [10], ainsi que les articles [5], [6], [7] et [8], éventuellement complétés par des démonstrations.

Il m'est difficile de donner une idée sur la démonstration de la conjecture d'Arnold qui soit courte et utile. Toutefois, on peut en donner quelques éléments cruciaux. On dit que $\alpha \in \mathbb{T}$ est de type Roth, si on a :

$$\forall \varepsilon > 0,\ \inf_{p/q \in \mathbb{Q}} q^{2+\varepsilon} \left\| \alpha - \frac{p}{q} \right\| > 0.$$

Cette condition est satisfaite par presque tout nombre. Une version précise du théorème de conjugaison local, prouvée dans [12] et s'inspirant d'un travail de Helmut Rüssmann [55], montre que si $f \in \mathrm{Diff}^{\infty}(\mathbb{T})$ a un nombre de rotation $\rho(f)$ de type Roth et est assez proche de $R_{\rho(f)}$ dans la topologie $\mathrm{C}^{2+\varepsilon}$, pour un $\varepsilon > 0$, alors, le difféomorphisme f est C^{∞} conjugué à $R_{\rho(f)}$. Il reste alors à conjuguer un difféomorphisme f pour le rendre proche de $R_{\rho(f)}$. Pour cela, Herman introduit une suite g_n de difféomorphismes du cercle donnés par $\tilde{g}_n = \frac{1}{n}\sum_{i=0}^{n-1}(\tilde{f}^i - ia)$, où $\tilde{g}_n$ et $\tilde{f}$ sont les relevés de g_n et f au revêtement universel $\mathbb{R}$ de $\mathbb{T}$, et $a = \lim_{n\to\infty} \tilde{f}^n(x)/n$ est donc un représentant ad hoc dans $\mathbb{R}$ de $\rho(f) \in \mathbb{T}$. Les g_n sont des difféomorphismes C^{∞} et convergent uniformément vers l'homéomorphisme g donné par le théorème de Denjoy qui conjugue topologiquement f à $R_{\rho(f)}$; de plus, on a $\tilde{g}_n \circ \tilde{f} \circ \tilde{g}_n^{-1} = R_a + (\tilde{f}^n - \mathrm{Id} - na)/n \circ \tilde{g}_n^{-1}$. Il suffit donc de voir que $(\tilde{f}^n - \mathrm{Id} - na)/n \circ \tilde{g}_n^{-1}$ tend vers zéro dans la $\mathrm{C}^{2+\varepsilon}$ topologie. Il est clair que la dérivée seconde de cette dernière application fait intervenir en particulier le quotient de la dérivée seconde $D^2 f^n$ de l'itérée d'ordre n par la dérivée Dg_n. Comme on peut écrire :

$$D^2 f^n = Df^n \cdot \Big(\sum_{i=1}^{n-1} \frac{D^2 f}{Df} \circ f^i \cdot Df^i \Big),$$

$$Dg_n = \sum_{i=1}^{n-1} Df^i,$$

on voit que l'on fait intervenir par ces quotients des quantités qui peuvent être traitées par le théorème ergodique de Chacon-Ornstein. Ce qui montre l'intervention naturelle de la théorie ergodique. En fait, Herman démontre que la convergence de $(\tilde{f}^n - \mathrm{Id} - na)/n \circ \tilde{g}_n^{-1}$ vers zéro dans la $\mathrm{C}^{2+\varepsilon}$ topologie est vraie dès que f est $\mathrm{C}^{1+\varepsilon'}$ conjugué à une rotation pour un $\varepsilon' > \varepsilon$. Il reste à voir que f est $\mathrm{C}^{1+\varepsilon'}$ conjugué à une rotation ; c'est surtout là que les conditions arithmétiques sur $\rho(f)$ interviennent. Évidemment cette esquisse ne dit rien de la finesse des arguments d'estimation utilisés pour obtenir ces résultats.

À la suite de sa thèse Michel Herman a séjourné à Warwick pendant le printemps 76. La publication [13] est de cette époque. Il y considère l'équation fonctionnelle $\psi - \psi \circ \mathbb{R}_\alpha = \varphi$, où $\psi, \varphi : \mathbb{T} \to \mathbb{C}$ sont mesurables. Anosov a montré que pour certains choix de φ dans L^1, cette équation a des solutions ψ mesurables, mais aucune d'elles n'est dans L^1. Dans cette note, Michel Herman montre que si $\varphi \in L^1$ a une série de Fourier lacunaire, alors, il y a des solutions ψ dans L^2.

Pendant ce séjour à Warwick, Michel a aussi complètement résolu la conjecture d'Arnold, puisqu'il trouve un ensemble de mesure pleine dans $\mathbb{T}$, qu'il appelle d'ailleurs A et tel que tout $f \in \mathrm{Diff}^{\infty}(\mathbb{T})$ avec $\rho(f) \in A$ est conjugué à une rotation. Ce résultat est annoncé dans [14]. La démonstration complète se

trouve dans [18], la version améliorée des parties non publiées précédemment de sa thèse [12]. La lecture de [18] montre bien le talent de Michel pour étudier une question de manière exhaustive et lui donner le meilleur traitement possible. Notons aussi qu'au chapitre 9 de [18], Herman donne une démonstration de la conjecture d'Arnold qu'il qualifie d'élémentaire puisqu'elle n'utilise plus un théorème de conjugaison local.

Une caractérisation complète de l'ensemble des nombres de rotations pour lesquels la conjugaison de Denjoy est toujours C^∞ a été donnée par Jean-Christophe Yoccoz [56].

Les deux articles suivants sont parus dans les comptes-rendus de la « Troisième École de Mathématiques d'Amérique Latine » qui a eu lieu en juillet 1976 à Rio de Janeiro à l'IMPA, dans les anciens locaux près de la place Tiradentes, au cœur du quartier chaud de Rio. J'y assistai sur la recommandation de Michel. J'ai évidemment des souvenirs extrêmement vifs de ce premier congrès dans le monde des systèmes dynamiques, en particulier, du premier déjeuner où j'ai été ébloui par une conversation mathématique très animée entre trois mathématiciens et surtout menée par un gamin volubile qui buvait énormément de Coca-Cola : Charles Pugh, Sheldon Newhouse et le « gamin » Ricardo Mañé. La baie de Rio vue du Pain de Sucre et du Corcovado dès les premiers jours m'a tellement marqué, que j'y suis retourné le week-end avec Michel qui n'y avait pas encore été. J'ai souvenir d'un Michel radieux au sommet du Pain de Sucre, se protégeant du soleil. Je me souviens surtout de trois conférences de mathématiques : une de José Adem qui faisait un cours en espagnol [57], la conférence de Michel sur l'invariant de Godbillon-Vey et une conférence de Jürgen Moser (si impressionnant dans cette première rencontre comme dans les suivantes) qui, il me semble, portait sur KdV dont j'entendais parler pour la première fois (ma mémoire me trahit-elle encore une fois, aucun des papiers publiés par Moser dans les proceedings ne porte sur cette question ?). Je vois Michel, à la sortie de sa conférence, discutant avec Jürgen Moser, qui avait ce regard souriant et si infiniment doux dont il gratifiait ceux dont il appréciait les mathématiques. Je vois aussi Michel avec Jaco Palis, qui l'entourait de cette chaleur et de cette sollicitude (peut-être) brésilienne (mais dans laquelle je reconnais aussi une composante moyen-orientale), dont il fait si souvent preuve. Michel a été très heureux là ; ceux qui lui sont proches savent combien les moments de bonheur ont été précieux dans sa vie. Son plaisir d'être à Rio n'a fait que s'accroître au fil des années, jusqu'à décider d'y aller début 2001 pour plusieurs années. Quel dommage qu'il n'ait pas pu aller vivre à Rio, où il était ces dernières années toujours moins malheureux qu'ailleurs.

Pour en revenir au contenu mathématique des publications de Michel, l'article [15] étudie les chemins $(f_t)_{t\in[0,1]}$ de classe C^1 dans $\mathrm{Diff}^\infty(\mathbb{T})$ (par exemple). On désigne par $M(f_t)$ l'ensemble des points $t \in [0,1]$ tels que f_t soit C^∞ conjugué à une rotation. Si $\rho(f_t)$ n'est pas une constante, Herman montre que la mesure de Lebesgue de $M(f_t)$ est > 0. De plus, cette mesure tend vers 1 si la famille f_t tend en topologie C^3 vers la famille des rotations R_t, $t \in [0,1]$. En fait, un raisonnement élémentaire montre que l'on a :

$$\forall f \in \mathrm{Homeo}_+(\mathbb{T}),\ \forall \alpha \in \mathbb{T},\ \|\rho(f) - \alpha\| \leqslant \sup_{x\in\mathbb{T}} \|f(x) - R_\alpha(x)\|.$$

Si $g = h^{-1}R_\alpha h$ avec h un C^1-difféomorphisme, il en résulte que :

$$\forall f \in \mathrm{Homeo}_+(\mathbb{T}), \|\rho(f) - \rho(g)\| \leqslant \sup_{x\in T} |h'(x)| \sup_{x\in\mathbb{T}} \|f(x) - g(x)\|.$$

On peut, alors, écrire $M(f_t) = \cup_{k\in\mathbb{N}} D_k$, où l'ensemble D_k est formé par les $t \in [0,1]$ tels que $f_t = h^{-1}R_\alpha h$, avec h tel que $\sup_{t\in T} |h'(t)| \leqslant k$. Si on pose $C = \sup_{(t,x)\in[0,1]\times\mathbb{T}} \partial f_t/\partial t(x)$, l'application continue $t \mapsto r(t) = \rho(f_t)$ est lipschitzienne sur D_k de constante Ck, donc $m(r(D_k)) \leqslant Ckm(D_k)$. L'image $r(M(f_t))$ est un intervalle $[a,b]$ non réduit à un point. Par la conjecture d'Arnold $r(M(f_t)) = \cup_{k\in\mathbb{N}} r(D_k) \supset [a,b] \cap A$, qui est de mesure $b - a > 0$. Il en résulte que l'on a $0 < m(r(D_k)) \leqslant Ckm(D_k)$, pour au moins un indice k, et donc $m(M(f_t)) > 0$. Le fait que $m(M(f_t)) \to 1$ quand f_t tend vers la famille des rotations R_t, $t \in [0,1]$, résulte encore une fois d'une version précise du théorème de conjugaison local.

L'article [16] répond positivement à une question de Rosenberg et Thurston : l'invariant de Godbillon-Vey d'un feuilletage par plan du tore $\mathbb{T}^3$, de classe C^2 et transversalement orientable est nul. Cet invariant de Godbillon-Vey a été le premier invariant numérique des feuilletages qui ne venait pas des classes caractéristiques du fibré normal d'un feuilletage de codimension 1, puisque Thurston avait montré qu'il pouvait prendre n'importe quelle valeur réelle en prenant des exemples ad hoc de feuilletage de codimension 1 sur des variétés de dimension 3 « plutôt hyperboliques » que plates. Le travail de Herman a été une étape importante pour la compréhension de cet invariant de Godbillon-Vey jusqu'aux travaux de Duminy. Il faut toutefois ajouter qu'au même moment que Herman, Guy Wallet obtenait aussi un résultat un peu plus faible de ce type [58]. Pour démontrer la nullité de l'invariant de Godbillon-Vey, Herman utilise un travail de Rosenberg et Roussarie : un tel feuilletage sur $\mathbb{T}^3$ s'obtient par une opération de type suspension, à partir de deux difféomorphismes $f, g \in \mathrm{Diff}^2_+(\mathbb{T})$ qui commutent. Comme les feuilles du feuilletage sont des plans, les nombres de rotation de f et g sont irrationnels et rationnellement indépendants. Ensuite, utilisant une classe de cohomologie du groupe $\mathrm{Diff}^2_+(\mathbb{T})$ introduite par Thurston pour exprimer l'invariant de Godbillon-Vey dans ce cadre, il montre que l'invariant est donné par :

$$-\int_{\mathbb{T}} \Big[\frac{\log Dg^n}{n} \log Df^{-1} + \frac{\log Dg^{-n}}{n} \log Df\Big]\, dm$$

où m est la mesure de Lebesgue et n est un entier quelconque. Il reste à observer que $\log Dg^{\pm n}/n \to 0$, quand $n \to \infty$. Or, on peut écrire :

$$\frac{\log Dg^n}{n} = \frac{1}{n}\sum_{i=0}^{n-1} (\log Dg) \circ g^i,$$

par l'unique ergodicité de g (de nombre de rotation irrationnel), on en déduit que $\log Dg^n/n$ tend vers une constante, cette constante est nécessairement nulle car $\int_{\mathbb{T}} Dg^n\, dm = 1$ (puisque g^n est un difféomorphisme de $\mathbb{T}$).

Mike Shub était en 76-77 en sabbatique à Orsay et il y avait une activité assez forte en systèmes dynamiques autour de son cours. C'est donc tout naturellement en automne 76 que Michel est venu à Orsay nous parler de son dernier

travail : la construction d'un difféomorphisme minimal sur $\mathbb{S}^3$, c'est-à dire d'un difféomorphisme dont toute orbite est dense. C'était une construction qui était assez subtile et où des propriétés fines de théorie de la mesure intervenaient. À la suite de son exposé, il m'avait encouragé à regarder le cas des variétés produits du type $\mathbb{T} \times M$. M'inspirant du travail d'Anosov-Katok [59] que Michel m'avait fait découvrir, je réussis à traiter ce cas dans les mois suivants. Finalement, nous joignîmes nos forces et les derniers arguments ont été échangés dans une conversation téléphonique entre Orsay où je me trouvais et l'Institut Mittag Leffler où Michel a dû passer deux mois au printemps 77. Nous avions obtenu qu'une variété compacte, munie d'une action localement libre du cercle, admettait toujours un difféomorphisme de classe C^∞ dont toute orbite est dense [18]. En fait, si on désigne par R_t, $t \in \mathbb{T}$, les éléments de l'action de $\mathbb{T}$ sur la variété M, alors, dans la fermeture de l'orbite $O^\infty(\mathbb{T}) = \{f \circ R_t \circ f^{-1} \mid f \in \mathrm{Diff}^\infty_+(M)\}$ pour la topologie C^∞, les difféomorphismes minimaux forment un G_δ dense. Ce travail s'inspire fortement de [59] et, comme nous l'apprîmes par la suite, l'existence de difféomorphismes minimaux était due à Anatole Katok, qui l'avait déjà annoncée dans [60], voir [61], §11.44 pages 102–105, pour la preuve. Notre travail a beaucoup popularisé cette méthode d'Anosov-Katok en la mettant à l'œuvre dans son cas le plus simple.

C'est malheureusement à cette époque que Michel a été victime d'une agression qui devait le laisser perpétuellement handicapé. Jusqu'à la fin de sa vie, il a gardé un genou très douloureux et il devait utiliser une béquille. L'article [18] a été rédigé par Michel à la fin du printemps 77, pendant qu'il était hébergé chez Harold Rosenberg, à sa sortie de l'hôpital après cette agression. Avec François Laudenbach, nous allions le voir en lui emmenant des gâteaux au chocolat de chez Lenôtre, un de nos vices communs. Cet article fut publié dans les comptes rendus du colloque de systèmes dynamiques à Varsovie de l'été 77. Comme Michel n'avait pas pu se rendre à ce colloque, à la suite de ses problèmes de genou, et que je n'avais pas prévu d'y aller, c'est Mike Shub qui a présenté le papier pour nous.

C'est à la rentrée 77, que Pierre Arnoux et Jean-Christophe Yoccoz sont devenus les deux premiers élèves de Michel (Larry Siebenmann est resté mon patron de thèse. Jusqu'à ce jour, il s'est toujours intéressé de très près à mes progrès). Michel a fait travailler Pierre sur les échanges d'intervalles et Jean-Christophe sur les difféomorphismes du cercle, voir leurs notices [62, 63].

L'exposé de Michel Herman [19] au Congrès International des Mathématiciens, qui eut lieu à Helsinki pendant l'été 1978, donne un panorama de ses travaux qu'il est intéressant de consulter. Il y réfère déjà à des travaux qu'il n'a pas publié de son vivant et qui ont pourtant circulé sous forme de notes manuscrites. Pour ce premier exemple, il s'agit des constructions de difféomorphismes ergodiques ayant diverses propriétés.

En 78-79, Michel a co-organisé, avec François Ledrappier et Jean-Paul Thouvenot le séminaire de théorie ergodique de Paris VI sur les travaux de Yasha Pesin [64, 65] qui permettent d'étendre la théorie des difféomorphismes d'Anosov et celles des difféomorphismes Axiome A, introduite par Smale, en mettant en lumière l'universalité du rôle des exposants de Lyapunov et leur importance dans les calculs d'entropie. Il s'agit là d'une pierre angulaire de la

théorie moderne des systèmes dynamiques. Ce séminaire, ainsi que les travaux de David Ruelle [66], ont permis de diffuser rapidement les travaux de Pesin, au delà du cercle des dynamiciens d'Europe de l'Est. Des notes manuscrites ont été distribuées pendant ce séminaire et ont beaucoup circulé ensuite. Une des forces des travaux de Pesin est d'avoir réalisé que les variétés stables existaient presque sans aucune condition, ce qui permet de trouver des coordonnées produits stable-instable mesurables, dans lesquels la théorie ergodique pouvait intervenir de manière bien plus efficace. Les exposés de Michel Herman, Jean-Christophe Yoccoz et moi-même concernaient surtout la partie existence de variété stable. Ils montraient comment adapter les arguments du cas uniformément hyperbolique tels que nous les avions appris des travaux d'Irwin [67] et du cours de Mike Shub [68]. La version finale en anglais des notes [20] est parue en 1983, dans les comptes-rendus du colloque d'inauguration des nouveaux locaux de l'IMPA à Rio en 1981. Elle avait été rédigée par Jean-Christophe et moi-même dans un de ces bureaux « cagibi » qui servaient aux thésards de l'Université de Warwick, où nous séjournions pendant l'été 1980.

L'article [21] (dédié à Laurent Schwartz pour son 65ème anniversaire) utilise la théorie de Pesin pour montrer l'existence d'un difféomorphisme minimal d'entropie non nulle sur une variété compacte. Michel considérait cet article comme un de ses tout meilleurs. C'est effectivement un véritable tour de force : jusque là les difféomorphismes minimaux avaient plutôt tendance à être d'entropie nulle (translations sur les tores, flots horocycliques, exemples donné par les techniques Anosov-Katok comme dans [17]). Par contre, l'entropie non nulle se rencontrait plutôt dans les situations d'hyperbolicité à la Anosov-Smale, ce qui implique l'existence d'orbites périodiques. La construction de [22] se fait à l'aide d'un produit croisé (skew-product) au-dessus d'une translation irrationnelle de $\mathbb{T}$. On considère l'application $A : \mathbb{T} \to \mathrm{SL}(2,\mathbb{R})$, de classe C^∞, définie par :

$$A(\theta) = A_\theta = \begin{pmatrix} \cos 2\pi\theta & -\sin 2\pi\theta \\ \sin 2\pi\theta & \cos 2\pi\theta \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix},$$

où $\lambda > 1$ est fixé. On se donne $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ un sous-groupe discret tel que $M_1 = \mathrm{SL}(2,\mathbb{R})/\Gamma$ soit le fibré unitaire tangent d'une surface compacte à courbure négative (en particulier $-\mathrm{Id} \in \Gamma$). Le groupe $\mathrm{SL}(2,\mathbb{R})$ opère sur M_1 à gauche. Pour tout $\alpha \in \mathbb{T}$, on peut définir le produit croisé $F_\alpha : \mathbb{T} \times M_1 \to \mathbb{T} \times M_1, (\theta, x) \mapsto (\theta + \alpha, A_\theta x)$. Michel Herman montre que pour $\alpha \in \mathbb{T}$, l'entropie topologique de F_α est $\geqslant 2\log(\lambda/2 + 1/2\lambda) > 0$, et pour un G_δ dense de α, le difféomorphisme F_α est minimal. Je ne résiste pas à la tentation d'en donner pratiquement la démonstration, car elle est bien représentative d'un style de mathématiques qu'affectionnait particulièrement Michel. En fait, la mesure de Haar sur $\mathrm{SL}(2,\mathbb{R})$, qui est unimodulaire, donne sur M_1 une mesure de probabilité qui vient d'une forme volume analytique. Cette mesure ν est préservée par l'action de $\mathrm{SL}(2,\mathbb{R})$, grâce à l'unimodularité. Il en résulte que F_α préserve la mesure produit $\mu = m \otimes \nu$, où m est la mesure de Lebesgue sur $\mathbb{T}$. C'est l'entropie métrique de F_α pour μ que l'on peut

minorer par $2\log(\lambda/2+1/2\lambda)>0$. Par la théorie de Pesin, il suffit de minorer la quantité :

$$\lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{T}\times M}\log\|D_{(\theta,x)}F_\alpha^n\|\,d\mu(\theta,x).$$

On a $F_\alpha^n(\theta,x)=(\theta+n\alpha,A_{\alpha,\theta}^n x)$, où $A_{\alpha,\theta}^n=A_{\theta+(n-1)\alpha}\dots A_{\theta+\alpha}A_\theta$. En ne regardant que la norme de la restriction de l'action de la dérivée de F_α sur la partie tangente à M_1 dans le produit $M=\mathbb{T}\times M_1$ et en comparant des normes, on voit que la quantité ci-dessus se minore par :

$$2\lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{T}\times M_1}\log\|A_{\alpha,\theta}^n\|\,dm(\theta).$$

Le théorème ergodique sous-additif montre que la limite $a_\alpha(\theta)$ de la suite $n^{-1}\log\|A_{\alpha,\theta}^n\|$ existe pour presque tout $\theta\in\mathbb{T}$, que la fonction a_α est invariante par R_α et que $\lambda_+(\alpha,A)=\int_{\mathbb{T}}a_\alpha\,dm$ est $\inf_{n\geqslant 1}\frac{1}{n}\int_{\mathbb{T}\times M_1}\log\|A_{\alpha,\theta}^n\|\,dm(\theta)$, la limite de la suite sous-additive $\frac{1}{n}\int_{\mathbb{T}\times M_1}\log\|A_{\alpha,\theta}^n\|\,dm(\theta)$. Comme A est à valeurs dans $\mathrm{SL}(2,\mathbb{R})$, la fonction a_α est $\geqslant 0$. Pour α irrationnel, l'ergodicité de R_α montre que $a_\alpha=\lambda_+(\alpha,A)$ presque partout. Si $\alpha=p/q \mod 1$, alors $a_\alpha(\theta)$ existe partout et est égale $q^{-1}\log|\beta_{p,q}(\theta)|$, où $\beta_{p,q}(\theta)$ est la valeur propre de plus grand module de la matrice $A_{p/q,\theta}^q$. Ensuite, pour α irrationnel, par la stricte ergodicité de R_α, on montre que $n^{-1}\log[\sup_\theta\|A_{\alpha,\theta}^n\|]$ converge aussi vers $\lambda_+(\alpha,A)$. En prenant, une diagonalisation sur $\mathbb{C}$ de la partie rotation de A_θ et en effectuant des conjugaisons dans $\mathrm{SL}(2,\mathbb{R})$ on se ramène à un cocycle $C_{\alpha,\theta}^n$ à valeurs matricielles, avec $\|C_{\alpha,\theta}^n\|\leqslant K\|A_{\alpha,\theta}^n\|$, où K est une constante, et tel que $\int_{\mathbb{T}}C_{\alpha,\theta}^n\,dm(\theta)$ est une matrice dont un des coefficients est $(\lambda/2+1/2\lambda)^n$. Il en résulte aisément que $\lambda_+(\alpha,A)\geqslant\log(\lambda/2+1/2\lambda)$, pour tout α irrationnel. Comme $\lambda(\alpha,A)$ est semi-continue supérieurement en α (car c'est la borne inférieure des quantités $n^{-1}\int_{\mathbb{T}\times M}\log\|A_{\alpha,\theta}^n\|\,dm(\theta)$, qui sont continues en α), on en conclut que $\lambda_+(\alpha,A)\geqslant\log(\lambda/2+1/2\lambda)$, pour tout $\alpha\in\mathbb{T}$. Ce qui finit de montrer que l'entropie de F_α est toujours >0.

Il reste à voir que pour un G_δ dense de $\alpha\in\mathbb{T}$, le difféomorphisme F_α de $M=\mathbb{T}\times M_1$ est minimal. Comme M admet une base dénombrable d'ouverts non vides de la forme $I\times V$, où I est un intervalle non vide de $\mathbb{T}$ et V est un ouvert de M_1, il suffit de voir que l'ouvert $\mathcal{U}_{I\times V}=\{\alpha\in\mathbb{T}\mid M=\bigcup_{n\in\mathbb{Z}}F_\alpha^n(I\times V)\}$ de $\mathbb{T}$ est dense. On remarque alors que cet ouvert $\mathcal{U}_{I\times V}$ contient l'intersection de l'ouvert $\mathcal{V}_{I\times V}=\{\alpha\in\mathbb{T}\mid\exists\theta_0\in\mathbb{T},\{\theta_0\}\times M_1\subset\bigcup_{n\in\mathbb{Z}}F_\alpha^n(I\times V)\}$ avec les irrationnels $\mathbb{T}\setminus\mathbb{Q}/\mathbb{Z}$. En effet, soit α dans cette intersection comme $\bigcup_{n\in\mathbb{Z}}F_\alpha^n(I\times V)$ est ouvert et M_1 compact, il existe un intervalle ouvert $J\ni\theta_0$ avec $J\times M_1\subset\bigcup_{n\in\mathbb{Z}}F_\alpha^n(I\times V)$. Il en résulte que l'ensemble invariant $\bigcup_{n\in\mathbb{Z}}F_\alpha^n(I\times V)$ contient l'orbite $\bigcup_{n\in\mathbb{Z}}F_\alpha^n(J_1\times M_1)=[\bigcup_{n\in\mathbb{Z}}R_\alpha^n(J)]\times M_1$, comme J est un ouvert non vide et que α est irrationnel, on a $\bigcup_{n\in\mathbb{Z}}R_\alpha^n(J)=\mathbb{T}$. On est ramené à voir que $\mathcal{V}_{I\times V}$ est dense dans $\mathbb{T}$. On va montrer qu'il contient tous les rationnels p/q, avec p et q premiers entre eux et $1/q$ plus petit que la longueur de I. On a $F_{p/q}^q(\theta,x)=(\theta,A_{p/q,\theta}^q(x))$, il suffit de voir qu'il existe $\theta\in I$ tel que l'action de la matrice $A_{p/q,\theta}^q$ soit minimale sur M_1. En fait, comme $A_{p/q,\theta}^q=A_{\theta+(q-1)p/q}A_{\theta+(q-2)p/q}\dots A_{\theta+p/q}A_\theta$ et $A_{p/q,\theta+i/q}^q=A_{\theta+i/q+(q-1)p/q}A_{\theta+i/q+(q-2)p/q}\dots A_{\theta+i/q+p/q}A_{\theta+i/q}$ sont deux

produits de matrices inversibles qui diffèrent par une permutation cyclique, la minimalité de l'action $A^q_{p/q,\theta}$ est équivalente à celle de l'action de n'importe laquelle des matrices $A^q_{p/q,\theta+i/q}$. Or $1/q$ est inférieur à la longueur de I et donc pour tout $\theta \in \mathbb{T}$, l'un des $\theta + i/q$ sera dans I, on voit donc que l'on doit montrer qu'il existe $\theta \in \mathbb{T}$ tel que l'action de la matrice $A^q_{p/q,\theta}$ sur M_1 soit minimale. Herman utilise alors la minimalité du flot horocyclique sur M_1. L'action du flot horocyclique h_t sur M_1 est l'action des matrices $\left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$, $t \in \mathbb{R}$. Depuis, les travaux de Hedlund, on sait que le flot horocyclique est minimal sur M_1 et même que l'action de chaque difféomorphisme h_{t_0} sur M_1, pour tout $t_0 \neq 0$, est minimale. Une matrice parabolique de $\mathrm{SL}(2,\mathbb{R})$ est une matrice conjuguée dans $\mathrm{SL}(2,\mathbb{R})$ à une des h_t, avec $t \neq 0$. Une matrice de $\mathrm{SL}(2,\mathbb{R})$ est parabolique si elle n'est pas l'identité et ses deux valeurs propres sont confondues et égales à 1. Il faut alors trouver un θ tel que $A^q_{p/q,\theta}$ soit parabolique. Le groupe $\mathrm{SL}(2,\mathbb{R})$ agit bien sûr sur l'espace projectif $\mathbb{P}_1(\mathbb{R}) \approx \mathbb{T}$ des droites vectorielles de $\mathbb{R}^2$, on note par $\hat{B} : \mathbb{T} \to \mathbb{T}$, l'action de $B \in \mathrm{SL}(2,\mathbb{R})$ sur $\mathbb{P}_1(\mathbb{R}) \approx \mathbb{T}$. Tout $\hat{B}$ est un homéomorphisme de $\mathbb{T}$ homotope à l'identité, on peut donc parler de son nombre de rotation $\rho(\hat{B})$. On a $\rho(\hat{B}) = 0$ si et seulement si $\hat{A}$ admet un point fixe c'est-à-dire si A a une valeur propre réelle. Revenons alors à $A^q_{p/q,\theta}$, par ce que l'on a vu plus haut $a_{p/q}(\theta)$ existe partout et est égale $q^{-1}\log|\beta_{p,q}(\theta)|$, où $\beta_{p,q}(\theta)$ est la valeur propre de plus grand module de la matrice $A^q_{p/q,\theta}$, et $\int_{\mathbb{T}} a_{p/q}(\theta)\,d\theta = \lambda_+(p/q, A) > 0$. Il en résulte qu'il existe θ telle que $A^q_{p/q,\theta}$ admette une valeur propre strictement plus grande que 1 en module, cette valeur propre est alors réelle, car $A^q_{p/q,\theta} \in \mathrm{SL}(2,\mathbb{R})$. Notons alors par O le sous-ensemble de $\mathbb{T}$ formé par les θ tels que $A^q_{p/q,\theta}$ ait une valeur propre strictement plus grande que 1 en module. C'est un ouvert de $\mathbb{T}$, de plus $\rho(\hat{A}^q_{p/q,\theta}) = 0$, pour tout $\theta \in O$. On a $O \neq \mathbb{T}$, car $\hat{A}_\theta = R_{2\theta}\hat{\Lambda}$, où $\Lambda = \left(\begin{smallmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{smallmatrix}\right)$ et donc $\theta \mapsto \rho(\hat{A}_\theta)$ est surjectif. Prenons $\theta_0 \in \bar{O} \setminus O$ alors $\rho(A^q_{p/q,\theta_0}) = 0$ et $A^q_{p/q,\theta_0}$ a une valeur propre réelle de module 1. Montrons que $\hat{A}^q_{p/q,\theta_0}$ n'est pas l'identité. Puisque $\hat{A}_\theta = R_{2\theta}\hat{\Lambda}$, il n'est pas difficile de voir que, pour tout $t \in \mathbb{P}_1(\mathbb{R}) \approx \mathbb{T}$ fixé, l'application $\theta \mapsto \hat{A}_\theta(t)$ est une application de $\mathbb{T}$ dans lui-même (localement) strictement croissante. Par composition, cette propriété est encore vrai pour $\hat{A}^q_{p/q,\theta_0}$ il en résulte que $\theta \mapsto \rho(A^q_{p/q,\theta})$ est strictement croissante en tout point θ tel que $\hat{A}^q_{p/q,\theta}$ est l'identité. Comme $\theta_0 \in \bar{O} \setminus O$ et que $\theta \mapsto \rho(A^q_{p/q,\theta})$ est constante sur $\bar{O}$, on en conclut que $A^q_{p/q,\theta_0}$ a ses deux valeurs propres réelles et égales (donc égales soit à 1 soit à -1) et qu'elle n'est pas diagonale. Comme $M_1 = \mathrm{SL}(2,\mathbb{R})/\Gamma$, avec $-\mathrm{Id} \in \Gamma$, on voit que l'action de $A^q_{p/q,\theta_0}$ est l'action d'une matrice parabolique et qu'elle est donc minimale.

C'est en 1980 que Michel a commencé son séminaire de Systèmes Dynamiques à l'École polytechnique. Ce séminaire s'est poursuivi en différents endroits jusqu'à la fin de sa vie. J'y ai découvert une grande partie de toutes les nouvelles avancées des Systèmes Dynamiques (du moins avant mon départ pour les États-Unis). Je crois que ce séminaire a non seulement marqué tous ses participants, mais qu'il a eu un rôle très important pour la dissémination des connaissances en dehors du cercle des participants réguliers.

Au début des années 80, Michel Herman a fait un cours à l'École Normale Supérieure de la rue d'Ulm sur les théorèmes KAM. Il a produit des notes manuscrites qui ont largement circulé et exerce jusqu'à aujourd'hui une grande influence. Une trace publiée de ce cours se trouve dans un séminaire Bourbaki de Jean-Benoît Bost [69].

Dans [22], Michel, avec Jean-Christophe Yoccoz, établit, pour des corps non-archimédiens, les analogues du théorème de Siegel sur la linéarisation des germes de fonctions holomorphes, ainsi qu'un analogue du théorème de conjugaison local d'Arnold et Moser.

Ce dernier travail [22] comme [20] a été publié dans les proceedings de la conférence qui eut lieu à Rio de Janeiro l'été 1981 pour l'inauguration des nouveaux locaux de l'IMPA. C'était une très grande conférence qui a couvert un très large spectre des Systèmes Dynamiques. C'était aussi le premier de nombreux séjours de Michel dans le quartier de Leblon, dont il apprit à connaître tous les restaurants et surtout la churrascaria Plataforma qu'il affectionnait particulièrement et où il m'a entraîné plusieurs fois. Nous sommes allés ensemble un des premiers jours de nouveau au sommet du Pain de Sucre, ce mois de juin 1981 avait été particulièrement pourri à Paris et lézarder au soleil de Rio nous avait remonté le moral. Je me souviens aussi d'un dîner avec Michel, Charles Conley (en grande forme) et Eddy Zehnder.

L'article [23] a été publié en 1983, mais, nous en avions corrigé la première version pendant l'été 1982 assis tous les deux sur le muret du jardin du CIRM à Luminy pendant le colloque de théorie ergodique que Michel avait organisé avec François Ledrappier. C'est un article foisonnant, où Michel récolte tous les bénéfices possibles des méthodes qu'il a mises au point dans [21]. Des versions abstraites et générales de minorations d'exposants de Lyapunov, ainsi que des études fines de produits croisés lui permettent de construire plein d'exemples et de contrexemples qui délimitent les domaines de ses théorèmes préférés : Siegel, Arnold-Moser, Denjoy. Il introduit dans ce papier une notion de nombre de rotations pour les produits croisés d'homéomorphismes du cercle au-dessus d'un homéomorphisme minimal qui se révèle bien utile (elle apparaît aussi dans des travaux de R. Johnson et Moser [70]). Les exposés que Michel devait faire dans son séminaire en décembre 2000 (et qui n'ont donc pas eu lieu) concernaient les exposants de Lyapunov, ce qui montre bien que ce sujet a fait partie des préoccupations de Michel jusqu'au bout.

C'est vers cette époque, que pour la seule et unique fois de sa vie Michel s'est fâché avec moi. Il reprochait à une tierce personne de ne pas le remercier suffisamment dans son dernier papier. Comme je les aimais bien tous les deux, j'ai eu le malheur de lui dire que tout ça n'avait pas tant d'importance et que ce n'était pas si grave que cela, il s'est alors fermé comme une huître. J'avais visiblement touché une corde bien plus sensible que je ne le pensais. Quelques jours après, j'ai eu droit à un coup de téléphone de Michel où il m'a fait la morale sur nous autres petits jeunes (seules huit années nous séparent) qui ne reconnaissent pas les mérites de leurs aînés qui les ont tant aidés et ainsi de suite. Afin de finir de me clouer le bec, il a ajouté que d'ailleurs François (Laudenbach) lui donnait raison. J'ai dû balbutier quelques mots, après lesquels, il a raccroché. L'incident était clos et nos rapports ont après toujours été au

beau fixe. Je suis bien conscient que beaucoup de lecteurs qui ont connus Michel ne l'ont jamais considéré comme un monument de gentillesse. Il a souvent fait preuve d'une grande agressivité tant dans ces relations professionnelles que personnelles. Toutefois quand il décidait d'exercer son charme celui-ci pouvait être infini. Ces élèves ont tous parlé de sa grande sollicitude et de sa gentillesse à leur égard. Personnellement, en dehors de l'incident mentionné ci-dessus, mes relations avec lui ont toujours été agréables et enrichissantes, tant sur le plan mathématique que sur le plan personnel.

Le deux volumes de [24] et [30] représentent la partie publiées des travaux de Herman sur le théorème de la courbe invariante pour les « twist-maps » de l'anneau. La découverte par John Mather [71] et indépendamment par Serge Aubry [72] des ensembles dits d'Aubry-Mather avait, à cette époque du début des années 80, bien relancé les travaux sur ces questions difficiles. Pour expliquer un peu ce dont il s'agit, considérons sur l'anneau $\mathbb{T}\times\mathbb{R}$ les coordonnées naturelles (θ, r) et la 1-forme $rd\theta$ dont la dérivée extérieur $-d\theta\wedge dr$ est une forme volume qui donne la mesure de Lebesgue sur $\mathbb{T}\times\mathbb{R}$. Un difféomorphisme $F : \mathbb{T}\times\mathbb{R}\to\mathbb{T}\times\mathbb{R}$ est dit globalement canonique si la forme $F^*(rd\theta) - rd\theta$ est exacte, un tel F préserve les aires. C'est une sous-classe de l'ensemble des difféomorphismes préservant les aires, par exemple $G_{r_0}(\theta, r) = (\theta, r + r_0)$ préserve les aires mais n'est pas globalement canonique. L'exemple fondamental de difféomorphisme globalement est $T(\theta, r) = (\theta + r, r)$, c'est ce qu'on appelle le cas intégrable (plus précisément un difféomorphisme conjugué à T est dit intégrable). C'est effectivement un cas simple et intéressant puisque T préserve chaque cercle horizontal $r = \alpha$ et que sur ce cercle T n'est rien d'autre que la rotation R_α. Le théorème de la courbe invariante de Moser, qui fait partie de la famille des théorèmes KAM, dit que si $\alpha\in\mathbb{T}$ satisfait à une condition diophantienne, pour tout difféomorphisme globalement canonique F, qui est une petite perturbation de T en topologie C^k pour un certain $k = k(\alpha)$, il existe une courbe invariante par F, non homotopiquement triviale et sur laquelle F est conjugué à R_α. Rüssmann a beaucoup amélioré la valeur de $k(\alpha)$, voir par exemple [73]. Le but de [24] et [30] est de montrer que dans le cas où α est de type constant (c'est-à-dire que $\inf_{p/q\in\mathbb{Q}} q^2|\alpha - p/q| > 0$), alors le plus petit $k(\alpha)$ possible est 3 (Rüssmann avait aussi obtenu la valeur $3+\varepsilon$, pour tout $\varepsilon > 0$). Dans le premier des deux volumes, Michel reprend les travaux de Birkhoff qui montrent qu'une courbe invariante par un tel F et non homotopiquement triviale est nécessairement de la forme $\{(\theta,\psi(\theta)) \mid \theta\in\mathbb{T}\}$, où $\psi : \mathbb{T}\to\mathbb{R}$ est une application lipschitzienne. Dans les chapitres 2 et 3, il montre que si on fixe un nombre de rotation $\alpha\in\mathbb{T}$, il existe en suite de difféomorphismes F_n, $n\in\mathbb{N}$ de classe C^∞ tels que $F_n\to T$ en topologie $\mathrm{C}^{3-\varepsilon}$ et qui n'ont pas de courbe invariante non homotopiquement triviale sur laquelle F_n est conjugué à R_α. Je ne peux pas esquisser ici la démonstration. Je voudrais juste donner une idée du début pour montrer comment exclure des courbes invariantes. Soit F de la forme $F(\theta, r) = (\theta + r, r + \varphi(\theta + r))$, où $\varphi : \mathbb{T}\to\mathbb{R}$. Si une courbe non homotope à 0 est invariante par F, elle est de la forme $\{(\theta,\psi(\theta)) \mid \theta\in\mathbb{T}\}$. Si on note par $g : \mathbb{T}\to\mathbb{T}$ la restriction de F à ce cercle identifié à $\mathbb{T}$ par la projection, Herman montre que :

$$x + \frac{1}{2}\varphi(x) = \frac{\tilde{g}(x) + \tilde{g}^{-1}(x)}{2},$$

où $\tilde{g}$ est un relevé ad hoc de g en un homéomorphisme de $\mathbb{R}$. Je me souviens du sourire radieux de Michel le jour, où dans son séminaire, il nous écrivit cette équation. C'était clairement pour lui une observation géométrique clé. Cette équation permet alors d'exclure les courbes invariantes, en s'arrangeant pour que φ soit choisi tel que $\theta \mapsto \theta + 1/2\varphi(\theta)$ évite le sous-ensemble formé par les $(\tilde{g} + \tilde{g}^{-1})/2$, où g est le relevé d'un homéomorphisme de $\mathbb{T}$ de nombre de rotation α.

Dans les chapitres 2 et 3, Michel donne deux constructions pour la suite F_n. Dans le chapitre 4, il démontre le théorème de Moser pour α de type constant, en topologie $C^{3+\varepsilon}$, pour tout $\varepsilon > 0$. Si j'avais à peu près suivi le contenu de [24], surtout à l'occasion des nombreux séminaires donnés par Michel sur cette partie, il est vrai que je n'ai pas lu du tout le second volume [30]. Dans le premier chapitre de ce second volume (chapitre 5), Michel démontre le cas C^3, il lui faut près de 90 pages, l'une des difficultés est qu'il faut renoncer aux espaces Hölder et introduire des espaces de Sobolev. Michel considérait ce chapitre comme étant de loin la meilleure partie de ces deux volumes. Le chapitre 6 introduit les espaces de Besov et montre que le théorème d'existence de courbe invariante s'applique sous des hypothèses sur les dérivées au sens des distributions parfois plus faibles. Le chapitre 7 estime diverses meilleures constantes intervenant dans les théorèmes précédents. Le dernier chapitre démontre l'existence de courbe invariante pour une classe d'applications qui ne sont pas nécessairement C^2, couvrant ainsi des cas qui avaient été étudiés (numériquement) par Froeschlé en 1968.

Dans [25], Michel Herman montre que pour un difféomorphisme de classe C^2 de nombre de rotation α de type constant et assez proche de R_α en topologie C^2, la conjugaison topologique donnée par le théorème de Denjoy est absolument continue. Jane Hawkins et Klaus Schmidt [74] avaient montré auparavant qu'un résultat de ce type est faux si le nombre de rotation n'est pas de type constant. On voit apparaître dans ce papier les espaces de Sobolev ainsi que BMO. Ce papier précède [30] et le chapitre 5 de [30] utilise le résultat de [25] et en donne aussi une autre démonstration. Dans ces deux démonstrations le travail [75] d'Yves Meyer est crucial.

Les travaux [26, 27, 28, 31] et [32] sont les travaux publiés de Michel en dynamique holomorphe. Ce travail a culminé en sa collaboration avec Jean-Christophe Yoccoz pour décrire presque entièrement ce qui se passe au bord du disque de Siegel d'un point fixe, malheureusement ce travail n'a pas été entièrement rédigé sous une forme publiable. Il m'est impossible de donner une vue d'ensemble et de placer ces travaux dans le contexte des autres travaux sur la dynamique holomorphe. C'est un sujet qui a connu une grande explosion dans les vingt dernières années et dont je n'ai suivi que les premiers balbutiements au début des années 80, quand j'étais encore à Orsay, vu que toutes les informations passaient par la trinité locale Douady-Herman-Sullivan (avant que Yoccoz ne revienne de son service militaire à l'IMPA). Je renvoie au séminaire Bourbaki d'Adrien Douady sur la question [76], ainsi qu'à l'article de Herman [32] qui est un article de type « survey » à la fois des résultats anciens et récents (en 1986) sur le problème de linéarisation de Siegel et l'étude du bord du disque maximal de linéarisation.

Dans [26], Michel considère les itérations de fractions rationnelles sur la sphère $\mathbb{S}^2$. Dennis Sullivan [77] venait de montrer qu'il n'y avait pas de composante connexe errante dans le complémentaire de l'ensemble de Julia ; il résolvait ainsi magistralement un problème qui remontait aux travaux de Fatou. Ceci permet alors d'appliquer la classification obtenue par Fatou sur la dynamique d'une fraction rationnelle f sur une composante C du complémentaire de l'ensemble de Julia $J(f)$. Par le résultat de Sullivan, une telle composante C est prépériodique sous l'action de f et quitte à la remplacer par une de ses images, il existe un entier $q \geqslant 1$ tel que $f^q(C) = C$. Il résulte des travaux de Fatou que l'on peut distinguer deux cas. Dans le premier cas (la composante connexe C est alors appelé un domaine de Fatou), il existe un point critique $c \in C$ de f^q et on peut trouver un point fixe fixe de f^q dans la fermeture de C, tel que $f^{nq}(y) \to x$, pour tout $y \in C$. Dans le second cas (la composante C est appelé un domaine singulier), Fatou montre que C est conformément équivalente soit à $\mathbb{D}$, le disque unité, soit à un anneau $\{z \in \mathbb{C} \mid r < |z| < 1\}$, avec $r > 0$, et sur un tel domaine singulier f^q est holomorphiquement conjugué à une rotation. Le fait qu'il existe (beaucoup) de domaines singuliers qui sont des disques est donné par le théorème de Siegel de linéarisation au voisinage d'un point fixe où la dérivée est de la forme $\exp(2i\pi\alpha)$, avec α diophantien (ou même satisfaisant une condition plus générale obtenue par Brjuno), un tel disque est appelé disque de Siegel. Michel montre comment obtenir ici des anneaux singuliers. Il suffit pour cela de trouver des fractions rationnelles f qui laissent invariantes le bord $\mathbb{S}^1$ de $\mathbb{D}$ (par exemple des quotients de deux produits finis de type Blaschke utilisés dans la théorie classique des fonctions d'une variable complexe) et telles que $f_{|\mathbb{S}^1}$ est un difféomorphisme. On considère alors $f_\alpha = e^{2i\pi\alpha} f, \alpha \in \mathbb{T}$. Si on note par $\rho(\alpha)$ le nombre de rotation de $f_{\alpha|\mathbb{S}^1}$, l'application $\alpha \mapsto \rho(\alpha)$ est surjective et on peut donc trouver un α tel que l'on puisse appliquer la conjecture d'Arnold au difféomorphisme $\mathbb{R}$-analytique f_α qui est donc analytiquement conjugué à la rotation $z \mapsto e^{2i\pi\alpha} z$. Par le principe de prolongement analytique, cette conjugaison analytique s'étend automatiquement à un voisinage de $\mathbb{S}^1$. Les anneaux invariants du type donnés par la classification de Fatou ont été par la suite nommés anneaux de Herman.

Ce travail [26] contient aussi des exemples de fractions rationnelles ayant des orbites denses (comme son titre l'indique). Lattès avait construit de tels exemples en passant au quotient des endomorphismes complexes de $\mathbb{T}^2$. Michel Herman montre qu'il y en a beaucoup d'autres.

Dans l'appendice de [26], Michel donne une démonstration du théorème de conjugaison local d'Arnold et Moser pour les nombres de type constant sans utiliser le théorème des fonctions implicites dans les espaces de Fréchet. Il le remplace par le théorème point fixe de Schauder-Tykhonov. Dans [29], Michel Herman montre comment généraliser cette démonstration de l'appendice de [26] du théorème de conjugaison locale des difféomorphismes du cercle de nombre de rotation de type constant au cas du nombre de rotation diophantien. La dérivée schwartzienne y est explicitement utilisée, elle était implicite dans [18].

Dans [27], Michel étudie le bord d'un disque de Siegel ou d'un anneau de Herman invariant par une fraction rationnelle f. Par un argument remarquable [78] (dont Michel aurait bien aimé pouvoir revendiquer la paternité), Étienne Ghys

avait montré que si un bord d'un tel domaine était une courbe de Jordan et f restreinte à ce domaine était conjuguée à $z \mapsto e^{2i\pi\alpha}z$, avec α diophantien, alors f a un point critique sur la courbe de Jordan. Dans [27], Michel généralise ce résultat au cas où le bord n'est plus une courbe de Jordan, il doit toutefois ajouter que l'hypothèse que f est injective sur le bord. Je me rappelle que lors du premier exposé de ce travail, Michel pensait que l'injectivité sur le bord résultait aisément de celle de f sur l'intérieur (qui est, elle, une conséquence du fait que sur le domaine singulier f est conjuguée à la multiplication par un nombre complexe). Dennis Sullivan a directement soulevé le lièvre pendant l'exposé. C'est la seule et unique fois où j'ai vu Michel coincé par une erreur stupide lors d'un exposé.

Dans [28], Michel Herman montre qu'un difféomorphisme f du cercle $\mathbb{T}$ du type $x + \varphi(x) \mod 1$, où φ est un polynôme trigonométrique réel de degré au plus n a au plus $2n$ cycles périodiques, ce qui répond à une question d'Arnold. La démonstration passe par le domaine complexe et la théorie de l'itération de Fatou et Julia.

Dans [31], revenant sur la méthode d'Anosov-Katok [59], Michel construit un difféomorphisme C^∞ de la sphère $\mathbb{S}^2$ ayant un 0 comme point fixe, qui est holomorphe sur un ouvert invariant U contenant 0, holomorphiquement conjugué sur U à une rotation irrationnelle sur le disque unité et tel que le bord de U soit un pseudo-cercle (c'est un compact pathologique du plan). Les nombres de type constant sont de nouveau présents à la fin de l'article.

Les articles [33, 34, 35, 36, 37] et [38] sont les travaux publiés de Michel Herman sur les tores invariants en dimension supérieure. Le séminaire Bourbaki [79] de Jean-Christophe Yoccoz est une excellente référence sur la question.

La publication [33] est la rédaction d'un exposé fait en janvier 88 au Séminaire sur les EDP de l'École Polytechnique annonçant les résultats contenus dans [34] (article dédié à René Thom pour son 65ème anniversaire). Il s'agit de généraliser les théorèmes de Birkhoff sur les courbes invariantes par les applications de type « twist ». Vers la même époque (fin 88), John Mather obtenait les généralisations de la théorie d'Aubry-Mather [80, 81].

Je me souviens parfaitement bien de la première fois que Michel m'a parlé des résultats de [34]. C'était en janvier 1989, j'étais à l'Université de Floride à Gainesville. Michel passait la majeure partie de l'année académique 88-89 à l'Institut à Princeton et il est venu faire une visite en Floride. Il devait bien sûr faire un exposé et il m'a dit qu'il savait de quoi il allait parler parce qu'il était sûr que ça me plairait beaucoup. Il a donc exposé la première partie de [34].

Avant d'en venir à la partie mathématique, je voudrais rapporter une autre anecdote concernant ce séjour. Cette semaine là nous avions aussi la visite à Gainesville de Howie Weiss que je cherchais à faire recruter par notre université. Howie était un peu effrayé par la présence de Michel, car les premiers contacts qu'il avait eus avec Michel avaient été aussi rêches que Michel pouvaient les rendre. Toutefois lors de ce séjour commun, Michel a tout fait pour exercer son charme sur Howie lui parlant de mathématique et de bien d'autres choses, en particulier, de vin. Ce qui a le plus impressionné Howie Weiss est certainement la visite que nous avons faite tous les trois au « Cheese and Wine Gallery » (le meilleur commerce de vins et fromages du coin) où Michel a montré par ses commentaires qu'il était un grand œnologue et a dégotté en quelques minutes

une bouteille de vin à un prix ridiculement bas, qui s'est révélée excellente plus tard dans la journée lors de la dégustation.

Pour en revenir à l'exposé et au contenu de [34], introduisons sur $\mathbb{A}^n = \mathbb{T}^n \times \mathbb{R}^n$ les coordonnées canoniques (θ, r) et la forme $\lambda = \sum_{i=1}^n r_i d\theta_i$. La forme $\omega = -d\lambda$ est la forme symplectique usuelle. On suppose que F est un difféomorphisme de $\mathbb{A}^n$ symplectique exact, c'est-à-dire que la forme $F^*\lambda - \lambda$ est exacte. La condition de monotonie sur F (qui étend naturellement la condition « twist » pour les difféomorphismes de l'anneau) est que pour tout θ l'image de $r \to F(\theta, r)$ soit transverse aux fibres de la projection $\mathbb{A}^n \to \mathbb{T}^n$. Par des exemples bien choisis, Michel montre que cette condition ne suffit pas pour généraliser la théorie de Birkhoff, mais qu'il faut lui ajouter une condition de positivité et de croissance à l'infini (qui en fait correspondent aux conditions de lagrangien défini positif imposées par Mather dans [81]). Il montre de plus que les tores invariants (qui remplacent les courbes de l'anneau pour les « twists ») doivent être lagrangiens. Sous ces hypothèses, il montre que l'on peut contrôler effectivement la constante lipschitzienne d'une fonction $\psi : \mathbb{T}^n \to \mathbb{R}^n$ continue dont le graphe est lagrangien (au sens des distributions) et invariant par F. Dans la seconde partie, il montre que pour les F qui sont des petites perturbations (en topologie C^1) du cas intégrable (par exemple du temps 1 du flot géodésique d'une métrique plate sur le tore), si T est un tore lagrangien homotope à la section nulle, invariant par F et tel que la dynamique de la restriction à T soit récurrente par chaînes, alors T est un graphe. Pour sa démonstration, Michel doit supposer que la classe de Maslov est nulle, ce qui est donné par un théorème de Claude Viterbo [82]. Des généralisations non-perturbatives de cette partie de la théorie de Birkhoff ont été donnée par Mischa Bialy et Leonid Polterovitch, voir par exemple [83].

Sur un plan plus personnel, Michel avait parfaitement raison de programmer cet exposé à Gainesville en janvier 89. Depuis plusieurs années mon programme de recherches tourne autour des thèmes de cet article, des travaux de John Mather [81] ainsi que de leur relation avec les solutions de viscosité de l'équation de Hamilton-Jacobi.

Dans [35], Michel Herman considère le difféomorphisme L de $\mathbb{A}^n$ défini par $L(\theta, r) = (\theta + r, r)$. Généralisant une idée de Yoccoz dans le cas $n = 1$, il construit un voisinage ouvert $\mathcal{U}$ de L dans les difféomorphismes de $\mathbb{A}^n$ muni de la topologie C^∞ dans lequel on peut appliquer les résultats de [34]. Pour tout $F \in \mathcal{U}$, il note $Y^\infty(F)$ l'ensemble formé par les tores T de classe C^∞ invariants, lagrangiens, homotopes à $\mathbb{T}^n \times \{0\}$ et pour lesquels la restriction de F à T est C^∞ conjuguée à une rotation diophantienne. Tous ces tores sont par [34] des graphes lipschitziens (avec borne de la constante lipschitzienne de graphes intersectant un compact fixé), la fermeture $Y(F)$ en topologie C^0 de $Y^\infty(F)$ est donc formée par des graphes qui sont des tores lagrangiens et invariants. Il démontre alors que, quitte à rétrécir $\mathcal{U}$, il existe $\mathcal{G} \subset \mathcal{U}$ un G_δ dense, tel que pour tout $F \in \mathcal{G}$, il existe un ensemble $G_1 \subset Y(F)$, dense en topologie C^0, et tel que tout $T \in G_1$ soit de classe C^1, avec $F|T$ minimal et uniquement ergodique. De plus, si $n \geqslant 2$, on peut s'arranger pour que $F|T$ soit faiblement mélangeant pour son unique mesure de probabilité. En particulier, la dynamique sur un tel tore n'est même pas boréliennement conjuguée à une translation sur un tore (ou sur un autre groupe compact). C'est un résultat

assez étonnant puisqu'il dit que, dans l'adhérence des tores invariants donnés par le théorème KAM, on trouve des tores avec de la dynamique exotique.

Dans [36], Michel Herman considère un difféomorphisme L de $\mathbb{A}^n$ de la forme $L(\theta, r) = (\theta + D\ell(r), r)$, où $\ell : \mathbb{R}^n \to \mathbb{R}$ est de classe C^∞. Un tel difféomorphisme est toujours symplectique exact. Il est monotone si $D^2l(r)$ est une forme bilinéaire non dégénéré. Si cette forme est partout définie positive (ou partout définie négative), alors, les résultats de [34] s'appliquent pour toutes les perturbations suffisamment petite en topologie C^1. Par contre s'il existe r_0 tel que $D^2l(r_0)$ soit indéfinie (ce qui est parfaitement compatible avec la non-dégénérescence pour $n \geqslant 2$), il montre qu'il existe une suite F_i de difféomorphismes de $\mathbb{A}^n$ de classe C^∞, tendant vers L en topologie C^∞, et une suite de tores lagrangiens T_i de classe C^∞, homotopes à $\mathbb{T}^n \times \{0\}$ et tels que T_i soit invariant par F_i, de plus T_i n'est pas un graphe et T_i tend vers $\mathbb{T}^n \times \{r_0\}$ dans la topologie de Hausdorff. C'est une manière très forte de montrer que la théorie de Birkhoff ne peut être généralisée que dans le cadre de [34]. L'article contient aussi des conséquences de ce théorème sur des difféomorphismes symplectiques génériques qui ont un tore invariant sur lequel la dynamique est conjuguée à une rotation diophantienne.

Dans la note [37], Michel donne un exemple explicite d'une structure symplectique sur $\mathbb{T}^{2p}$ et d'un hamiltonien $H_0 : \mathbb{T}^{2p} \to [-1, 1]$ tel que pour toute petite perturbation $H : \mathbb{T}^{2p} \to \mathbb{R}$ de H_0 dans une topologie C^k (avec k ad hoc, dans [38] il donne la valeur optimale de ce k), le flot hamiltonien de H n'a pas d'orbite périodique dans $H^{-1}[-1/2, 1/2]$. Ceci montre qu'on ne peut pas obtenir « le closing lemma » dans le cadre hamiltonien de Pugh et Robinson [84] au-delà de la topologie C^2 (sur les hamiltoniens) qu'ils considèrent. Décrivons un peu précisément l'exemple, on se donne $\alpha = (\alpha_1, \ldots, \alpha_{2p-1}) \in \mathbb{R}^{2p-1}$ satisfaisant une condition diophantienne. On définit la forme symplectique w_α sur $\mathbb{R}^{2p}$ dont la matrice dans la base canonique est A avec :

$$-A^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & \alpha_1 \\ -1 & 0 & 0 & 0 & \cdots & 0 & \alpha_2 \\ 0 & 0 & 0 & 1 & \cdots & 0 & \alpha_3 \\ 0 & 0 & -1 & 0 & \cdots & 0 & \alpha_4 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & 0 & \alpha_{2p-1} \\ -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 & \cdots & -\alpha_{2p-1} & 0 \end{pmatrix}$$

On prend comme hamiltonien $H_0(\theta_1, \ldots \theta_{2p}) = \sin \theta_{2p}$. Le flot hamiltonien est donné par le champ de vecteurs $X_H(\theta_1, \ldots \theta_{2p}) = \cos(\theta_{2p}) \sum_{i=1}^{2p-1} \alpha_i \partial/\partial\theta_i$. On voit alors que le niveau $H_0^{-1}(c), c \in]-1, 1[$ est formé par les deux tores $\mathbb{T}^{2p-1} \times \{\beta_i(c)\}, i = 1, 2$, avec $\sin(\beta_i(c)) = c$, invariants par le flot et sur lesquels l'action du flot est donnée par le champ de vecteurs $\cos(\beta_i(c))\alpha$. En particulier, pour $c \in]-1, 1[$, le flot est minimal sur chacun de ces deux tores et donc n'a pas d'orbites périodiques dès que $p \geqslant 2$. Bien sûr, si H est une petite perturbation de H_0, pour tout $c \in [-1/2, 1/2]$ la surface de niveau $H^{-1}(c)$ est encore formée par deux copies $T_{c,1}$ et $T_{c,2}$ de $\mathbb{T}^{2p-1}$ qui se projettent chacune difféomorphiquement sur $\mathbb{T}^{2p-1}$ (projection donnée par les $2p-1$ premières coordonnées). Chacun de ces deux tores est invariant par le champ de vecteurs hamiltonien

X_H de H, et la situation est une perturbation de celle pour $H = H_0$. Par le théorème de Liouville sur chacun des deux tores $T_{c,j}$, on peut trouver une forme volume invariante par X_H qui est donc un multiple (par une fonction) de la forme volume usuelle de $\mathbb{T}^{2p-1}$. On reparamètre X_H sur $T_{c,j}$ pour préserver la forme volume usuelle. Un calcul explicite montre alors que le vecteur de rotation (pour la forme volume usuelle) du flot ainsi paramétré est encore α. Comme la situation est une perturbation de la situation $H = H_0$, un argument un peu délicat (mais presque standard pour Michel) permet d'utiliser le théorème local de conjugaison d'Arnold et Moser pour obtenir qu'un temps de ce flot reparamétré est conjugué, dans le groupe des difféomorphismes préservant la forme volume, à une rotation diophantienne. Par un simple argument de densité, cette conjugaison envoie le champ reparamétré sur un champ constant ; par un argument de vecteur de rotation, ce champ constant est α. Il en résulte bien que l'on ne peut pas avoir d'orbite périodique dans la surface de niveau $H^{-1}(c)$. On a en fait obtenu qu'un reparamétrage de X_H restreint à $T_{c,j}$ est conjugué au flot du champ de vecteurs constant α, qui satisfait à une condition diophantienne. Michel va un peu plus loin : un argument dû essentiellement à Kolmogorov permet de se ramener à un reparamétrage constant, et donc la restriction de X_H à $T_{c,j}$ est conjuguée à $C(T_{c,j})\alpha$, avec $C(T_{c,j}) \in \mathbb{R}^*$.

Des méthodes un peu plus raffinées montrent comment donner des contre-exemples à l'hypothèse ergodique. Michel n'a pas publié cette partie, mais on peut consulter [79] pour une description.

Michel a bien sûr souligné que les structures symplectiques dans ces exemples sont « exotiques », toutefois ces résultats sont assez extraordinaires comme l'explique John Mather [48]. Ils ont été annoncés au cours d'un Colloque à l'ENS-Lyon qui était le point culminant d'une année spéciale de Systèmes Dynamiques en France qui eut lieu pendant l'année académique 90-91. Les auditeurs se rappellent que ce colloque avait démarré en grande fanfare par une conférence de Mischa Gromov, qui d'entrée a représenté par deux patatoïdes son opinion sur l'importance relative des systèmes dynamiques et de la géométrie. Michel n'était définitivement pas d'accord avec cette représentation (sur ce genre de choses, il lui manquait une bonne dose d'humour !).

Je me rappelle aussi qu'à l'occasion de ce colloque Michel m'a fait découvrir le Saint-Joseph (un crû de de la vallée du Rhône) au cours d'un dîner chez Gamboni. Nous étions assis tous les deux seuls à une table, j'avais bien apprécié le vin et Michel encouragé par mes commentaires a commandé une seconde bouteille. J'ai rapidement dépassé mon niveau de saturation. Michel ne voulait évidemment pas laisser une bonne bouteille à moitié pleine et insistait pour que j'en reprenne. Heureusement le rire tonitruant d'Helmut Hofer, installé à une autre table avec plusieurs personnes, m'a sauvé ce jour la face (et peut-être la vie) : Michel est allé leur porter la bouteille.

Dans [39], Michel esquisse en une page des exemples de hamiltoniens sur $\mathbb{R}^{2p}$ dont une hypersurface de niveau n'a pas d'orbite périodique. Ce résultat est aussi dû à Victor Guinzburg [73, 74].

Le dernier article [40] est une liste de problèmes ouverts que Michel considérait comme important dans son domaine des systèmes dynamiques (remarquer la dédicace). Il a été publié dans les Comptes rendus du Congrès International de Berlin en 1998, c'est la rédaction de l'exposé qu'il y a fait. J'étais dans la

salle, pleine, quand il a fait cet exposé. À la sortie, il m'a demandé « comment c'était ? », comme il le faisait toujours anxieusement après un exposé. Jürgen Moser qui assistait à la conférence au premier rang est venu lui dire quelques mots d'encouragement. C'est la dernière fois que je les ai vus réunis.

Michel a demandé que ses cendres soient dispersées au-dessus de la baie de Rio. Cela a été fait lors d'un congrés à l'IMPA, du 9 au 12 avril 2001. Une bonne partie des mathématiciens proches de Michel y participaient, ainsi que de plus jeunes représentants de cette école brésilienne de systèmes dynamiques que Michel considérait comme un second chez lui. Le mardi 10 avril en fin de matinée nous sommes conduits dans la forêt de Tijuca au-dessus de l'IMPA pour la cérémonie de dispersion des cendres. Nous arrivons à un belvédère dominant la baie, un escalier descend vers un petit balconnet, où je vais rejoindre Jaco Palis regardant au loin la baie de Rio, la main posée sur l'urne contenant les cendres de Michel. Près de lui, Marguerite Flexor qui a courageusement ramené toute seule cette urne depuis Paris. Aussi présents sur ce balconnet devant moi César Camacho, François Laudenbach, Harold Rosenberg et Jean-Christophe Yoccoz, derrière moi le long des marches et sur le grand balconnet beaucoup de monde : Alain Chenciner, Raphaël Krikorian, Patrice Le Calvez, Raphaël Douady, John Mather, Jack Milnor, Eddy Zehnder, François Ledrappier, Anatole Katok, Jean-Paul Thouvenot, tous les dynamiciens de Rio et d'autres, impossible de tous les citer. Les voitures sous les directives de Marcello Viana, partent chercher d'autres personnes. Nous attendons sous un soleil de plomb, regardant la baie au loin. Je n'oublierai jamais Jaco Palis très ému, la main sur les cendres de celui qu'il aurait dû voir s'installer vers cette époque pour quelques années à l'IMPA. Les derniers participants arrivent. Jaco part, transversalement au balconnet le long d'un rebord. Il disperse les cendres de Michel vers le bas le long de ce rebord. Quand il a fini, il remet l'urne vide à Marguerite, nous remontons, pas un mot n'a été dit. L'après-midi les mathématiques reprennent leurs droits. De la dernière journée, je voudrais retenir la phrase d'Eddy Zehnder : « I liked Michel for his mathematics, and also for his sharp tongue ». Hakan Elliasson, dans le dernier exposé, nous fait part de la conjecture sur laquelle il travaille, qu'il considère comme presque démontrée puisqu'il l'a déjà mentionnée deux fois dans des exposés devant Michel sans que celui-ci ne la conteste ! Juste avant ce dernier exposé, Marcello Viana nous lit avec une voix pleine d'émotion, un texte admirable de Lennart Carleson sur Michel (il est bien dommage que Carleson ne veuille pas en laisser une trace). La conférence est finie. Le soir plusieurs d'entre nous se retrouvent à dîner autour d'une table. Nous ne parlons que de Michel : le mathématicien (sans concession, aimant s'attaquer aux problèmes difficiles), l'homme (à la fois charmeur et cultivé, pas toujours gentil) et puis le bon vivant, fin gourmet avec qui nous avions partagés tant de bons moments.

Cette nuit là, je me réveille vers trois heures du matin, le chagrin m'envahit, les larmes sortent enfin. En regardant la baie de Rio par la fenêtre, je réalise : Michel est parti.

Au revoir Michel, tu me manques infiniment.

Lyon, le 4 Juin 2001.

Références

[1] Michael-Robert Herman. Le groupe linéaire des espaces de Hilbert de dimension infinie est égal à son groupe des commutateurs. *C. R. Acad. Sci. Paris Sér. A-B,* 273 : A15–A17, 1971.

[2] Michael-Robert Herman. Simplicité du groupe des difféomorphismes de classe C^∞, isotopes à l'identité, du tore de dimension n. *C. R. Acad. Sci. Paris Sér. A-B,* 273 : A232–A234, 1971.

[3] Michael Herman and Francis Sergeraert. Sur un théorème d'Arnold et Kolmogorov. *C. R. Acad. Sci. Paris Sér. A-B,* 273 : A409–A411, 1971.

[4] M.R. Herman . Intersection complète, algébrique, affine, non singulière, en géométrie algébrique réelle. Thèse de 3e cycle, Université Paris Sud, Centre d'Orsay, Mars, 1972.

[5] Michael R. Herman. Sur le groupe des difféomorphismes du tore. *Ann. Inst. Fourier (Grenoble),* 23 (2) : 75–86, 1973. Colloque International sur l'Analyse et la Topologie Différentielle (Colloques Internationaux CNRS, Strasbourg 1972).

[6] M. R. Herman. Sur le groupe des difféomorphismes **R**-analytiques du tore. In *Differential topology and geometry (Proc. Colloq., Dijon, 1974)*, pages 36–42. Lecture Notes in Math., Vol. 484. Springer, Berlin, 1975.

[7] M. R. Herman. Sur l'algèbre de Lie des champs de vecteurs **R**-analytiques du tore. In *Differential topology and geometry (Proc. Colloq., Dijon, 1974)*, pages 43–49. Lecture Notes in Math., Vol. 484. Springer, Berlin, 1975.

[8] Michael-Robert Herman. Sur le groupe des difféomorphismes **R**-analytiques de **R**. *Nederl. Akad. Wetensch. Proc. Ser. A* **78** = *Indag. Math.,* 37 (4) : 351–355, 1975.

[9] Michel R. Herman. Sur la conjugaison des difféomorphismes du cercle à des rotations. Journées sur la Géométrie de la Dimension Infinie et ses Applications à l'Analyse et à la Topologie (Univ. Claude-Bernard–Lyon I, Lyon, 1975). *Bull. Soc. Math. France, Mém.,* 46 : 181–188, 1976.

[10] M. R. Herman . Sur les mesures invariantes. In *International Conference on Dynamical Systems in Mathematical Physics (Rennes, 1975)*, pages 103–104. Astérisque, No. 40 Soc. Math. France, Paris, 1976.

[11] Michael-Robert Herman. Conjugation C^∞ des difféomorphismes du cercle dont le nombre de rotations satisfait à une condition arithmétique. *C. R. Acad. Sci. Paris Sér. A-B,* 282 (10) : Ai, A503–A506, 1976.

[12] M. R. Herman . Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. Thèse d'État, Université Paris Sud, Centre d'Orsay, avril, 1976.

[13] Michael-Robert Herman. L^2 regularity of measurable solutions of a finite difference equation of the circle. Technical report, University of Warwick, May, 1976.

[14] Michael-Robert Herman. Conjugaison C^∞ des difféomorphismes du cercle pour presque tout nombre de rotation. *C. R. Acad. Sci. Paris. Sér. A-B,* 283 (8) : Aii, A579–A582, 1976.

[15] Michael-Robert Herman. Mesure de Lebesgue et nombre de rotation. In *Geometry and topology (Proc. III Latin Amer. School of Math., Inst. Mat. Pura Aplicada CNPq, Rio de Janeiro, 1976)*, pages , 271–293. Lecture Notes in Math., Vol 597 Springer, Berlin, 1977.

[16] Michael-Robert Herman. The Godbillon-Vey invariant of foliations by planes of T^3. In *Geometry and topology (Proc. III Latin Amer. School of Math., Inst. Mat. Pura Aplicada CNPq, Rio de Janeiro, 1976)*, pages 294–307. Lecture Notes in Math., Vol. 597 Springer, Berlin, 1977.

[17] Albert Fathi and Michael R. Herman. Existence de difféomorphismes minimaux. In *Dynamical systems, Vol. I—Warsaw*, pages 37–59. Astérisque, No. 49 Soc. Math. France, Paris, 1977.

[18] Michael-Robert Herman. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Inst. Hautes Études Sci. Publ. Math.,* 49 : 5–233, 1979.

[19] Michael-Robert Herman. Résultats récents sur la conjugaison différentiable. In *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, pages 811–820 Acad. Sci. Fennica, Helsinki, 1980.

[20] A. Fathi, M.-R. Herman and J.-C. Yoccoz. A proof of Pesin's stable manifold theorem. In *Geometric dynamics (Rio de Janeiro, 1981)*, pages 177–215 Springer, Berlin, 1983.
[21] M.-R. Herman. Construction d'un difféomorphisme minimal d'entropie topologique non nulle. *Ergodic Theory Dynamical Systems,* 1 (1) : 65–76, 1981.
[22] M. Herman and J.-C. Yoccoz. Generalizations of some theorems of small divisors to non-Archimedean fields. In *Geometric dynamics (Rio de Janeiro, 1981)*, pages 408–447 Springer, Berlin, 1983.
[23] Michael-R. Herman. Une méthode pour minorer les exposants de Lyapounov et quelques exemples montrant le caractère local d'un théorème d'Arnol′d et de Moser sur le tore de dimension 2. *Comment. Math. Helv.*,, 58 (3) : 453–502, 1983.
[24] Michael-R. Herman. Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. 1. Société Mathématique de France, Paris, i+221, 1983. With an appendix by Albert Fathi, With an English summary.
[25] M.-R. Herman. Sur les difféomorphismes du cercle de nombre de rotation de type constant. In *Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I, II (Chicago, Ill., 1981)*, pages 708–725 Wadsworth, Belmont, CA, 1983.
[26] Michael-R. Herman. Exemples de fractions rationnelles ayant une orbite dense sur la sphère de Riemann. *Bull. Soc. Math. France,* 112 (1) : 93–142, 1984.
[27] Michael-R. Herman. it Are there critical points on the boundaries of singular domains ? *Comm. Math. Phys.,* 99 (4) : 593–612, 1985.
[28] Michael-R. Herman., Majoration du nombre de cycles périodiques pour certaines familles de difféomorphismes du cercle. *An. Acad. Brasil. Ciênc.,* 57 (3) : 261–263, 1985.
[29] Michael-R. Herman. Simple proofs of local conjugacy theorems for diffeomorphisms of the circle with almost every rotation number. *Bol. Soc. Brasil. Mat.,* 16 (1) : 45–83, 1985.
[30] Michael-R. Herman. Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. 2. *Astérisque,* (144) : 248, 1986. With a correction to : *On the curves invariant under diffeomorphisms of the annulus, Vol. 1* (French) [Astérisque No. 103-104, Soc. Math. France, Paris, 1983 ; MR 85m : 58062].
[31] Michael-R. Herman. Construction of some curious diffeomorphisms of the Riemann sphere. *J. London Math. Soc. (2),* 34 (2) : 375–384, 1986.
[32] Michael-R. Herman. Recent results and some open questions on Siegel's linearization theorem of germs of complex analytic diffeomorphisms of $\mathbf{C}^n$ near a fixed point. In *VIIIth international congress on mathematical physics (Marseille, 1986)*, pages 138–184 World Sci. Publishing, Singapore, 1987.
[33] Michael-R. Herman. Existence et non existence de tores invariants par des difféomorphismes symplectiques. In *Séminaire sur les Équations aux Dérivées Partielles 1987–1988*, pages Exp. No. XIV, 24 École polytech., Palaiseau, 1988.
[34] Michael-R. Herman. Inégalités « a priori » pour des tores lagrangiens invariants par des difféomorphismes symplectiques. *Inst. Hautes Études Sci. Publ. Math.,* 70 : 47–101 (1990), 1989.
[35] M.-R. Herman. On the dynamics of Lagrangian tori invariant by symplectic diffeomorphisms. In *Progress in variational methods in Hamiltonian systems and elliptic equations (L'Aquila, 1990)*, pages 92–112 Longman Sci. Tech., Harlow, 1992.
[36] Michael-R. Herman. Dynamics connected with indefinite normal torsion. In *Twist mappings and their applications*, pages 153–182, Springer, New York, 1992.
[37] Michael-R. Herman. Exemples de flots hamiltoniens dont aucune perturbation en topologie C^∞ n'a d'orbites périodiques sur un ouvert de surfaces d'énergies. *C. R. Acad. Sci. Paris Sér. I Math.,* 312 (13) : 989–994, 1991.
[38] Michael-R. Herman. Différentiabilité optimale et contre-exemples à la fermeture en topologie C^∞ des orbites récurrentes de flots hamiltoniens. *C. R. Acad. Sci. Paris Sér. I Math.,* 313 (1) : 49–51, 1991.
[39] Michel R. Herman. Examples of compact hypersurfaces in $\mathbf{R}^{2p}$, $2p \geq 6$, with no periodic orbits. In *Hamiltonian systems with three or more degrees of freedom (S'Agaró, 1995)*, pages 126 Kluwer Acad. Publ., Dordrecht, 1999.

[40] Michael Herman. Some open problems in dynamical systems. In *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)* number Extra Vol. II, pages 797–808 (electronic) 1998.

[41] Alain Chenciner. Michel Herman, la mécanique céleste et quelques souvenirs. *Gazette des Mathématiciens,* 88 : 84–90, 2001.

[42] Henri Cartan and Harold Rosenberg. Actualité de A. Denjoy. In *Arnaud Denjoy : évocation de l'homme et de l'oeuvre*, pages 70–78. Astérique, Vol. 28–29 SMF, Paris, 1975.

[43] Harold Rosenberg. quelques souvenirs d'avant la thèse de Michael. *Gazette des Mathématiciens,* 88 : 67–68, 2001.

[44] L. C. Siebenmann. Le fibré tangent. Centre de Mathématiques de l'Ecole Polytechnique, Paris, 1969. Cours donné à la Faculté des Sciences d'Orsay, premier seemestre 1966-1967, notes de A. Chenciner, M. Herman et F. Laudenbach.

[45] D. B. A. Epstein. The simplicity of certain groups of homeomorphisms. *Compositio Math.,* 22 165–173, 1970.

[46] John N. Mather. On Haefliger's classifying space. I. *Bull. Amer. Math. Soc.,* 77 : 1111–1115, 1971.

[47] John N. Mather. A curious remark concerning the geometric transfer map. *Comment. Math. Helv.,* 59 (1) : 86–110, 1984.

[48] John N. Mather. Michael Herman. *Gazette des Mathématiciens,* 88 : 55–58, 2001.

[49] V. I. Arnol′d. Small denominators. I. Mapping the circle onto itself. *Izv. Akad. Nauk SSSR Ser. Mat.,* 25 : 21–86, 1961.

[50] Harold Rosenberg. Les difféomorphismes du cercle (d'après M. R. Herman). In *Séminaire Bourbaki, Vol. 1975/76, 28ème année, Exp. No. 476*, pages 81–98. Lecture Notes in Math., Vol. 567 Springer, Berlin, 1977.

[51] Jean-Paul Thouvenot. Michel Herman et la théorie ergodique. *Gazette des Mathématiciens,* 88 : 70–71, 2001.

[52] Dennis Sullivan. Reminiscences of Michel Herman's first great theorem. *Gazette des Mathématiciens,* 88 : 84–90, 2001.

[53] Raphaël Douady. Herman ou la passion des mathématiques et de la vie. *Gazette des Mathématiciens,* 88 : 75–77, 2001.

[54] Pierre Deligne. Les difféomorphismes du cercle (d'après M. R. Herman). In *Séminaire Bourbaki, Vol. 1975/76, 28ème année, Exp. No. 477*, pages 99–121. Lecture Notes in Math., Vol. 567 Springer, Berlin, 1977.

[55] Helmut Rüssmann. Kleine Nenner. I. Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes. *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II,* 1970 : 67–105, 1970.

[56] J.-C. Yoccoz. Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne. *Ann. Sci. École Norm. Sup. (4),* 17 (3) : 333–359, 1984.

[57] José Adem. Algebra lineal, campos vectoriales e immersiones. IMPA, Rio de Janeiro, 1978. III Escola Latino Americana de Mathmática, Rio de Janeiro, julio 1976.

[58] Guy Wallet. Nullité de l'invariant de Godbillon-Vey d'un tore. *C. R. Acad. Sci. Paris Sér. A-B,* 283 (11) : Aii, A821–A823, 1976.

[59] D. V. Anosov and A. B. Katok. New examples in smooth ergodic theory. Ergodic diffeomorphisms. *Trudy Moskov. Mat. Obšč.,* 23 : 3–36, 1970.

[60] A. B. Katok. Minimal diffeomorphisms on principal $\mathbf{S}^1$-fiber spaces. In *Teszisy VI Vsesojuznoi Topol. Konf. Tbilisi*, pages 63, 1972.

[61] I. U. Bronšteĭn. Extensions of minimal transformation groups. Martinus Nijhoff Publishers, The Hague, viii+319, 1979. Translated from the Russian.

[62] Pierre Arnoux. Quelques souvenirs de Michel Herman. *Gazette des Mathématiciens,* 88 : 74–75, 2001.

[63] Jean-Christophe Yoccoz. Souvenirs de Michel. *Gazette des Mathématiciens,*, 88 : 58–60, 2001.

[64] Ja. B. Pesin. Families of invariant manifolds that correspond to nonzero characteristic exponents. *Izv. Akad. Nauk SSSR Ser. Mat.,* 40 (6) : 1332–1379, 1440, 1976.

[65] Ja. B. Pesin. Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat. Nauk,* 32 (4 (196)) : 55–112, 287, 1977.

[66] David Ruelle. Ergodic theory of differentiable dynamical systems. *Inst. Hautes Études Sci. Publ. Math.,* (50) : 27–58, 1979.

[67] M.C. Irwin. On the stable manifold theorem. *Bull. London Math. Soc.,* 2 : 196–198, 1970.,

[68] Michael Shub. Stabilité globale des systèmes dynamiques. With an English preface and summary., Société Mathématique de France, Paris, iv+211, 1978.

[69] Jean-Benoît Bost. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnol′d, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...). *Astérisque,* (133-134) : 113–157, 1986. Seminar Bourbaki, Vol. 1984/85.

[70] R. Johnson and J. Moser. The rotation number for almost periodic potentials. *Comm. Math. Phys.,* 84 (3) : 403–438, 1982.

[71] John N. Mather. Existence of quasiperiodic orbits for twist homeomorphisms of the annulus. *Topology,* 21 (4) : 457–467, 1982.

[72] Serge Aubry. The devil's staircase transformation in incommensurate lattices. In *The Riemann problem, complete integrability and arithmetic applications (Bures-sur-Yvette/New York, 1979/1980)*, pages 221–245 Springer, Berlin, 1982.

[73] Helmut Rüssmann. On the existence of invariant curves of twist mappings of an annulus. In *Geometric dynamics (Rio de Janeiro, 1981)*, pages 677–718 Springer, Berlin, 1983.

[74] Jane Hawkins and Klaus Schmidt. On C^2-diffeomorphisms of the circle which are of type III_1. *Invent. Math.,* 66 (3) : 511–518, 1982.

[75] Yves Meyer. Sur un problème de Michael Herman. In *Conference on harmonic analysis in honor of Antoni Zygmund, Vol. I, II (Chicago, Ill., 1981)*, pages 726–731 Wadsworth, Belmont, CA, 1983.

[76] Adrien Douady . Disques de Siegel et anneaux de Herman. Séminaire Bourbaki, Vol. 1986/87. *Astérisque,* (152-153) : 4, 151–172 (1988), 1987.

[77] Dennis Sullivan. Quasiconformal homeomorphisms and dynamics. I. Solution of the Fatou-Julia problem on wandering domains. *Ann. of Math. (2),* 122 (3) : 401–418, 1985.

[78] Étienne Ghys. Transformations holomorphes au voisinage d'une courbe de Jordan. *C. R. Acad. Sci. Paris Sér. I Math.,* 298 (16) : 385–388, 1984.

[79] , Jean-Christophe Yoccoz. Travaux de Herman sur les tores invariants. Séminaire Bourbaki, Vol. 1991/92. *Astérisque,* (206) : Exp. No. 754, 4, 311–344, 1992.

[80] John N. Mather. Minimal action measures for positive-definite Lagrangian systems. In *IXth International Congress on Mathematical Physics (Swansea, 1988)*, pages 466–468 Hilger, Bristol, 1989.

[81] John N. Mather. Action minimizing invariant measures for positive definite Lagrangian systems. *Math. Z.,* 207 (2) : 169–207, 1991.

[82] Claude Viterbo. Generating functions, symplectic geometry, and applications. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 537–547 Birkhäuser, Basel, 1995.

[83] M. Bialy and L. Polterovich. Hamiltonian diffeomorphisms and Lagrangian distributions. *Geom. Funct. Anal.,* 2 (2) : 173–210, 1992.

[84] Charles C. Pugh and Clark Robinson. The C^1 closing lemma, including Hamiltonians. *Ergodic Theory Dynam. Systems,* 3 (2) : 261–313, 1983.

[85] Viktor L. Ginzburg. A smooth counterexample to the Hamiltonian Seifert conjecture in $\mathbf{R}^6$. *Internat. Math. Res. Notices,* (13) : 641–650, 1997.

[86] Viktor L. Ginzburg. Hamiltonian dynamical systems without periodic orbits. In *Northern California Symplectic Geometry Seminar*, pages 35–48 Amer. Math. Soc., Providence, RI, 1999.

Algorithmic classification of 3-manifolds and knots

S. V. Matveev [1] (*Chelyabinsk State University*)

1. Introduction

The following is known as *the recognition problem for* 3*-manifolds:*

Does there exist an algorithm to decide whether or not two given 3*-manifolds are homeomorphic?*

Why is this problem important? It is because the positive answer would imply the existence of algorithmic classification of 3-manifolds. Indeed, one can easily construct an algorithm which enumerates step by step all compact 3-manifolds. Using it, we could create a list $M_1, M_2, \dots$ of all 3-manifolds without duplicates by inquiring if each next manifold has been listed before. It is this list that is considered as a *classifying list* of 3-manifolds. Certainly, this is a classification in a very week sense; the knowledge that the classifying list exists does not help to answer questions. It is the proof of the existence that is important since the search for it would inevitably lead one to deeper understanding of the intrinsic structure of 3-manifolds.

Our goal is to overview the positive solution of the above problem for the class of so-called *sufficiently large* 3-manifolds. This case is especially important, since it implies the positive solution of the algorithmic classification problem for knots, one of the most intriguing problems of low-dimensional topology.

Recall that a *knot* is a circle embedded in S^3. Two knots are *equivalent*, if there is an isotopy $S^3 \to S^3$ taking one knot to the other. Knots are usually presented by *knot diagrams*, i.e., by generically immersed plane curves such that at each crossing point it is shown which strand goes over.

If we know at advance that two given knots are equivalent, then one can rigorously prove that by certain local modifications (called *Reidemeister moves*) that transform one diagram to the other. If they are distinct, then sometimes one can prove that by calculating different polynomial or numerical invariants. But what can we do, if both potentially infinite procedures (comparing the knots via Reidemeister moves and via knot invariants) do not stop? The right strategy consists in considering knot complement spaces, which are sufficiently large 3-manifolds.

The history of the positive solution of the recognition problem for sufficiently large 3-manifolds is very interesting. In 1962 W. Haken suggested an approach

[1] The paper is written under financial support of RFBR (grant N 99-01-00813), INTAS (project 97-808), and the program "Universities of Russia".

for solving the problem [3]. However, his proof contained a heavy gap. Thanks to efforts of several mathematicians, to the early seventies a crucial obstacle was singled out, and, when in 1978 G. Hemion overcame it [5], it was broadly announced that the problem was solved [8], [14]. Later on many topologists used extensively this result.

Let me explain my interest in the problem. Trying to understand in detail the proof, I discovered that there was no written complete proof at all. All papers and even books ([8], [14], [6]) devoted to this subject were written according to the same scheme: they contained a description of Haken's approach, of the obstacle, of Hemion's result, and the claim that these three ingredients give a proof. But no paper contained a proof of the claim! I undertook an investigation of the question and came to the following conclusions:

1. The statement that the recognition problem for sufficiently large 3-manifolds is algorithmically solvable is true.

2. There is another obstacle of similar nature that cannot be overcome by the same tools as the first one.

3. It can be overcome by using an algorithmic version of W. Thurston's theory of surface homeomorphisms that appeared only in 1995 [1].

Thus for more than 15 years mathematicians relied on an unproven theorem. In this paper we fill the gap by describing a modified proof that is based on the same ideas but is much shorter and simpler than the original one. The paper is based on my talk given at Caen university during my stay at IHES in January 1999.

2. Haken's theory of normal surfaces

The theory of normal surfaces was developed by W. Haken in the early 1960s. Its fundamental importance to the 3-manifold topology cannot be overestimated. Most of the papers on 3-manifolds in recent years are based on or related to it. W. Haken is one of the first topologists who realized that the right strategy to investigate 3-manifolds is to look over surfaces that are contained in it.

Let M be a compact 3-manifold with a fixed triangulation T. Later on we will always assume M is *irreducible* and *boundary irreducible*. It means that every 2-sphere $S \subset M$ must bound a 3-ball, and the boundary of every proper disc $D \subset M$ must bound a disc in ∂M. The restriction has a technical nature. It allows us to avoid considering connected sums, boundary connected sums, and problems related to the Poincaré conjecture.

By a *surface* in M we mean a 2-dimensional proper submanifold of M.

Definition 1. A surface $F \subset M$ is called *normal* if the following holds:

(1) F is in general position with respect to T.
(2) The intersection of F with every tetrahedron consists of discs. Those discs are called *elementary*.
(3) The boundary of every elementary disc crosses at least one edge and crosses each edge at most once.

It is easy to show that for each tetrahedron there are seven types of allowed elementary discs: four triangles and three quadrilaterals, which correspond to plane sections of the tetrahedron (see Fig. 1). Note that elementary discs as

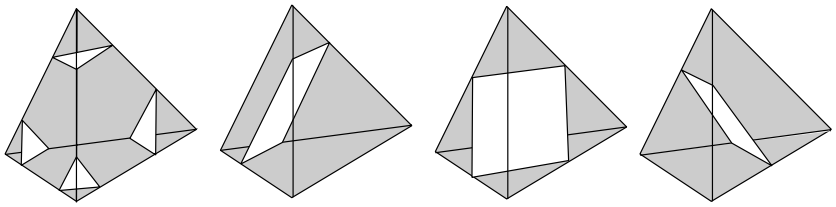


FIGURE 1. Seven plane sections of the tetrahedron

well as normal surfaces are usually considered up to *normal isotopy* of M, which preserves the triangulation.

It turns out that the set $\mathcal{N}$ of all normal surfaces in M (considered up to normal isotopy) possesses the following two non-formal properties:

(1) $\mathcal{N}$ is informative in the sense that it contains representatives of all interesting classes of surfaces in M. Certainly, the notion of an interesting class depends heavily on the problem we are trying to solve. For our purposes interesting surfaces are *incompressible* ones, see below.
(2) $\mathcal{N}$ admits a more or less explicit description.

As we have mentioned earlier, knowledge of surfaces contained in a given 3-manifold is useful for understanding its structure. Let us think a little on surfaces that are contained in R^3. It is known that any such surface can be obtained from a 2-sphere in R^3 by successive addition of tubes. The tubes may be knotted and linked (see Fig. 2), and run inside each other.

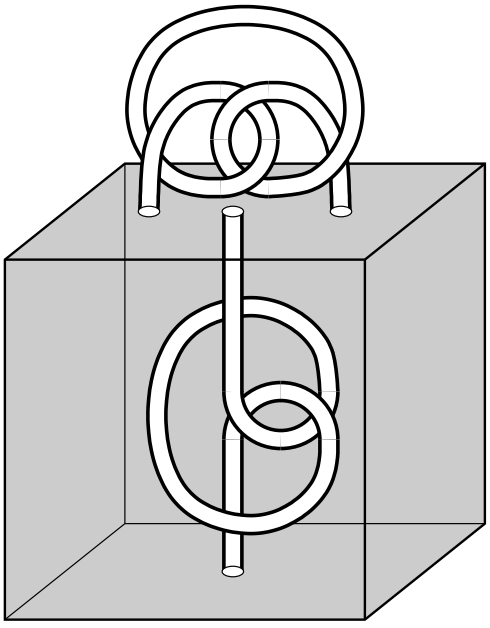


FIGURE 2. Knotted tubes

Now let us return to the case of arbitrary 3-manifold M. The same procedure (adding tubes) works here as well, but all surfaces obtained in this way seem to be not very interesting. One possible explanation is that all surfaces that are contained in R^3 are contained in every 3-manifold and hence carry no information on M. Thus surfaces without tubes are most interesting. To give a formal description of surfaces without tubes, note that if the tubes are present, then the meridional disc D of the last attached tube meets the surface only along ∂D. We come naturally to the notion of incompressible surface. By definition, a surface F in a 3-manifold M is *incompressible*, if the boundary of every disc $D \subset M$ with $D \cap F = \partial D$ bounds a disc in F.

Similarly, a *boundary incompressible* proper surface F satisfies the following property: for every disc $D \subset M$ which meets F along an arc $l \subset \partial D$ and meets ∂M along the remaining arc of ∂D, the arc l is trivial in F, i.e., cuts off a disc from F.

Proposition 1. *If an incompressible boundary incompressible surface F in an irreducible boundary irreducible triangulated 3-manifold contains no spherical or disc components, then it is isotopic to a normal surface.*

The proof is easy: we simply improve the intersection of F with every tetrahedron Δ^3 of the triangulation by isotopy of F until getting conditions 1 - 3 of Definition 1.

Corollary 1. (First surface finiteness theorem). *Let F be an incompressible closed orientable surface in a triangulated irreducible orientable 3-manifold M. Assume that F contains no spherical and no parallel components. Then $\#(F) \leqslant 10t$, where $\#(F)$ is the number of connected components of F and t is the number of tetrahedra in the triangulation.*

Proof. By Proposition 1 we may assume that F is normal. Connected components of the intersection of F with any tetrahedron Δ^3 of the triangulation are called *patches*. A patch is called *good*, if it lies between two parallel neighboring patches. It is easy to see that each tetrahedron contains no more than 10 bad patches. Since each component of F must contain at least one bad patch, $\#(F) \leqslant 10t$. □

To present an explicit description of $\mathcal{N}$, denote by $E_1, E_2, \ldots, E_n$ all the types of elementary discs in all the tetrahedra. Here $n = 7t$, where t is the number of tetrahedra in T. To each normal surface F we assign an n-tuple $\bar{x}(F) = (x_1, x_2, \ldots, x_n)$ of nonnegative integer numbers in the following natural way: we take the number of elementary discs (triangles or quadrilaterals) of each type E_i in the intersection of F with the tetrahedra. Obviously, any two surfaces having the same vector are normally isotopic. This gives us a parameterization of $\mathcal{N}$ by a subset of the set of all points in R^n with nonnegative integer coordinates. Already this parameterization can be considered as an explicit description of $\mathcal{N}$, but we want more.

Consider an angle of a common triangle face of two tetrahedra Δ^3_1 and Δ^3_2 of T. Each of the tetrahedra contains two types of elementary discs that intersect both sides of the angle: one triangle and one quadrilateral. Let $E_i, E_j \subset \Delta^3_1$ and $E_k, E_l \subset \Delta^3_2$ be the corresponding types. Then we write the equation $x_i + x_j = x_k + x_l$. Doing the same for all the angles, we get the so-called *matching system* of $3m$ linear homogeneous equations, where m is the number

of triangles in the interior of M (if M is closed, then $m = 2t$). Clearly, for any normal surface F the n-tuple $\bar{x}(F) = (x_1, x_2, \ldots, x_n)$ is a solution of the matching system. Such solutions are called *admissible*. Thus we have a parameterization of $\mathcal{N}$ by admissible solutions of the matching system. Moreover, algebraic summation of solutions has a geometrical meaning (see [4], [2]), so it makes sense to take sums of surfaces. Note that the Euler characteristic is additive with respect to the summation: if $F = F_1 + F_2$, then $\chi(F) = \chi(F_1) + \chi(F_2)$.

It seems difficult to apply these facts to solving algorithmic problems, since the set of admissible solutions is infinite. However, the set has a finite basis consisting of fundamental solutions. We say that a nonnegative integer solution to the matching system is *fundamental*, if it cannot be presented as a nontrivial sum of two other nonnegative integer solutions.

Proposition 2. *The set of fundamental solutions of any system of linear homogeneous equalities with integer coefficients is finite, and can be constructed algorithmically.*

The proof of this purely algebraic fact is based on the observation that all fundamental solutions are contained in a compact region of R^n, see, for example, [4], [2].

Let F be an incompressible torus or an incompressible boundary incompressible annulus in a 3-manifold M. Then F is called *essential*, if F is not parallel to a surface (torus or annulus) in ∂M. It is convenient to measure *the complexity* of an arbitrary proper surface $F \subset M$ by the number $c(F) = -\chi(F)$, where $\chi(F)$ is the Euler characteristic.

Corollary 2. (Second surface finiteness theorem). *Assume that an irreducible boundary irreducible 3-manifold M contains no essential tori and annuli. Then for any number c there exist only finitely many incompressible boundary incompressible surfaces in M of complexity $\leqslant c$ (up to isotopy). The surfaces can be constructed algorithmically.*

Proof. It follows from the main result of [10] that any incompressible boundary incompressible surface F in M can be presented as an integer linear combination $\sum_i k_i F_{i_1}, k_i \geqslant 0$, of fundamental surfaces F_i such that F_i are also incompressible and boundary incompressible. Moreover, since M is irreducible and boundary irreducible, among the surfaces F_i there are no 2-spheres, projective planes, and discs. By the assumption, M contains no essential tori and annuli. Therefore $c_i = c(F_i) > 0$ for all i. The conclusion of the corollary follows from the evident fact that for any number c there are only finitely many integer linear combinations $\sum_i k_i c_i, k_i \geqslant 0$, which do not exceed c. □

3. An idea of classifying sufficiently large 3-manifolds.

A 3-manifold M is called *sufficiently large*, if it contains a two-sided closed incompressible surface $F \neq S^2, RP^2$. The class of sufficiently large 3-manifolds is "sufficiently large". For example, every irreducible 3-manifold M with $\partial M \neq \varnothing$ is either sufficiently large or homemomorphic to a solid pretzel. Every irreducible closed 3-manifold M having infinite first homology group $H_1(M; Z)$ is also sufficiently large. Indeed, to construct an incompressible surface in M, consider a map $f\colon M \to S^1$ that induces a surjective homomorphism $H_1(M; Z) \to Z$. The inverse image of a regular point $\{*\} \in S^1$ contains a nonseparating surface F in M. After compressing F, i.e., after cutting all tubes, we get a nonseparating incompressible surface.

A similar construction shows that if $M \neq B^3$ is irreducible and $\partial M \neq \varnothing$ is incompressible, then M contains an incompressible boundary incompressible surface which does not separate M.

Definition 2. An orientable 3-manifold is called *Haken*, if it is irreducible, boundary irreducible, and sufficiently large.

Main Theorem. *There exists an algorithm to decide whether or not two Haken 3-manifolds are homeomorphic.*

We present here a modified version of Haken's original idea of the proof. Let M be a Haken 3-manifold. If M is closed, we choose a minimal (i.e., having the maximal Euler characteristic) incompressible surface $F \neq S^2, RP^2$ in M, which we will consider as *a partition wall*. Here is the only place where we use the assumption that M is sufficiently large. If $\partial M \neq \varnothing$, we take $F_0 = \partial M$.

Then we begin to erect new and new partition walls inside M. Each next wall F_{k+1} is a minimal incompressible boundary incompressible surface attached to the union $P_k = \cup_{0 \leqslant i \leqslant k} F_i$ of the previous walls along its boundary curves. The boundary curves of F_{k+1} should be in general position with respect to the set of singular points of P_k. This requirement guarantees us that each P_k is *a simple polyhedron* in the sense that it has the simplest possible singularities: *triple lines* (where one wall is attached to another one) and *true vertices* (crossing points of triple lines). The polyhedron P_k decomposes M into parts called *chambers*.

At last we get a 2-dimensional simple polyhedron P such that (1) all chambers of the pair (M, P) are 3-balls, and (2) the set of nonsingular points of P consists of open 2-cells. Any simple subpolyhedron of M possessing properties 1, 2 above is called a *special skeleton* of M. P can be considered as a kind of *hierarchy* for M (see [3]).

Denote by $\mathcal{P}(M)$ the set of all special skeletons of M considered up to homeomorphisms of M. It follows from the definition that the set $\mathcal{P}(M)$ is *characteristic* in the following sense:

M and N are homeomorphic $\Longleftrightarrow$ $\mathcal{P}(M)$ and $\mathcal{P}(N)$ do coincide, i.e., there is a bijection $\mathcal{P}(M) \to \mathcal{P}(N)$ taking each polyhedron to a homeomorphic one.

Haken's idea consists in an attempt to replace the infinite set $\mathcal{P}(M)$ by a finite algorithmically constructible subset $\mathcal{P}_0(M)$ having the same characteristic property. This would reduce the recognition problem for Haken 3-manifolds

to algorithmic comparing two finite sets of 2-dimensional polyhedra, which is trivial.

Below we will describe a sort of transformations of simple subpolyhedra called *extension moves*. Let P_k be a simple subpolyhedron of a Haken 3-manifold$M$. Each extension move transforms P_k to a bigger simple subpolyhedron P_{k+1} of M. The following conditions should be satisfied:

C_1. The number of different extensions of P_k is finite (up to homeomorphisms of the pair (M, P_k)), and all of them can be constructed algorithmically;

C_2. Any sequence $P_0, P_1, P_2 \ldots$, where $P_0 = \partial M$ and each P_{k+1} is an extension of P_k, is finite;

C_3. If P_k has no extensions, then P_k is a special skeleton of M.

Proof of the Main Theorem. (Under assumption that the extension moves had been described). Suppose that M is a Haken 3-manifold. Let us apply to P_0 step by step all extension moves as long as possible. It follows from conditions $C_1 - C_3$ that we get a finite set $\mathcal{P}_0(M)$ of special skeletons of M depending only on the topological type of M. On the other hand, if for two Haken manifolds M, N the sets $\mathcal{P}_0(M)$ and $\mathcal{P}_0(N)$ contain at least one pair of homeomorphic polyhedra, then M and N are homeomorphic. Indeed, we can extend a homeomorphism of the special skeletons to a homeomorphism of the manifolds by applying the cone construction to each ball chamber. It follows that $\mathcal{P}_0(M)$ is characteristic. □

It remains to construct extension moves satisfying conditions $C_1 - C_3$. The idea consists in adding new surfaces (partition walls) having minimal complexity. Corollary 2 tells us that the first difficulty in realizing the idea is the presence of essential tori and annuli (they hinder us from having only a finite number of new walls). The famous JSJ-decomposition theorem helps us to overcome it.

4. Jaco-Shalen-Johannson decomposition.

It happens often in mathematics that a proof of some statement becomes easier if we formulate the statement in more general form. Having this in mind, we will consider 3-*manifolds with boundary pattern*, where a boundary pattern is a fixed graph (one-dimensional polyhedron) in the boundary of the manifold [7]. In particular, every manifold can be considered as a manifold with the empty boundary pattern. A map between manifolds with boundary pattern is a map $M_1 \to M_2$ which takes $\Gamma_1 \to \Gamma_2$, i.e., a map of pairs $(M_1, \Gamma_1) \to (M_2, \Gamma_2)$.

Manifolds with boundary pattern appear naturally by realizing Haken's idea: each chamber $Q \subset M, Q \cap P_k = \partial Q$, is a manifold with boundary pattern consisting of triple lines and vertices of P_k which lie in ∂Q.

Let (M, Γ) be a manifold with boundary pattern. A surface $F \subset M$ is called *proper* if $F \cap \partial M = \partial F$ and ∂F is in general position with respect to Γ. The complexity $c(F)$ of F is given by the formula $c(F) = -\chi(F) + \#(F \cap \Gamma)$, where $\#(F \cap \Gamma)$ is the number of points in $F \cap \Gamma$. A surface F is *clean* if $F \cap \Gamma = \varnothing$.

The notions of irreducible boundary irreducible 3-manifold, of incompressible boundary incompressible surface, and of essential torus or annulus admit natural generalizations to the case of manifold M with boundary pattern Γ. In particular, any essential torus must be not parallel to a clean torus in ∂M, and

any clean essential annulus must be not parallel to an annulus A in ∂M such that $A \cap \Gamma$ is a collection of several parallel copies of the core circle of A. Haken's theory of normal surfaces works here as well. Corollary 2 also remains true: if an irreducible boundary irreducible 3-manifold M with boundary pattern Γ contains no essential tori and clean essential annuli, then for any number c there exist only finitely many incompressible boundary incompressible surfaces in M of complexity $\leqslant c$.

Definition 3. Let M be an irreducible boundary irreducible 3-manifold with boundary pattern Γ, and F an essential torus or clean essential annulus in M. Then F is *rough* if any other essential torus and any other clean essential annulus in M is cleanly isotopic to a surface disjoint from F.

The terminology is related to an observation that everybody would avoid contacts with a rough person.

Let M be an irreducible boundary irreducible 3-manifold M with boundary pattern Γ. A finite collection of disjoint pairwise nonparallel rough tori and annuli in M is a *JSJ-system* if it is maximal, i.e., if any other rough torus or annulus in M is isotopic to a member of the collection.

We formulate now a version of the usual JSJ-decomposition theorem for manifolds with boundary pattern (JSJ stands for W. Jaco, P. Shalen, and K. Johannson [11] ,[7]). In the case of empty boundary pattern an easy proof can be found in [13].

JSJ-decomposition Theorem. *For any irreducible boundary irreducible 3-manifold M with boundary pattern Γ the JSJ-system exists and is unique up to clean isotopy.* □

Let S be a JSJ-system in (M, Γ). Then S decomposes M into a collection $\{Q_i\}$ of compact manifolds called 3-*components* of the pair (M, S). We equip every Q_i with a boundary pattern $\Delta_i \subset \partial Q_i$ consisting of the portion $\partial Q_i \cap \Gamma$ of the pattern for M, and of traces in ∂Q_i of the boundary circles of annuli from S.

Let us investigate the structure of Q_i. Note that the JSJ-system for (Q_i, Δ_i) is empty. It means that all rough tori and annuli in Q_i are inessential. It may happen that (Q_i, Δ_i) contains no essential tori and annuli at all. Then (Q_i, Δ_i) is called *simple*. On the other hand, Q_i may contain essential tori and annuli, which necessarily are not rough. Let us present two basic examples.

Example 1. Let Q be a Seifert manifold fibered over a surface F. Suppose that Q is equipped with boundary pattern Δ consisting of a collection of fibers. Then, with a few exceptions, Q contains tori or annuli which are essential but not rough. The exceptions are fibered solid tori and manifolds fibered over S^2 with $\leqslant 3$ exceptional fibers.

Example 2. Let Q be an I-bundle, i.e., an orientable direct or skew product of a surface F and an interval. Suppose that Q is equipped with a boundary pattern Δ consisting of a nonempty collection of nontrivial circles in each annulus of the induced I-bundle over ∂F. Then, with a few exceptions, Q contains annuli that are essential but not rough. The exceptions are I-bundles over surfaces with nonnegative Euler characteristics.

Let us introduce two types of essential annuli in manifolds with boundary pattern.

Definition 4. An essential annulus A in a 3-manifold Q with boundary pattern Δ is called *longitudinal* if any clean incompressible boundary incompressible annulus $A_1 \subset Q$ is cleanly isotopic to an annulus A_1' such that $\partial A \cap \partial A_1' = \varnothing$. Otherwise A is called *transverse*.

Note that all essential rough annuli are longitudinal, while transverse annuli are never rough.

Proposition 4. *Suppose that a manifold Q with boundary pattern Δ contains no rough annuli and tori. If Q contains an essential torus or a longitudinal annulus, then Q is a Seifert manifold (as in Example 1). If Q contains a transverse annulus, then Q is an I-bundle (as in Example 2).*

Proof. To be definite, assume that Q contains an essential annulus A. Since A is not rough, there is an essential annulus A' such that $A \cap A'$ is a nonempty collection of nontrivial circles or radial segments. Then $A \cup A'$ fibers into circles in the first case and into segments in the second, and so does a regular neighborhood N of $A \cup A'$. Applying the same procedure to an essential annulus or torus contained in ∂N, we extend the fibration to a larger submanifold of Q. The "no rough annuli" assumption tells us that there are no obstructions for extending the fibration over the whole Q. □

Corollary 3. *Any irreducible boundary irreducible manifold M with boundary pattern Γ contains only finitely many incompressible tori and longitudinal annuli up to homeomorphisms of M fixed on ∂M.*

Proof. Since the behavior of essential tori and annuli in Seifert manifolds is well known (up to homeomorphisms, there are only finitely many of them), the conclusion follows from the fact that every such surface can be isotoped into a Seifert part of M. □

5. Realization of Haken's idea.

Let M be a Haken 3-manifold. We describe an algorithmic construction of a finite characteristic set $\mathcal{P}_0(M)$ of special skeletons. Starting with $P_0 = \partial M$, we will erect new partition walls. Let P_k be a simple subpolyhedron of M, obtained in an intermediate step of the construction. Consider a chamber (Q_i, Γ_i) of (M, P_k).

Extension move E_1. *Assume that (Q_i, Γ_i) contains an essential torus T or a longitudinal annulus A. Then we enlarge P_k by adding T or A, respectively.*

Although (Q_i, Γ_i) is boundary irreducible as a manifold with boundary pattern, ∂Q_i can be compressible if we forget about Γ_i.

Extension move E_2. *Suppose that (Q_i, Γ_i) contains no essential tori and clean annuli, ∂Q_i is compressible, and Q_i is not a solid torus with a clean longitude on the boundary. Then we extend P_k by adding a nontrivial compressing disc having the smallest complexity.*

Extension move E_3. *Assume that Q_i is not a 3-ball, (Q_i, Γ_i) contains no essential tori and clean annuli, and that ∂Q_i is incompressible. Then Q_i, considered as a manifold without boundary pattern, contains an incompressible boundary incompressible surface $F \neq S^2, RP^2$ that, if $\partial M \neq \varnothing$, does not*

split M. Among all such surfaces we choose a surface of the minimal complexity and insert it as a new partition wall.

Let us now perform moves $E_1 - E_3$ as long as possible. Since each time we get simpler chambers, the process terminates after a finite number of steps. By Corollary 2 for manifolds with boundary pattern and Corollary 3, at each step we have only finitely many choices of new partition walls, and the walls can be constructed algorithmically. Therefore, using all of them (and multiplying (M, P_k) in corresponding number of exemplars), we get a finite set of simple subpolyhedra of M.

Let us analyze the structure of chambers for a polyhedron P_k obtained at this stage. By construction, we can have chambers of the following three types: *balls, I-bundles, and solid tori having a clean longitude in the boundary.*

Ball type chambers are exactly what we want. Our goal is to decompose into balls the chambers of the other two types.

Denote by U the union of the I-bundle chambers. Then for any connected component U_0 of U one of the following holds:

(a) U_0 is an I-bundle $F\tilde{\times}I$, such that $F\tilde{\times}\partial I \subset \partial M$;
(b) U_0 is a Stallings manifold M_f obtained from $F \times I$, where F is a surface, by identifying $F \times \{0\}$ and $F \times \{1\}$ via a homeomorphism $f\colon F \times \{0\} \to F \times \{1\}$;
(c) U_0 is a quasi-Stallings manifold $M_{\alpha,\beta}$ obtained from $F \times I$ by identifying $F\times\{0\}$ with $F\times\{0\}$ and $F\times\{1\}$ with $F\times\{1\}$ via free involutions $\alpha\colon F\times\{0\} \to F \times \{0\}$ and $\beta\colon F \times \{1\} \to F \times \{1\}$, respectively.

Extension move E_4. *Assume that a connected component U_0 of U is an I-bundle $F\tilde{\times}I$. Then we decompose it into balls by inserting discs of the type $l\tilde{\times}I$, where l is a proper arc in F.*

This can be done in a finite number of ways. Indeed, up to homeomorphisms $U_0 \to U_0$ fixed on $\partial F\tilde{\times}I$, there are only finitely many such discs. On the other hand, since $F\tilde{\times}\partial I \subset \partial M$, any such homeomorphism $U_0 \to U_0$ can be extended to a homeomorphism $M \to M$.

Let U_0 be a Stallings manifold as in (b). The set of homeomorphisms $U_0 \to U_0$ can be very poor, so the above trick does not work. After a little thinking we come to the conclusion:

In this situation Haken's idea does not work, and one should find an independent solution of the recognition problem for Stallings manifolds.

It is the key problem that was overlooked by W. Haken and solved by G. Hemion. Note that two Stallings manifolds M_f, M_g with the same fiber F are fiber-preserving homeomorphic if and only if the monodromy maps $f, g\colon F \to F$ are conjugate (modulo isotopy). Since in our situation the manifolds contain no essential tori and annuli, one can assume that f and g are pseudo-Anosov, i.e., admit no essential periodic curves.

Theorem. (Hemion [5]). *For any two pseudo-Anosov homeomorphisms $f, g\colon F \to F$ there exists an algorithmically constructible finite set $\mathcal{H}(f, g) = \{h_1, \dots, h_m\}$ of homeomorphisms $F \to F$ such that the following holds:*

(1) f and g are conjugate $\iff g = h_i f h_i^{-1}$ for some $h_i \in \mathcal{H}(f, g)$;

(2) Every homeomorphism $h\colon F \to F$ that conjugates f to g has a form $h = h_j f^n$, where $h_j \in \mathcal{H}(f,g)$ and n is an integer. □

This theorem solves the recognition problem for Stallings manifolds by reducing it to simple checking whether some h_i conjugates f to g. Simultaneously it helps us to subdivide any Stallings chamber into balls in essentially canonic way (see [12] for details).

Next EXTENSION MOVE E_5 consists in doing that.

Now let U_0 be a quasi-Stallings manifold as in (c). Just as above, we must find an independent solution of the recognition problem for such manifolds. Hemion's theorem is insufficient for the purpose (that fact was overlooked in [8], [14], [6]). Note that two quasi-Stallings manifolds $M_{\alpha,\beta}, M_{\alpha',\beta'}$ with the same fiber F are fiber-preserving homeomorphic if and only if there exists a homeomorphism $\psi\colon F \to F$ such that $\psi\alpha\psi^{-1} = \alpha'$ and $\psi\beta\psi^{-1} = \beta'$. Multiplying the equalities, we get the equality $\psi\alpha\beta\psi^{-1} = \alpha'\beta'$, which has a clear geometric meaning: it determines a homeomorphism $M_{\alpha\beta} \to M_{\alpha'\beta'}$ of Stallings manifolds $M_{\alpha\beta}, M_{\alpha'\beta'}$ that double cover $M_{\alpha,\beta}, M_{\alpha',\beta'}$, respectively. By Hemion's theorem there are only finitely many candidates for ψ (modulo postcomposing with $(\alpha\beta)^n$), and they can be enumerated algorithmically. Thus we can assume that $\psi = \psi_0(\alpha\beta)^n$, where ψ_0 is known.

Let us transform the first equality:

$$\psi_0(\alpha\beta)^n\alpha(\alpha\beta)^{-n}\psi_0^{-1} = \alpha' \Longleftrightarrow \psi_0(\alpha\beta)^n\alpha(\beta\alpha)^n\psi_0^{-1} = \alpha' \Longleftrightarrow$$
$$\psi_0(\alpha\beta)^{2n}\alpha\psi_0^{-1} = \alpha' \Longleftrightarrow (\alpha\beta)^{2n} = \psi_0^{-1}\alpha'\psi_0\alpha \quad .$$

We have run into the following problem:

Does there exist an algorithm to decide whether one of two given pseudo-Anosov homeomorphisms $f, g\colon F \to F$ is isotopic to an integer power of another? (Here f and g stand for $\alpha\beta$ and $\psi_0^{-1}\alpha'\psi_0\alpha$, respectively; both of them are known.

The positive answer to this question would imply a solution of the recognition problem for quasi-Stallings manifolds. It turns out that the answer can be obtained by Thurston's theory of surface homeomorphisms [**?**], more precisely, by its algorithmic version [1]. Let us comment this.

Every pseudo-Anosov homeomorphism $f\colon F \to F$ has the so-called stretching factor $\lambda(f)$ possessing the following properties:

(1) $\lambda(f)$ is an algebraic number greater than 1, and $\lambda(f^n) = \lambda(f)^n$;
(2) Isotopic homeomorphisms have the same stretching factor;
(3) $\lambda(f)$ can be calculated algorithmically.

These properties allow us to find n such that $\lambda(f^n) = \lambda(g)$ (if it exists), and reduce the problem to a routine checking whether f^n and g are isotopic. Certainly, the same should be done for the pair (g, f).

Having solved the recognition problem for quasi-Stalling manifolds, we can introduce next EXTENSION MOVE E_6, which consists in subdividing into balls the quasi-Stallings chambers.

Applying moves $E_1 - E_6$ as long as possible, we get a partition of M into balls and solid tori having clean longitudes. Let U_0 be a connected component of the union of the solid tori. Then U_0 is a Seifert manifold without exceptional fibers, i.e., a direct or twisted product of a surface and a circle. It is easy to

subdivide U_0 into balls by inserting a cross-section of the fibration. To be sure that it can be done in a finite number of ways, we take a cross-section S whose boundary curves intersect singular points of P_k in a minimal number of points. The intersection of S with each solid torus is a meridional disc of the torus.

In general, the polyhedron $P_k \cup S$ is not simple, since it may contain four-fold lines. To transform it to a special skeleton we replace S by a collection of disjoint meridional discs of the tori such that the discs are close and parallel to S. The next and last EXTENSION MOVE E_7 consists in decomposing U_0 into balls by choosing a minimal cross-section and perturbing it as above.

By construction, moves $E_1 - E_7$ possess properties $C_1 - C_3$, see Section 3. This completes the proof of Main Theorem. □

6. Concluding remarks.

Remark 1. Let us show how Main Theorem gives us the algorithmic classification of knots. By the same philosophy as in the introduction, it is sufficient to construct a recognition algorithms for knots.

Let K be a knot. Denote by $C(K)$ the complement manifold obtained from S^3 by removing an open tubular neighborhood of K. We supply $C(K)$ with a boundary pattern Γ consisting of a meridional circle of K. It is easy to see that two nontrivial knots K_1, K_2 are equivalent if and only if $C(K_1), C(K_2)$, considered as manifolds with boundary pattern, are homeomorphic. It remains to note that both manifolds are Haken, and apply Main Theorem. Certainly, the same method classifies links in S^3.

Remark 2. The technic of hierarchies described above is sufficient to solve nearly all problems about Haken manifolds. For example, one can calculate the Heegaard genus [9], solve the word problem in the fundamental group [15], and so on.

References

[1] Bestvina M., Handel M. *Train tracks for surface homeomorphisms*, Topology, **34** (1995), N. 1, 109–140.

[2] Fomenko A., Matveev S. *Algorithmic and computer methods for three-manifolds*, Mathematics and Its Applications, Kluver Academic Publishers, 1997, 334 p.

[3] Haken W. *Über das Homöomorphieproblem der 3-Mannigfaltigkeiten I.* Math. Z., **80** (1962), 89–120.

[4] Haken W. *Theorie der Normalflächen. Ein Isotopiekriterium für der Kreisknoten*, Acta Math., **105** (1961), 245–375.

[5] Hemion G. *On the classification of homeomorphisms of 2- manifolds and the classification of 3-manifolds*, Acta math., **142** (1979), 123– 155.

[6] Hemion G. *The classification of knots and 3-dimensional spaces*, Oxford Univ. Press, 1992, 162 p.

[7] Johannson K. *Homotopy equivalences of 3-manifolds with boundaries*, Springer LNM **761** (1979), 303 p.

[8] Johannson K. *Topologie und Geometrie von 3- Mannigfaltigkeiten*, Jahresber. Deutsch. Math.–Ver., **86** (1984), 37–68.

[9] Johannson K. *Topology and combinatorics of 3-manifolds*, Springer LNM **1599** (1995), 446 p.

[10] Jaco W., Oertel U. *An algorithm to decide if a 3-manifold is a Haken manifold*, Topology, **23** (1984), 195–209.

[11] Jaco W., Shalen P. *Seifert fibered spaces in 3-manifolds*, Mem. Amer. Math. Soc., **220** (1979), Providence, RI.

[12] Matveev S. *Classification of sufficiently large three-dimensional manifolds*, Uspekhi Mat. Nauk **52** (1997), N. 5, 147-174 (Russian; English translation in Russian Math. Surveys, **52** (1997), N. 5, 1029-1055).

[13] Neumann W., Swarup G. *Canonical decompositions of 3-manifolds*, Geometry and Topology, **1** (1997), 21-40.

[14] Waldhausen F. *Recent results on sufficiently large 3-manifolds*, Proc. of Symp. in Pure Math. AMS, **32** (1978), 21–38.

[15] Waldhausen F. *The word problem in fundamental groups of sufficiently large irreducible 3-manifolds*, Ann. of Math., (2) **88** (1968), 272 –280.

Une approche statistique de l'analyse des génomes

Florence Muri-Majoube et Bernard Prum [1]

1. Introduction

Depuis la fin du XIX$^{\text{ème}}$ siècle les mécanismes de l'hérédité, c'est à dire de la transmission des caractères des parents à leurs enfants, sont de mieux en mieux compris. Il s'avère d'abord que ces mécanismes sont extrêmement similaires tout au long de l'échelle des espèces, depuis les bactéries jusqu'à l'Homme, en passant par toutes les plantes et tous les animaux : chaque individu porte dans chacune de ses cellules un patrimoine génétique qui détermine un grand nombre de ses caractéristiques. Ce patrimoine, qui est appelé le « génome » de cet individu, est composé d'une ou plusieurs longues chaînes, appelées les « chromosomes ». Les bactéries n'ont qu'un seul chromosome, tandis que dans l'espèce humaine on compte 46 chromosomes.

Chacun de ces chromosomes est donc une chaîne le long de laquelle se succèdent quatre molécules différentes, des bases (ou nucléotides), notées `a`, `c`, `g` et `t` (ce sont les initiales des noms de ces quatre molécules : adénine, cytosine, guanine, et thymine) — on parle de « séquence d'ADN ». Un chromosome est donc un texte « écrit » dans l'alphabet constitué de ces quatre lettres. La longueur des chromosomes est très variable d'une espèce à l'autre : le chromosome de la bactérie qui a été la plus étudiée, le colibacille ou, de son nom savant *Escherichia coli*, est composé d'un peu moins de 5 millions de bases. Les 46 chromosomes humains totalisent environ 3 milliards de bases, et, de fait, seule une infime partie du génome varie d'un homme à l'autre.

La façon selon laquelle l'information est transmise de génération en génération est en partie connue : une part de génome permet, selon un « code » aujourd'hui parfaitement déchiffré, la fabrication par la cellule des protéines qui sont les molécules qui participent à tous les mécanismes du vivant, la respiration, l'alimentation, etc. Ce qui code pour une protéine est appelé « un gène ». Mais seule une proportion limitée du génome « code pour les protéines » ; chez l'Homme elle représente environ 5% du génome.

Notons enfin que les êtres vivants se répartissent en deux catégories : les procaryotes (essentiellement les bactéries) qui sont dépourvus de noyau cellulaire et les eucaryotes (les animaux, les plantes, les champignons...) dont les cellules

[1] Laboratoire Statistique et Génome, Upresa CNRS 8071, La genopole, Evry, 523 place des Terrasses de l'Agora, 91000 Évry. E-mail : prum@genopole.cnrs.fr et florence.muri-majoube@genopole.cnrs.fr .

possèdent un noyau. Chez les eucaryotes, et contrairement aux procaryotes, la plupart des gènes sont interrompus : chacun est composé de parties codant pour des parties successives de la protéine, les exons, séparées par des parties qui ne sont pas traduites, les introns, et qui sont omises lors de la synthèse de la protéine.

Une des questions essentielles en biologie aujourd'hui concerne donc les « parties non codantes » des génomes (soit 95% du génome humain). On peut penser qu'il s'agit là d'un texte inutile, vestige peut-être d'une évolution passée. On peut aussi penser que se cachent encore dans ces parties qui ne codent pas pour les protéines un ou plusieurs autres codes aujourd'hui encore mystérieux. Le mathématicien, par l'étude statistique des structures de ces textes peut contribuer à la compréhension du fonctionnement du génome.

Avant de développer quelques exemples de contribution du statisticien à l'étude du génome, remarquons qu'il intervient dans de nombreuses autres problématiques liées à la génétique :

- Longtemps l'étude de l'héritabilité des caractères s'est faite sans que l'on ait accès au génome : on avait comme seules données la répartition des caractères au sein d'une même famille. Cette approche a suscité — et suscite encore — beaucoup de méthodes statistiques.
- En particulier, une attention spéciale a été consacrée aux caractères quantitatifs, comme par exemple le poids de grain produit par une céréale ou la quantité de lait que donne une vache. Des modèles, souvent fondés sur le modèle linéaire à effets fixes ou à effets aléatoires, ont alors été utilisés.
- Depuis quelques années, la recherche d'un message précis régissant de tels caractères quantitatifs mobilise l'énergie des généticiens comme celle des mathématiciens (l'emplacement d'un message sur le génome étant appelé un locus, on parle de QTL, pour *Quantitative Trait Locus*).
- Du calcul — à préciser — de la proximité entre individus (en particulier de la proximité entre leurs génomes), on peut inférer des conclusions sur l'évolution des espèces et construire ce que l'on appelle des arbres phylogéniques, censés retracer l'histoire des règnes végétaux et animaux depuis l'apparition de la vie sur Terre.
- Enfin l'analyse des génomes eux mêmes demande une forte intervention du mathématicien (et de l'informaticien) : reconstitution des séquences à partir de fragments déterminés expérimentalement (séquençage), constitution de « cartes génétiques » de plus en plus fines situant les gènes sur les chromosomes, comparaison des séquences provenant d'espèces différentes (par « alignement »), voici quelques domaines où l'interaction mathématique-biologie trouve matière à recherche.

Centrons notre attention sur un problème précis, celui de la recherche des mots exceptionnels. Une des justifications de cette recherche est la suivante : les virus sont essentiellement des morceaux de génome (des chaînes d'ADN) dépourvus de tout le mécanisme cellulaire l'environnant et permettant sa vie, en particulier sa reproduction. Leur fonctionnement consiste à pénétrer dans une cellule animale ou végétale, à détourner à leur profit les outils et l'énergie de cette cellule pour l'obliger à les dupliquer en grand nombre. Lorsque la cellule épuisée éclate, elle libère dans le milieu ambiant quelques centaines de virus,

prêts à parasiter de nouvelles cellules. Le virus du Sida (virus HIV) est ainsi composé d'une séquence de 9718 bases, et il parasite — entre autres — certains globules blancs de l'Homme.

Pour se protéger des attaques de virus, les cellules disposent d'un système de défense approprié (mais malheureusement parfois inefficace). Prenons l'exemple de *E. coli* : il contient un complexe, appelé RecBCD, dont le rôle est d'attaquer toute chaîne d'ADN et de la détruire. En particulier RecBCD est apte à interdire l'entrée de virus dans la cellule. Le problème provient du risque que RecBCD s'attaque au chromosome de la cellule, dont on vient de dire qu'il était constitué comme un virus. Pour éviter cette auto-destruction, *E. coli* a réparti tout au long de son propre génome un mot-clé, analogue à un mot de passe ou à un code secret. Il s'agit du mot `gctggtgg`, appelé le Chi de *coli*. Quand RecBCD rencontre ce mot, il « comprend » qu'il est en train de détruire le génome de sa propre cellule, et il cesse la dégradation. Si cette dégradation est limitée, des mécanismes de réparation peuvent alors sauver la cellule.

Il importe donc que le mot-clé soit abondant sur le génome de *coli* pour jouer son rôle protecteur. On conçoit que l'on pourra constater un nombre d'occurrences supérieur à ce qui peut être prévu par un modèle raisonnable. Bien sûr le Chi de *coli* est connu, mais pour d'autres organismes, en particulier pour d'autres bactéries, la recherche de mots beaucoup plus fréquents qu'il n'est prévisible peut être une façon de détecter « leur Chi ». La détermination des « Chi » de divers organismes peut aider à appréhender correctement ce mode d'auto-protection des cellules contre les virus et donc participer à la compréhension des phénomènes biologiques impliqués.

2. Les mots exceptionnels

2.1. Nombre d'occurrences

On ne peut comparer un nombre d'occurrences à un nombre attendu que si l'on dispose d'un modèle pour calculer ce nombre attendu.

Nous avons choisi de modéliser la succession des lettres par une chaîne de Markov [2]. On suppose donc que l'on observe une séquence $X = X_1 X_2 \dots X_n$ où les X_i appartiennent à un alphabet fini $\mathcal{A}$ et que cette séquence est une réalisation stationnaire de la chaîne de Markov sur $\mathcal{A}$ de matrice de transition π :

$$\pi(a,b) = \mathbb{P}(X_i = b \,|\, X_{i-1} = a) \qquad \text{Modèle } M1$$

Fixons un mot $W = w_1 w_2 \dots w_h$ et notons $N_W(X)$ le nombre d'occurrences de W dans X et, de même, $N_{ab}(X)$ le nombre d'occurrences du mot de deux

[2] Ce choix est guidé d'abord par ce qu'il conduit à des calculs effectifs. Il résulte aussi de la notion de « pollution » : si un mot, par exemple de 4 lettres, `catg`, est extrêmement présent dans une séquence, les mots obtenus en ajoutant une lettre derrière (`catga`, par exemple) ou devant (`tcatg`, par exemple), « auront tendance » à être plus fréquents qu'il n'est attendu. Il convient donc d'évaluer le caractère exceptionnel d'un mot — dans notre exemple de longueur 5 — *conditionnellement* à ses « sous-mots »— dans notre exemple de longueur 4. Or les nombres de mots de longueur $h-1$, constituent la statistique exhaustive du modèle markovien d'ordre $h-2$. Travailler conditionnellement à ces nombres, c'est se placer dans ce modèle. Nous restreindrons ici l'exposé au modèle markovien d'ordre 1, on montre (cf. section 2.3) que le cas général se ramène à celui-ci.

lettres ab dans X. Pour décider du caractère exceptionnel de $N_W(X)$, cherchons à calculer son espérance et sa variance dans le modèle *M1*.

Le point de départ est une formule due à Whittle (1955) donnant le cardinal de l'ensemble $\mathcal{X}$ des séquences X' telles que la première lettre soit X_1 et, pour tout couple (a,b) on ait $N_{ab}(X) = N_{ab}(X')$:

$$(1) \qquad \#\mathcal{X} = K(X)H(X) \qquad \text{avec} \qquad K(X) = \prod_{a\in\mathcal{A}} \frac{N_{a+}(X)!}{\prod_{b\in\mathcal{A}} N_{ab}(X)!}$$

où $N_{a+}(X)$ est la somme sur b des $N_{ab}(X)$, tandis que $H(X)$ est le cofacteur du terme (X_n, X_n) de la matrice de transition empirique $\hat{\pi}$ définie par $\hat{\pi}(a,b) = N_{ab}(X)/N_{a+}(X)$.

En outre, dans le modèle *M1*, toutes les séquences de $\mathcal{X}$ ont même probabilité

$$\prod_{a,b\in\mathcal{A}} \pi(a,b)^{N_{ab}(X)} \ .$$

Donc si $\beta_i(W,X)$ est le nombre de séquences de $\mathcal{X}$ telles que W apparaisse en position i,

$$(2) \qquad \mathbb{E}(N_W(X) \,|\, \mathcal{X}) = \frac{1}{\#\mathcal{X}} \sum_{i=1}^{n} \beta_i(W,X)$$

L'idée (Cowan, 1991) est d'associer à chaque séquence X où W apparaît en position i la séquence notée (de façon légèrement abusive, puisque i n'est pas rappelé) $X\backslash W$ obtenue en remplaçant l'occurrence de W en position i par le mot $uv = w_1 w_h$ composé de deux lettres, la première et la dernière de W. On établit ainsi une bijection entre les séquences de $\mathcal{X}$ telles que W apparaisse en position i et celles de $\mathcal{X}\backslash W$ telles que uv apparaisse en position i et l'on a donc $\beta_i(W,X) = \beta_i(uv, X\backslash W)$.

Il suffit alors d'écrire l'analogue de l'équation (2) pour les occurrences du mot uv dans $X\backslash W$ et d'utiliser l'égalité des sommes de β_i pour obtenir le résultat :

$$\mathbb{E}(N_W(X) \,|\, \mathcal{X}) = N_{uv}(X\backslash W) \ \frac{\#\mathcal{X}\backslash W}{\#\mathcal{X}}$$

On n'a pas écrit ici d'espérance conditionnelle pour $N_{uv}(X\backslash W)$, car cette quantité n'est plus aléatoire : $N_{ab}(X\backslash W) = N_{ab}(X) - w_{ab}$ où $w_{ab} = N_{ab}(W)$ pour tout (a,b), sauf pour le couple (u,v) pour lequel $w_{uv} = N_{uv}(W) - 1$.

Le même raisonnement répété deux fois permet le calcul de $Var(N_W(X) \,|\, \mathcal{X})$. Soit $\delta(W;d)$ la variable qui vaut 1 si une occurrence de W en position i et une occurrence de W en position $i+d$ sont compatibles [3] et 0 sinon. Si l'on note

[3] C'est à dire si les $h-d$ dernières lettres de W sont les mêmes que ses $h-d$ premières lettres, de sorte qu'il puisse y avoir « chevauchement » sur $h-d$ lettres. Par exemple $W =$ `catgcat` peut se chevaucher lui-même sur $d = 3$ positions ; le mot $W^{(3)}W$ est alors `catgcatgcat`.

$W^{(d)}W$ le mot de $2h-d$ lettres dont les h premières et les h dernières lettres coïncident avec W, le résultat est :

$$\begin{aligned}\mathbb{E}(N_W^2(X)\,|\,\mathcal{X}) = N_{uv}(X\backslash W)\ \frac{\#\mathcal{X}\backslash W}{\#\mathcal{X}} \\ + N_{uv}(X\backslash W\backslash W)(N_{uv}(X\backslash W\backslash W)-1)\ \frac{\#\mathcal{X}\backslash W\backslash W}{\#\mathcal{X}} \\ + 2\ \sum_{d=1}^{h-2}\delta(W;d)\ N_{uv}(X\backslash W^{(d)}W)\ \frac{\#\mathcal{X}\backslash W^{(d)}W)}{\#\mathcal{X}}\end{aligned}$$

On trouvera dans Prum *et al.*(1995) le détail de ce calcul, la formule donnant $\mathbb{E}(N_W(X)N_{W'}(X)\,|\,\mathcal{X})$ permettant bien sûr d'obtenir $Cov(N_W(X), N_{W'}(X)\,|\,\mathcal{X})$ ainsi que des expressions asymptotiques de toutes les quantités précédemment déterminées, dans le cas où la longueur de la séquence observée tend vers l'infini. On y trouvera aussi la démonstration du fait que, W étant fixé, $U_W = (N_W(X) - \mathbb{E}(N_W(X)\,|\,\mathcal{X}))/Var(N_W(X)\,|\,\mathcal{X})^{1/2}$ tend en loi vers la gaussienne $\mathcal{N}(0,1)$.

Les mots exceptionnels d'une séquence d'ADN donnée, sont détectés et identifiés en utilisant le logiciel *R′MES*[4]

2.2. Exemple

Le virus HIV du Sida comporte 9718 « lettres », et la matrice des $N_{ab}(HIV)$ vaut [5]

$$N(HIV) = \begin{pmatrix} 1112 & 561 & 1024 & 713 \\ 795 & 413 & 95 & 470 \\ 820 & 457 & 661 & 432 \\ 684 & 342 & 590 & 548 \end{pmatrix} \qquad \text{total/ligne} = \begin{pmatrix} 3410 \\ 1773 \\ 2370 \\ 2164 \end{pmatrix}$$

la dernière lettre étant un `a`. Le mot $W =$ `gatc` apparaît 33 fois dans cette séquence. Est-ce « normal », « excessif » ou « excessivement peu » ?

On pourrait craindre que l'apparition dans (1) de factorielles de nombres aussi grands que $N_{\mathtt{a}+}(HIV) = 3410$ rende totalement impraticable cette approche. Montrons sur cet exemple comment le quotient de tels nombres inaccessibles peut être manipulé sans difficulté.

Remplacer `gatc` par le mot constitué de sa première et de sa dernière lettre, `gc`, fait varier $N_{\mathtt{ga}}(HIV)$ de -1, $N_{\mathtt{at}}(HIV)$ de -1, $N_{\mathtt{tc}}(HIV)$ de -1 et $N_{\mathtt{gc}}(HIV)$ de $+1$, les autres demeurant invariants. Donc $N_{\mathtt{a}+}(HIV)$ varie de -1, $N_{\mathtt{c}+}(HIV)$ varie de 0, $N_{\mathtt{g}+}(HIV)$ varie de 0 et $N_{\mathtt{t}+}(HIV)$ varie de -1.

Dans $\dfrac{K(X\backslash W)}{K(X)}$ intervient

$$\frac{N_{\mathtt{a}+}(X\backslash W)!\ N_{\mathtt{c}+}(X\backslash W)!\ N_{\mathtt{g}+}(X\backslash W)!\ N_{\mathtt{t}+}(X\backslash W)!}{N_{\mathtt{a}+}(X)!\ N_{\mathtt{c}+}(X)!\ N_{\mathtt{g}+}(X)!\ N_{\mathtt{t}+}(X)!}$$

[4] **R**echerche de **M**ots **E**xceptionnels dans une **S**équence. *R′MES* est disponible sur le site http://www.inra.fr/bia/J/AB/genome/RMES2/welcome.html .

[5] Nous rangerons toujours les lettres dans l'ordre alphabétique : `a`, `c`, `g`, `t`.

qui est égal à $\dfrac{3409!\ 1773!\ 2370!\ 2163!}{3410!\ 1773!\ 2370!\ 2164!}$

et se réduit à $\dfrac{1}{3410 \times 2164}$.

De même on a à calculer (dans l'ordre ga, at, tc, gc)

$$\frac{820!\ 713!\ 342!\ 457\ !}{819!\ 712!\ 341!\ 458!} = \frac{820 \times 713 \times 342}{458} .$$

Finalement $\dfrac{K(X\backslash W)}{K(X)} = 0,05916$ et [6]

$$\mathbb{E}(N_{\mathtt{gatc}}(X) \,|\, \mathcal{X}) = 0,05916 \times N_{\mathtt{gc}}(X\backslash \mathtt{gatc}) = 0,05916 \times 458 = 27,097$$

Le nombre observé, $N_{\mathtt{gatc}}(HIV) = 33$ est au dessus du « nombre attendu ». Pour savoir si l'écart est significatif, on calcule $Var(N_{\mathtt{gatc}}(X) \,|\, \mathcal{X})$, on trouve 23,583. Donc

$$U_{\mathtt{gatc}} = \frac{N_{\mathtt{gatc}}(X) - \mathbb{E}(N_{\mathtt{gatc}}(X) \,|\, \mathcal{X})}{Var(N_{\mathtt{gatc}}(X) \,|\, \mathcal{X})^{1/2}} = 1,2156$$

ce qui n'est pas, aux niveaux usuels (5% par exemple), une valeur excessive pour une $\mathcal{N}(0,1)$.

Notons pour terminer que si l'on considère, comme précédemment, une suite de séquences dont la longueur n tend vers l'infini, mais aussi une suite de mots W_n de plus en plus longs, de sorte que $\mathbb{E}(N_{W_n}(X) \,|\, \mathcal{X})$ tende vers une constante λ, la loi de $N_{W_n}(X)$ ne tend pas, comme on pourrait s'y attendre vers la loi de Poisson de paramètre λ : du fait des chevauchements, la loi limite est plus complexe. Schbath (1995) montre qu'il s'agit d'un mélange de lois de Poisson (loi de Poisson composée).

2.3. Quelques extensions

Famille de mots, distance

Chez la bactérie *Haemophilus Influenzae* la fonction de Chi est assumée par un ensemble de 4 mots

$$\mathcal{W} = \{W_1 = \mathtt{gatggtgg}, W_2 = \mathtt{gctggtgg}, W_3 = \mathtt{ggtggtgg}, W_4 = \mathtt{gttggtgg}\}$$

Puisque l'on dispose de la covariance $Cov(N_W, N_{W'} \,|\, \mathcal{X})$ entre les nombres d'occurrences de deux mots W et W', il est possible de calculer l'espérance et la variance de $N_{\mathcal{W}} = \sum_{W \in \mathcal{W}} N_W$ et donc de tester si cette famille de mots est exceptionnellement fréquente dans une séquence. Les méthodes présentées précédemment peuvent donc s'appliquer à la recherche de Chi multiples.

Un autre problème concerne la répartition d'un mot W (ou des mots appartenant à une famille $\mathcal{W}$) le long d'une séquence. Sous un modèle markovien on peut étudier la loi exacte de la distance séparant deux occurrences du ou des mots considérés et donc tester la régularité de cette répartition. Notons qu'en

[6] Le rapport $H(X\backslash W)/H(X)$ étant proche de 1, on ne l'a pas calculé ici, ce qu'il aurait fallu faire pour avoir un résultat totalement exact.

l'absence de « chevauchement » (voir Note 3), la répartition attendue est poissonnienne, mais que ceci devient faux dès qu'il y a « chevauchement » possible (Reinert *et al.*, 2000, Robin and Daudin, 2001).

Approche combinatoire

Une approche sensiblement différente a montré toute sa puissance quand la famille $\mathcal{W}$ peut-être décrite par une *expression régulière*, c'est à dire en employant un nombre fini de fois les opérations d'union, de produit et de fermeture [7] (c'est la cas, par exemple, pour les motifs caractéristiques de nombreuses familles de protéines tels qu'ils sont décrits dans la base de données biologiques ProSite). Cette approche, qui ne demande donc pas à $\mathcal{W}$ d'être fini (en pratique, qui permet à $\mathcal{W}$ de comprendre plusieurs centaines de mots), se fonde sur l'emploi de fonctions génératrices

$$F(l_1, \ldots, l_p) = \sum f_{i_1,\ldots,i_p} l_1^{i_1} \ldots l_p^{i_p}$$

où $f_{i_1,\ldots,i_p}$ est le nombre de mots de $\mathcal{W}$ ayant i_k lettres égales à l_k, pour k variant de 1 à p. Cette approche fournit d'excellentes approximations théoriques et, en s'appuyant fortement sur les logiciels de calcul formel, permet de traiter de gros jeux de données. On la trouvera décrite dans Nicodème *et al.* (1999), Nicodème (2000) et Nicodème (2001).

Modèles Markoviens d'ordre h

Les modèles de chaîne de Markov (d'ordre 1) *M1* tels que nous les avons utilisés ci-dessus calculent donc espérance et variance du nombre d'occurrences d'un mot W conditionnellement aux décomptes des mots de 2 lettres. Il importe souvent de prendre en compte les nombres d'occurrences de mots plus longs, disons de longueur $h+1$ donc de travailler dans le modèle markovien d'ordre h, que nous noterons *Mh*. Techniquement, on ramène ce cas à *M1* en réécrivant la séquence dans l'alphabet [8] « agrandi » $\mathcal{A}^h$. Pour une séquence d'ADN, la matrice de transition dans l'alphabet agrandi aura 4^h lignes et 4^h colonnes. Si les résultats théoriques découlent facilement d'un tel changement d'alphabet, la mise en œuvre des algorithmes pose de gros problèmes concrets dès que h dépasse la dizaine Schbath (1995, 2000).

Grandes déviations

Enfin, les approches précédentes se fondent essentiellement sur une approximation « centrale » de la loi de N_W telle qu'elle peut être déduite d'un théorème de la limite centrale (cas gaussien) ou d'une approximation de type poissonnien. Or, dans la pratique, on est bien souvent conduit très loin du « centre » des distributions (une variable « devant » suivre une loi $\mathcal{N}(0,1)$ peut prendre des valeurs dépassant 30 ou 50 !). Il est alors pertinent de remplacer l'approche « centrale » par une approche fondée sur la théorie des Grandes Déviations. Celle-ci décrit à quelle vitesse (de fait exponentielle) décroît la probabilité pour qu'une moyenne de fonctionnelles présente un écart supérieur à δ par

[7] Le produit de deux familles $\mathcal{W}$ et $\mathcal{W}'$ est l'ensemble des concaténations WW' où $W \in \mathcal{W}$ et $W' \in \mathcal{W}'$; la fermeture de $\mathcal{W}$ est la réunion infinie des puissances $\mathcal{W}^k$ pour $k \geqslant 0$.

[8] Par exemple pour $h = 2$, la séquence `gctggtgg` s'écrira `(gc)(ct)(tg)(gg)(gt)(tg)(gg)` et le modèle *M2* « dans » l'alphabet $\mathcal{A}$ sera un modèle *M1* dans $\mathcal{A}^2$.

rapport à l'espérance, quand δ est fixé. Pour de tels événements les approximations « centrales » sont d'autant moins précises que l'asymptotique n du problème est grande, ce qui montre indéniablement les limites de ces approches dans ce cas. Pour leur part, les grandes déviations proposent une approximation des probabilités de ces événements qui ne cessent de s'améliorer lorsque n augmente. En choisissant judicieusement les valeurs des fonctionnelles dont on examine la moyenne, on peut se ramener à l'étude du comptage N_W d'un mot, ou encore à celui d'un famille finie de mots dont les différents comptages peuvent éventuellement recevoir des pondérations. Les calculs font intervenir la plus grande valeur propre d'une matrice creuse d'ordre 4^h, ce qui induit une complexité exponentielle. On pourra consulter la thèse de Nuel (2001) pour obtenir de plus amples détails sur ces méthodes et les moyens de les mettre en œuvre.

3. Les Chaînes de Markov cachées

Une critique pertinente du modèle markovien utilisé dans ce qui précède est qu'il est *homogène* tout au long de la séquence : la matrice de transition π est supposée la même en toute position du génome étudié. Quand on sait que, aujourd'hui, on considère des séquences de plusieurs centaines de milliers de bases, on conçoit que cette hypothèse soit inadmissible pour un biologiste.

La modélisation par chaîne de Markov d'ordre 1 (ou d'ordre k plus élevé) permet uniquement d'intégrer des dépendances locales qui peuvent exister entre les bases, mais ne reflète pas l'hétérogénéité pouvant exister entre différentes régions d'un même génome. Il existe en effet des différences de composition, de structure, de fonction à l'intérieur d'une même séquence. Il est par exemple connu des biologistes que des régions « riches en $\mathtt{g}+\mathtt{c}$ [9] » alternent avec des régions « pauvres en $\mathtt{g+c}$ ». Les séquences d'ADN qui codent pour les protéines contiennent des parties appelées codantes que nous assimilerons à des gènes, qui alternent avec des régions intergéniques. Chez les eucaryotes, un gène est lui-même composé d'exons séparés par des introns. La structure secondaire des protéines comporte trois catégories d'éléments structuraux : des boucles, des hélices α et des feuillets β.

La prise en compte de ces divers niveaux d'hétérogénéité dans la modélisation revient à exhiber un modèle capable de segmenter une séquence biologique en un nombre fini de régions homogènes pouvant avoir des statuts biologiques différents (par exemple, on cherchera à segmenter une séquence en deux types de régions pour identifier les exons et les introns). Une approche possible consiste à utiliser les modèles de chaînes de Markov cachées, *HMM* (Hidden Markov Model). Il existe bien sûr d'autres méthodes statistiques de segmentation comme la détection de ruptures, des méthodes « à fenêtres glissantes », Audic and Claverie (1998) (on pourra consulter par exemple Braun and Müller, 1998, pour une revue), mais les *HMM* sont aujourd'hui très largement utilisés en biologie moléculaire, en raison de leur caractère adaptatif, leur grande flexibilité, et l'interprétation simple des paramètres du modèle (contrairement par exemple aux réseaux de neurones). Nous retrouvons en effet ces modèles dans des problématiques aussi diverses que les alignements simples et multiples de séquences, la

[9] Ce qui signifie $\mathtt{g}$ ou $\mathtt{c}$.

détection et prédiction de gènes, la recherche de motifs, la prédiction de structure secondaire, l'analyse par profil... (Durbin *et al.*, 1998, Baldi and Brunak, 1998). Nous nous limiterons dans un premier temps au cadre un peu vague de description de l'hétérogénéité d'une séquence biologique par *HMM*. Ces modèles ont été utilisés pour la première fois pour la recherche de régions homogènes par Churchill (1989, 1992), puis par Muri (1997, 1998) et Boys *et al.* (2000). Après une présentation des *HMM* dans ce cadre, nous décrirons quelques algorithmes permettant de répondre à trois grands problèmes classiques que se posent les biologistes utilisant les *HMM* : estimer les paramètres du modèle à partir d'une séquence préalablement segmentée, ou d'une séquence dont la segmentation est inconnue. Il s'agit de problèmes d'apprentissage du modèle qui se résolvent en utilisant, par exemple, l'algorithme *EM* ou les méthodes de Monte Carlo par Chaînes de Markov. La troisième situation, qui se rencontre très fréquemment en biologie moléculaire, consiste à segmenter une séquence lorsque le modèle est appris à partir d'un ensemble d'apprentissage, ce qui peut se faire en utilisant l'algorithme de Viterbi. Nous terminerons en présentant quelques applications des *HMM* correspondant à des hétérogénéités particulières, comme la recherche de transferts horizontaux de gènes chez les bactéries, et la détection de gènes chez les eucaryotes et chez les procaryotes.

3.1. Les modèles de chaînes de Markov cachées

Dans cette approche, la séquence biologique est modélisée, non plus par un seul modèle homogène, mais par un ensemble fini q de modèles qui vont alterner le long de la séquence : nous supposons donc que la séquence peut être découpée en régions homogènes (pouvant se répartir sur plusieurs plages) modélisées par différents modèles ou régimes, qui correspondront à des compositions locales en nucléotides différentes. L'alternance des plages le long de la séquence est gouvernée par une chaîne de Markov, qui est cachée puisqu'on ignore a priori, et la taille, et la position des plages. Ceci revient à dire que les q modèles que l'on peut ajuster correspondent aux q états d'une chaîne de Markov, qui ne sont pas observés.

Les chaînes de Markov cachées sont donc caractérisées par deux processus : le processus caché $\{S_i, i = 1, 2, \ldots\}$ et le processus observé $\{X_i, i = 1, 2, \ldots\}$, le premier étant une structure sous-jacente au second. Ainsi à chaque position i de la séquence X sera associée non seulement la base observée $X_i \in \mathcal{A} = \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}$, mais aussi un régime caché $S_i \in \mathcal{S} = \{1, 2, \ldots, q\}$ indiquant avec quelle loi la base X_i est générée.
Les paramètres du modèle sont d'une part, les transitions markoviennes homogènes d'ordre 1 (*M1*) des régimes cachés notées

$$\pi_e(u, v) = \mathbb{P}(S_i = v \,|\, S_{i-1} = u),\ u, v \in \mathcal{S}$$

gérant les changements de régimes, et la loi initiale (que l'on peut choisir comme la loi invariante),

$$\mu_e(v) = \mathbb{P}(S_1 = v),\ v \in \mathcal{S}$$

et d'autre part, les paramètres de la loi d'apparition des bases X_i conditionnellement aux régimes cachés. On peut supposer que dans le régime u les bases sont tirées indépendamment (*M0*) les unes des autres selon la loi

$$\pi_o(u,a) = \mathbb{P}(X_i = a \,|\, S_i = u),\ u \in \mathcal{S}, a \in \mathcal{A}$$

On parlera alors du modèle *M1-M0* qui prend en compte la composition locale en nucléotides. Plus généralement, le modèle *M1-Mk* suppose une dépendance markovienne d'ordre k des bases conditionnellement aux régimes cachés, de transitions notées

$$\pi_o(u,a_1,\ldots,a_k,a_{k+1}) = \mathbb{P}(X_i = a_{k+1} \,|\, X_{i-1} = a_k,\ldots,X_{i-k} = a_1, S_i = u),$$
$$u \in \mathcal{S}, a_1,\ldots,a_{k+1} \in \mathcal{A}$$

et de loi initiale

$$\mu_o(u,a_1,\cdots,a_k) = \mathbb{P}(X_1 = a_1,\cdots,X_k = a_k \,|\, S_k = u)$$

Ce modèle tient compte de la composition locale en mots de longueur $k+1$.

Nous nous placerons désormais sous le régime stationnaire de telle sorte que les seuls paramètres à estimer sont les matrices de transitions Π_e et Π_o, que nous regrouperons sous la notation θ.

Classiquement, on représente l'architecture d'un modèle de chaînes de Markov cachées par un graphe dont les ronds désignent les états, et les flèches les transitions autorisées entre états. La figure 1 représente un modèle *M1-M0* à 2 états cachés dont les paramètres sont connus : la longueur moyenne des plages du régime 1 est 100 *bp* et de 200 *bp* pour le régime 2 ; le régime 1 est « pauvre en $\mathtt{g}+\mathtt{c}$ » ($\pi_o(1, \mathtt{g}$ ou $\mathtt{c}) = 0.3$) alors que le régime 2 est « riche en $\mathtt{g}+\mathtt{c}$ » ($\pi_o(2, \mathtt{g}$ ou $\mathtt{c}) = 0.8$).

Une structure de modèle étant choisie, c'est à dire le nombre d'états cachés q, et l'ordre k du modèle *M1-Mk* étant fixés, les deux problèmes classiques qui se posent sont les suivants :

(1) estimer les paramètres du modèle θ : on parle d'apprentissage du modèle,
(2) restaurer la suite des régimes cachés $S = (S_1 \ldots S_n)$ à partir de la séquence observée $X = (X_1 \ldots X_n)$: il s'agit d'un problème de segmentation.

La résolution des problèmes 1 et 2 permet de découper la séquence d'ADN en régions homogènes et de caractériser, par une composition en oligonucléotides particulière, les régions identifiées. Cette caractérisation statistique des régions détectées ainsi que leurs localisations précises sur la séquence apporteront alors aux biologistes des renseignements susceptibles de révéler des différences structurelles et fonctionnelles à l'intérieur du génome étudié. De plus, cette segmentation permettra d'appliquer sur les régions homogènes identifiées des méthodes statistiques élaborées avec des modèles fondés sur une hypothèse d'homogénéité tout au long du génome, comme la recherche de motifs apparaissant avec une fréquence exceptionnelle.

Nous allons dans ce qui suit présenter quelques méthodes permettant de répondre au double problème posé, en nous plaçant exclusivement dans le modèle *M1-M1* : les méthodes classiques d'estimation, comme l'estimation par maximum de vraisemblance ou l'estimation bayésienne, sont difficiles à mettre en oeuvre pour les *HMM* en raison du caractère incomplet du modèle.

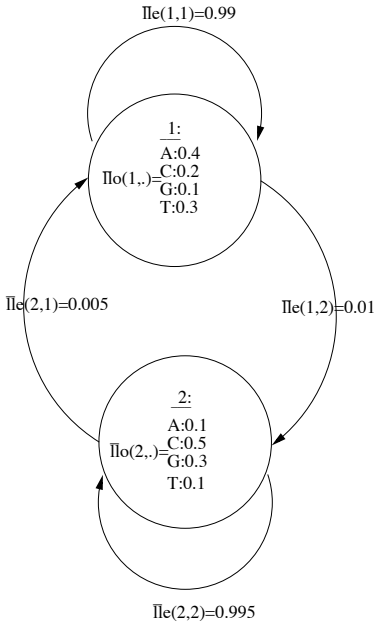


FIG. 1. Représentation d'un modèle *M1-M0* à 2 états cachés.

Les données *complètes* se décomposent en données observées *incomplètes* $X = (X_1, \cdots, X_n) \in \mathcal{A}^n$ et en données manquantes $S = (S_1, \cdots, S_n) \in \mathcal{S}^n$. Nous noterons $g(X \,|\, \theta)$, la vraisemblance des données incomplètes X et $f(X, S \,|\, \theta)$ la vraisemblance des données complètes (X, S) reliée à g par

$$\begin{aligned} g(X \,|\, \theta) &= \sum_{s \in \mathcal{S}^n} f(X, s \,|\, \theta) \\ &= \sum_{s_1, \cdots, s_n \in \mathcal{S}^n} \mu_e(s_1)\mu_o(s_1, X_1) \prod_{i=2}^{n} \pi_e(s_{i-1}, s_i)\pi_o(s_i, X_{i-1}, X_i) \end{aligned}$$

$g(X \,|\, \theta)$ n'est donc pas manipulable pour de grandes valeurs de n.
La vraisemblance $f(X, S \,|\, \theta)$ est en revanche tout à fait calculable dès lors que les régimes sont connus

$$f(X, S \,|\, \theta) = \mu_e(S_1)\mu_o(S_1, X_1)\Big(\prod_{u,v=1}^{q} \pi_e(u, v)^{N(uv)} \Big)\Big(\prod_{u=1}^{q} \prod_{a,b \in \mathcal{A}} \pi_o(u, a, b)^{N(u,ab)} \Big)$$

et peut s'exprimer (comme un modèle de chaînes de Markov) en fonction des comptages $N(uv) = \sum_{i=2}^{n} 1_{\{S_{i-1}=u, S_i=v\}}$ des régimes dans l'état u suivi d'un régime dans l'état v et $N(u, ab) = \sum_{i=2}^{n} 1_{\{S_i=u, X_{i-1}=a, X_i=b\}}$ du dinucléotide ab dans l'état u.

Une autre caractéristique des modèles de chaînes de Markov cachées, rendant l'estimation des paramètres difficile, est qu'ils peuvent être interprétés comme une généralisation des modèles de mélanges au cas où les observations sont dépendantes. En effet, un mélange classique de q distributions a une densité de la forme

$$g(X \mid \theta) = \prod_{i=1}^{n} g(X_i \mid \theta) = \prod_{i=1}^{n} \Big(\sum_{u=1}^{q} \mu_e(u) p(X_i \mid \varphi_u) \Big)$$

où le vecteur μ_e représente les proportions du mélange, p est une fonction de densité paramétrée par $\varphi_u \in \Phi_u \subseteq \mathbb{R}^{n_u}$ et $\theta = (\mu_e(1), \ldots, \mu_e(q), \varphi_1, \ldots, \varphi_q)$. Les variables S_i indiquent alors la provenance non observée de l'observation X_i et sont générées de manière indépendante avec probabilité μ_e. La densité du modèle incomplet s'écrit encore

$$g(X \mid \theta) = \sum_{s_1,\ldots,s_n=1}^{q} \Big(\prod_{i=1}^{n} \mu_e(s_i) p(X_i \mid \varphi_{s_i}) \Big) .$$

Le modèle de mélange est donc un « modèle multinomial caché ».

3.2. Apprentissage d'un modèle de chaînes de Markov cachées

3.2.1. Lorsque la segmentation est connue

Le premier problème que nous nous posons est d'estimer les paramètres du modèle θ à partir d'une séquence, ou d'un ensemble de séquences, dont la segmentation est connue. Ce cas est de loin le plus simple, puisqu'en chaque position i de la séquence la valeur de chaque régime S_i est connue. Pour tout $u, v \in \mathcal{S}$, la transition $\pi_e(u, v)$ sera alors naturellement estimée par [10]

$$\hat{\pi}_e(u, v) = \frac{N(uv)}{\sum_{v \in \mathcal{S}} N(uv)}$$

et $\forall u \in \mathcal{S}$, $\forall a, b \in \mathcal{A}$, la transition $\pi_o(u, a, b)$ par

$$\hat{\pi}_o(u, a, b) = \frac{N(u, ab)}{\sum_{b \in \mathcal{A}} N(u, ab)}$$

La vraie difficulté provient, non pas de l'estimation proprement dite des paramètres, mais de l'élaboration de l'ensemble d'apprentissage servant à déterminer la (ou les) structure cachée à partir de laquelle le modèle est appris. L'estimation obtenue est fortement liée à l'ensemble d'apprentissage utilisé qui doit donc être impérativement composé de séquences représentatives des séquences biologiques à modéliser. Bien plus grave, l'apprentissage du modèle n'est que très rarement remis en cause dans les applications, et les nouvelles séquences à segmenter (voir section 3.2.3) à partir de ce modèle « universel » devront obligatoirement être proches de l'ensemble d'apprentissage pour que la segmentation obtenue ait un sens. Nous reviendrons sur ce point fondamental dans les applications.

[10] Ces formules se généralisent facilement au cas où l'on dispose d'un ensemble d'apprentissage, en traitant les séquences de manière indépendante.

3.2.2. Lorsque la segmentation est inconnue

Lorsque la segmentation de la séquence est inconnue, l'estimation des paramètres est bien-entendu plus difficile à mettre en oeuvre. Nous présentons deux méthodes d'estimation : l'une par maximum de vraisemblance [11], l'autre bayésienne. Ces deux approches reposent sur la connaissance de la vraisemblance du modèle incomplet $g(X \mid \theta)$, qui, nous l'avons vu, est difficile à manipuler puisque les régimes ne sont pas observés. La solution consiste à augmenter les données en « attribuant » une valeur aux régimes cachés S_i et à travailler avec la vraisemblance des données complètes $f(X, S \mid \theta)$.

Plusieurs méthodes itératives ont été développées pour estimer θ de façon satisfaisante. On en trouvera une étude détaillée, entre autres, dans Rabiner (1989), Qian and Titterington (1990, 1991), Robert *et al.* (1993), Muri (1997,1998). Partant d'une valeur arbitraire du paramètre $\theta^{(0)}$, et sachant l'estimation courante $\theta^{(m)}$, ces algorithmes alternent deux étapes :

1. attribuer une valeur aux régimes cachés $S^{(m+1)} = (S_1^{(m+1)}, \ldots, S_n^{(m+1)})$ à partir de $\theta^{(m)}$ et de la séquence d'ADN observée X ;
2. sachant $S^{(m+1)}$, actualiser $\theta^{(m)}$ en $\theta^{(m+1)}$ sur la base de la vraisemblance des données complètes $f(X, S^{(m+1)} \mid \theta^{(m)})$.

Ces algorithmes répondent ainsi au double problème posé : ils permettent simultanément de reconstruire la suite des régimes cachés et d'estimer les paramètres du modèle.
Nous avons choisi de présenter deux algorithmes : l'algorithme *EM* [12] proposé par Dempster *et al.* (1977) pour approcher l'estimateur du maximum de vraisemblance dans un cadre général de modèles à données incomplètes, très largement utilisé en biologie moléculaire, ainsi qu'une estimation bayésienne utilisant l'échantillonnage de Gibbs (initialement proposée par Robert *et al.*, 1993 pour les *HMM*).
Notons qu'il existe d'autres méthodes d'estimation des paramètres d'un *HMM*, comme une estimation récursive fondée sur une optimisation directe de la log-vraisemblance proposée par Mevel (1997), des méthodes de recuit-simulé proposées par Vandekerkhove (1998)...

Algorithme EM

Pour les modèles de chaînes de Markov cachées, l'algorithme *EM* est connu sous le nom d'algorithme forward-backward de Baum-Welch (Baum *et al.*, 1970), et a été appliqué pour la première fois à la segmentation de l'ADN par Churchill (1989, 1992).

L'idée sous-jacente est simple : si l'on connaissait les S_i, on estimerait sans peine les matrices de transition Π_e et Π_o en maximisant en θ la vraisemblance des données complètes $f(X, S \mid \theta)$ (voir section précédente).

[11] Les résultats de consistance et de normalité, qui justifient l'approche par maximum de vraisemblance, ont été établis dans le modèle *M1-M0* par Baum *et* Petrie (1966) dans le cas stationnaire. On pourra consulter Mevel (1997), Le Gland and Mevel (2000), Douc (2000), pour des résultats dans le cas discret non stationnaire.
[12] Nous n'aborderons pas ici les versions stochastiques de *EM*, comme les algorithmes *SEM* et *EM à la Gibbs* (Muri 1998, Robert *et al.*, 1993).

Puisque les S_i ne sont pas observés, on remplace $\log f(X, S \mid \theta)$ par son espérance conditionnelle à la séquence X et à une valeur $\theta^{(m)}$ du paramètre, $Q(\theta \mid \theta^{(m)}) = \mathbb{E}_{\theta^{(m)}}(\log f(X, S \mid \theta) \mid X)$. Pour un modèle *M1-M1* [13],

$$Q(\theta \mid \theta^{(m)}) = \sum_{i=2}^{n} \sum_{u,v=1}^{q} \mathbb{P}_{\theta^{(m)}}(S_{i-1} = u, S_i = v \mid X) \log \pi_e(u, v) + \sum_{i=1}^{n} \sum_{u=1}^{q} \sum_{a,b \in \mathcal{A}} \mathbb{P}_{\theta^{(m)}}(S_i = u \mid X) 1_{\{X_{i-1}=a, X_i=b\}} \log \pi_o(u, a, b) \ .$$

Le calcul de cette espérance conditionnelle revient donc au calcul des probabilités conditionnelles de deux régimes successifs $\mathbb{P}_{\theta^{(m)}}(S_{i-1} = u, S_i = v \mid X)$ pour tout i.

On va donc mettre en jeu un algorithme où alternent deux phases :

Phase E : sachant $\theta^{(m)}$, calculer

$$\mathbb{P}_{\theta^{(m)}}(S_{i-1} = u, S_i = v \mid X) \quad \forall i = 2, \ldots, n, \ \forall u, v \in \mathcal{S}$$

Phase M : réactualiser $\theta^{(m)}$ en $\theta^{(m+1)} = \operatorname*{argmax}_{\theta} Q(\theta \mid \theta^{(m)})$.

La maximisation de la phase M est immédiate et conduit à calculer

$$\pi_e{}^{(m+1)}(u, v) = \frac{\sum_{i=2}^{n} \mathbb{P}_{\theta^{(m)}}(S_{i-1} = u, S_i = v \mid X)}{\sum_{i=2}^{n} \mathbb{P}_{\theta^{(m)}}(S_{i-1} = u \mid X)}$$

et $\pi_o^{(m+1)}(u, a, b)$ par une formule analogue.

Les probabilités de la *phase E* sont calculées par une récurrence « avant-arrière » sur les positions de la séquence. Si la notation X_1^i désigne les observations X_j pour $1 \leqslant j \leqslant i$, la récurrence avant permet d'obtenir les probabilités de filtrage et de prédiction, pour en déduire par récurrence arrière les probabilités de lissage [14].

Forward recurrence

1. Initialisation : $\mathbb{P}_{\theta^{(m)}}(S_1 = u \mid X_1^1) = \mu_e^{(m)}(u) \quad (u = 1, \ldots, q)$.
2. Pour $i = 2, \ldots, n$, calculer $\quad (u = 1, \ldots, q)$

$$\text{Filtrage} : \mathbb{P}_{\theta^{(m)}}(S_i = v \mid X_1^i) = \frac{\pi_o^{(m)}(v, X_{i-1}, X_i) \mathbb{P}_{\theta^{(m)}}(S_i = v \mid X_1^{i-1})}{\sum_{u=1}^{q} \pi_o^{(m)}(u, X_{i-1}, X_i) \mathbb{P}_{\theta^{(m)}}(S_i = u \mid X_1^{i-1})} \tag{3}$$

$$\text{Prédiction} : \mathbb{P}_{\theta^{(m)}}(S_i = v \mid X_1^{i-1}) = \sum_{u=1}^{q} \pi_e^{(m)}(u, v) \mathbb{P}_{\theta^{(m)}}(S_{i-1} = u \mid X_1^{i-1})$$

[13] Nous avons omis les premiers termes relatifs à S_1 et X_1.

[14] Cette formulation de la *phase E*, a été introduite par Churchill (1989) et est plus stable numériquement que celle proposée par Rabiner (1989).

Backward recurrence
1. Déduire $\mathbb{P}_{\theta^{(m)}}(S_n = u \,|\, X) = \mathbb{P}_{\theta^{(m)}}(S_n = u \,|\, X_1^n) \quad (u = 1, \cdots, q)$.
2. Pour $i = n, \ldots, 2$, calculer $\quad (u = 1, \ldots, q)$

$$\begin{aligned}\mathbb{P}_{\theta^{(m)}}(S_{i-1} = u \,|\, X) &= \sum_{v=1}^{q} \mathbb{P}_{\theta^{(m)}}(S_{i-1} = u, S_i = v \,|\, X) \\ &= \sum_{v=1}^{q} \frac{\pi_e^{(m)}(u,v)\mathbb{P}_{\theta^{(m)}}(S_{i-1} = u \,|\, X_1^{i-1})\mathbb{P}_{\theta^{(m)}}(S_i = v \,|\, X)}{\mathbb{P}_{\theta^{(m)}}(S_i = v \,|\, X_1^{i-1})}\end{aligned}$$

Redner and Walker (1984) ont démontré, dans le cadre de mélanges, que tout point limite de la suite $(\theta^{(m)})_{m\geqslant 0}$ générée par *EM* satisfait les équations de vraisemblance et que $(\theta^{(m)})_{m\geqslant 0}$ converge vers l'estimateur du maximum de vraisemblance si le point de départ $\theta^{(0)}$ n'est pas trop éloigné de la vraie valeur. Ces résultats s'étendent aux *HMM*, dès lors que $(\theta^{(m)})_{m\geqslant 0}$ est contenu dans un voisinage compact de la vraie valeur, *EM* peut donc converger vers un maximum local.

Les *phases E* et *M* sont itérées jusqu'à une itération M pour laquelle le critère d'arrêt $|\log g(X \,|\, \theta^{(M+1)}) - \log g(X \,|\, \theta^{(M)})| \leqslant \varepsilon$ est satisfait pour un ε donné : θ est alors estimé par $\theta^{(M)}$ et pour toute position i, la probabilité que le régime S_i soit dans l'état $v = 1, \ldots, q$ par $\hat{\mathbb{P}}(S_i = v) = \mathbb{P}_{\theta^{(M)}}(S_i = v \,|\, X)$. Remarquons que la log-vraisemblance des données complètes est facilement calculable (de manière récursive) à partir des probabilités de prédiction :

$$\log g(X \,|\, \theta) = \sum_{i=1}^{n} \log\Big(\sum_{v=1}^{q} \pi_o(v, X_{i-1}, X_i)\mathbb{P}_\theta(S_i = v \,|\, X_1^{i-1})\Big)$$

Estimation bayésienne par échantillonnage de Gibbs

Le nombre de paramètres d'un *HMM* peut être assez important (par exemple, un modèle *M1-M1* à 2 états cachés nécessite l'estimation de 26 paramètres si on travaille avec l'alphabet à 4 lettres `a, c, g, t`). Une estimation bayésienne paraît alors judicieuse pour utiliser et intégrer l'information a priori disponible sur le modèle.
Pour un coût quadratique, l'estimateur bayésien de θ est la moyenne a posteriori $\int_\Theta \theta\pi(\theta \,|\, X)d\theta$ où $\pi(\theta \,|\, X)$ désigne la loi a posteriori déduite de la vraisemblance et d'une loi a priori $\pi(\theta)$ sur θ par la formule de Bayes $\pi(\theta \,|\, X) \sim \pi(\theta)g(X \,|\, \theta)$.
Les lois a priori classiquement utilisées sont des lois (conjuguées) de Dirichlet (Robert *et al.* 1993, Muri, 1998, Boys *et al.*, 2000). Nous choisissons ainsi des lois de Dirichlet indépendantes pour chaque ligne $\pi_e(u)$ de la matrice de transition des régimes cachés Π_e, et chaque ligne $\pi_o(u,a)$ de la matrice Π_o : $\pi_e(u) \sim \mathcal{D}(\alpha(u,1), \ldots, \alpha(u,q))$ et $\pi_o(u,a) \sim \mathcal{D}(\beta(u,a,\mathtt{a}), \ldots, \beta(u,a,\mathtt{t}))$ pour tout $u \in \mathcal{S}$, et tout $a \in \mathcal{A}$.

Les méthodes de Monte Carlo par Chaînes de Markov, *MCMC*, (Robert, 1996) vont permettre d'échantillonner la loi d'intérêt $\pi(\theta \,|\, X)$ (non calculable, ni simulable), en construisant une chaîne de Markov dont la loi stationnaire est précisément la loi a posteriori $\pi(\theta \,|\, X)$. L'idée est encore d'augmenter les

données, et de simuler θ, non pas à partir de $\pi(\theta \,|\, X)$, mais à partir de la loi a posteriori conditionnelle

$$\pi(\theta \,|\, X, S) \sim \pi(\theta) f(X, S \,|\, \theta)$$

L'estimation bayésienne de θ par l'échantillonnage de Gibbs va alors consister à alterner les deux étapes suivantes :
Phase 1 : sachant $\theta^{(m)}$, simuler les régimes cachés $S^{(m+1)} = (S_1^{(m+1)}, \ldots, S_n^{(m+1)})$
Phase 2 : simuler $\theta^{(m+1)} \sim \pi(\theta \,|\, X, S^{(m+1)})$.
Les lois a posteriori (conditionnelles) $\pi(\theta \,|\, X, S^{(m+1)})$ sont encore des lois de Dirichlet et la *phase* 2 consiste à simuler chaque ligne, $1 \leqslant u \leqslant q$, $a \in \mathcal{A}$,

$$\pi_e(u) \sim \mathcal{D}\big(\alpha(u,1) + N^{(m+1)}(u\ 1), \ldots, \alpha(u,q) + N^{(m+1)}(u\ q)\big)$$
$$\pi_o(u,a) \sim \mathcal{D}\big(\beta(u,a,\mathtt{a}) + N^{(m+1)}(u,a\ \mathtt{a}), \ldots, \beta(u,a,\mathtt{t}) + N^{(m+1)}(u,a\ \mathtt{t})\big)\ .$$

$$\text{où}\ \ N^{(m+1)}(u\ v) = \sum_{i=2}^{n} 1_{\{S_{i-1}^{(m+1)}=u, S_i^{(m+1)}=v\}}$$

$$\text{et}\ \ N^{(m+1)}(u,a\ b) = \sum_{i=2}^{n} 1_{\{S_i^{(m+1)}=u, X_{i-1}=a, X_i=b\}}\ .$$

La *phase* 1 de simulation des régimes cachés peut s'effectuer de deux manières différentes (Robert *et al.* 1993, Muri, 1998). L'allocation peut être *globale* en simulant à chaque itération les régimes S_i selon leur loi jointe conditionnelle $\pi(S \,|\, X, \theta^{(m)})$ déduite de la relation

$$\mathbb{P}_{\theta^{(m)}}(S_1, \ldots, S_n \,|\, X) = \mathbb{P}_{\theta^{(m)}}(S_n \,|\, X) \ldots \mathbb{P}_{\theta^{(m)}}(S_i \,|\, S_{i+1}^n, X) \ldots \mathbb{P}_{\theta^{(m)}}(S_1 \,|\, S_2^n, X)$$

La *phase* 1 consiste alors à simuler les S selon $(1 \leqslant i \leqslant n)$

$$S_i \sim \mathcal{M}(1, \mathbb{P}_{\theta^{(m)}}(S_i = 1 \,|\, S_{i+1}^n, X), \ldots, \mathbb{P}_{\theta^{(m)}}(S_i = q \,|\, S_{i+1}^n, X))$$

où les probabilités $\mathbb{P}_{\theta^{(m)}}(S_i = u \,|\, S_{i+1}^n, X)$ se calculent par un algorithme « forward-backward » analogue à *EM*.

Backward recurrence
1. $\mathbb{P}_{\theta^{(m)}}(S_n = u \,|\, X) = \mathbb{P}_{\theta^{(m)}}(S_n = u \,|\, X_1^n) \quad (u = 1, \ldots, q)$.
2. Pour $i = n-1, \ldots, 1$, calculer $\quad (u = 1, \ldots, q)$

$$\mathbb{P}_{\theta^{(m)}}(S_i = u \,|\, S_{i+1}^n, X) \propto \pi_e^{(m)}(u, S_{i+1}) \mathbb{P}_{\theta^{(m)}}(S_i = u \,|\, X_1^i)\ ,$$

les probabilités de filtrage $\mathbb{P}_{\theta^{(m)}}(S_i = u \,|\, X_1^i)$ (et de prédiction) étant calculées par la récurrence avant (3).

L'allocation *locale* des régimes permet d'éviter cette double récurence en simulant les régimes $S_i^{(m+1)}$ composante par composante, selon la loi conditionnelle $\pi(S_i \,|\, S_{i'<i}^{(m+1)},\ S_{i''>i}^{(m)},\ \theta^{(m)}, X)$. Dans le modèle *M1-M1*, ces lois sont données par

$$\pi(S_1 \,|\, S_2^{(m)}, X_1, \theta) \propto \mu_e^{(m)}(S_1) \mu_o^{(m)}(S_1, X_1) \pi_e^{(m)}(S_1, S_2^{(m)})$$
$$\begin{aligned}\pi(S_i \,|\, S_{i'<i}^{(m+1)},\ S_{i''>i}^{(m)}, X, \theta^{(m)}) &= \pi(S_i \,|\, S_{i-1}^{(m+1)}, S_{i+1}^{(m)}, X_{i-1}, X_i, \theta^{(m)}) \\ &\propto \pi_e^{(m)}(S_{i-1}^{(m+1)}, S_i) \pi_o^{(m)}(S_i, X_{i-1}, X_i) \pi_e^{(m)}(S_i, S_{i+1}^{(m)})\end{aligned}$$
$$\pi(S_n \,|\, S_{n-1}^{(m+1)}, X_{n-1}, X_n, \theta^{(m)}) \propto \pi_e^{(m)}(S_{n-1}^{(m+1)}, S_n) \pi_o^{(m)}(S_n, X_{n-1}, X_n)$$

L'échantillonnage de génère ainsi une chaîne de Markov $(\theta^{(m)}, S^{(m)})_{m\geqslant 0}$ dont la loi stationnaire est par construction $\pi(\theta, S \,|\, X)$. La loi a posteriori $\pi(\theta \,|\, X)$ est la marginale en θ. Mais ces deux algorithmes génèrent deux chaînes en parallèle, $(S^{(m)})_{m\geqslant 0}$ qui est une chaîne de Markov uniformément géométriquement ergodique à espace d'états fini $\{1,\dots,q\}^n$, et la chaîne $(\theta^{(m)})_{m\geqslant 0}$, qui n'est pas une chaîne de Markov dans le cas de l'allocation locale des régimes cachés. Un principe de dualité entre ces deux chaînes, établi par Robert *et al.* (1993), permet de transférer les propriétés de $(S^{(m)})_{m\geqslant 0}$ à $(\theta^{(m)})_{m\geqslant 0}$: $(\theta^{(m)})_{m\geqslant 0}$ est aussi géométriquement uniformément ergodique et converge vers sa distribution limite $\pi(\theta \,|\, X)$. Par ergodicité, l'estimateur bayésien $\int_\Theta \theta\pi(\theta \,|\, y)d\theta$ sera approché par $\hat{\theta} = (1/M)\sum_{m=m_0}^{M+m_0} \theta^{(m)}$ et la probabilité de chaque régime caché S_i par $\hat{P}(S_i = u) = (1/M)\sum_{m=m_0}^{M+m_0} 1_{S_i^{(m)}=u\}}$ pour un m_0 assez grand pour avoir atteint la stationnarité [15].

3.2.3. Segmentation de la séquence lorsque le modèle est appris

Nous supposons ici que les paramètres du modèle θ sont parfaitement connus. On cherche à segmenter la séquence, c'est à dire à reconstruire la suite des régimes cachés S. L'algorithme le plus utilisé en biologie moléculaire est l'algorithme de Viterbi. Cet algorithme de programmation dynamique recherche la séquence d'états cachés la plus probable, s^*, pour un modèle *HMM* donné (voir par exemple Rabiner, 1989). L'algorithme consiste à calculer, par une équation de récurrence, la probabilité maximale P^* de générer la séquence X

$$\begin{aligned} P^* &= \max_{s_1,\dots,s_n} \mathbb{P}_\theta(S_1 = s_1,\dots,S_n = s_n | X_1,\dots,X_n) \\ &= \max_{s_1,\dots,s_n} \mathbb{P}_\theta(X_1,\dots,X_n, S_1 = s_1,\dots,S_n = s_n) \end{aligned}$$

pour en déduire le chemin le plus probable s^* permettant de générer X avec la probabilité P^*

$$s^* = \operatorname*{argmax}_{s_1,\dots,s_n} \mathbb{P}_\theta(X_1,\dots,X_n, S_1 = s_1,\dots,S_n = s_n)$$

P^* se calcule en fonction de la probabilité du chemin le plus probable (permettant de générer $X_1,\cdots,X_i$) qui se termine au site i dans l'état $S_i = v$. Nous noterons $z_i(v)$ cette probabilité

$$z_i(v) = \max_{s_1,\dots,s_{i-1}} \mathbb{P}_\theta(X_1,\dots,X_i, S_1 = s_1,\dots,S_{i-1} = s_{i-1}, S_i = v).$$

$\forall v \in \mathcal{S}$, $z_i(v)$ se calcule par la récurrence suivante

$$\begin{aligned} &z_1(v) = \mu_e(v)\mu_o(v, X_1) \\ &\forall i = 1,\cdots,n-1,\ z_{i+1}(v) = \pi_o(v, X_i, X_{i+1}) \max_{u\in\mathcal{S}} z_i(u)\pi_e(u, v) \end{aligned} \tag{4}$$

On en déduit P^* par $P^* = \max_{v\in\mathcal{S}} z_n(v)$.

[15] Des méthodes de contrôle de convergence des *MCMC* ont été implémentées dans le cadre des séquences biologiques par Muri *et al.* (1998), à la fois, par dualité, pour la chaîne des paramètres $(\theta^{(m)})_{m\geqslant 0}$, et pour la chaîne à espace d'états fini $(S^{(m)})_{m\geqslant 0}$, comme l'échantillonnage exact de la loi stationnaire de $(S^{(m)})_{m\geqslant 0}$ par la méthode de couplage arrière proposée par Propp and Wilson (1996).

Pour reconstruire le chemin le plus probable s^* réalisant la probabilité maximale P^*, il suffit de commencer par la dernière base s_n^*, puis de remonter étape par étape en suivant le cheminement inverse ayant mené au calcul de P^*. Pour cela, il est nécessaire de garder en mémoire l'état u permettant d'atteindre au site i l'état le plus probable v, c'est à dire celui qui maximise $z_i(v)$ dans la récurrence (4), et ce pour tout site i et tout état v.
Ainsi l'algorihtme de Viterbi s'écrit

Initialisation : $\forall v \in \mathcal{S}$,

$$\begin{aligned} z_1(v) &= \mu_e(v)\mu_o(v, X_1) \\ ptr_1(v) &= 0 \end{aligned}$$

Récurrence (avant) : $\forall i = 1, \ldots, n-1$, $\forall v \in \mathcal{S}$,

$$\begin{aligned} z_{i+1}(v) &= \pi_o(v, X_i, X_{i+1}) \max_{u \in \mathcal{S}} z_i(u)\pi_e(u, v) \\ ptr_{i+1}(v) &= \operatorname*{argmax}_{u \in \mathcal{S}} z_i(u)\pi_e(u, v) \end{aligned}$$

On en déduit

$$P^* = \max_{v \in \mathcal{S}} z_n(v) \quad \text{et} \quad s_n^* = \operatorname*{argmax}_{v \in \mathcal{S}} z_n(v)$$

puis par récurrence arrière $\forall i = n-1, \ldots, 1$

$$s_i^* = ptr_{i+1}(s_{i+1}^*)$$

En procédant de la même manière que la section 3.2.2, l'algorithme de Viterbi pourrait s'utiliser pour estimer simultanément les paramètres et les régimes cachés (en réactualisant les paramètres à chaque itération comme dans la section 3.2.1). Il est cependant largement utilisé à paramètres fixés en procédant en deux temps : le modèle est tout d'abord appris à partir d'un ensemble d'apprentissage (souvent difficile à constituer) de séquences les plus homogènes possibles dont la segmentation est connue. Une nouvelle séquence cible est alors segmentée par Viterbi à partir du jeu de paramètres préalablement estimé. Il est donc indispensable de recommencer la phase d'apprentissage du modèle dès que l'on cherche à segmenter une séquence « distante » de l'ensemble d'apprentissages. Il semble par exemple peu réaliste de segmenter une séquence humaine avec un modèle appris sur des bactéries ! De manière moins extrême, une séquence de bactérie pourra être segmentée à partir d'autres bactéries si ces dernières ne sont pas trop éloignées de la bactérie étudiée, par exemple la séquence cible et la bactérie utilisée pour l'apprentissage du modèle devront avoir un $\mathsf{g} + \mathsf{c}\%$ proche.

3.2.4. Quelques variantes des chaînes de Markov cachées

Nous nous proposons de présenter quelques variantes des *HMM* permettant de pallier certains défauts de ces modèles, essentiellement l'ordre fixé k de la chaîne de Markov utilisée pour modéliser la loi des observations conditionnellement aux régimes cachés, ainsi que la loi des temps séjour géométrique de chaque régime caché.

Les chaînes de Markov à ordre variable

Dans les *HMM* aujourd'hui mises en oeuvre, chaque régime correspond à une chaîne de Markov d'ordre fixe *Mk* (identique ou non d'un régime à un autre). Or, par exemple, avec un modèle *M5*, certains 5-mots peuvent être sous-représentés dans la séquence, et donc fournir une mauvaise estimation de la transition (à la base suivante), alors qu'il existe des mots de longueur 8 suffisamment présents pour donner une bonne estimation. Il est donc clairement nécessaire d'utiliser des chaînes dont l'ordre de dépendance varie selon ce qui précède chaque position.

Une première approche consiste à utiliser les *chaînes de Markov interpolées*, IMM_k (Interpolated Markov Model), introduites par Salzberg *et al.* (1998) pour la détection de gènes. L'idée des IMM est de trouver un compromis entre l'utilisation d'une chaîne de Markov d'ordre k le plus élevé possible pour avoir une qualité d'ajustement la meilleure possible, et un ordre k pas trop important pour estimer de manière satisfaisante les 4^k transitions. Une transition $\pi_o(w_k, a)$, d'un mot w_k de longueur k à une base a, sera bien estimée dès lors que les comptages $N(w_k\ a)$ et $N(w_k)$ seront suffisamment importants. A chaque position i de la séquence, on va donc se demander si le mot w_k est suffisamment représenté pour prédire correctement la base $X_i = a$. Un poids $\lambda_k(X_{i-1})$ est alors associé à chaque k-mot $w_k = X_{i-k} \dots X_{i-1}$, qui va refléter l'importance de w_k en termes de son nombre d'occurrences $N(w_k)$, et qui va mesurer une pondération entre les modèles IMM_k et IMM_{k-1}. On définit alors de proche en proche

$$IMM_k(X_i) = \lambda_k(X_{i-1})\pi_o(w_k, X_i) + (1 - \lambda_k(X_{i-1}))IMM_{k-1}(X_i)$$

Si w_k est fréquent, on choisit le modèle M_k pour estimer les transitions, et $\lambda_k(X_{i-1}) = 1$. Si w_k est « moins fréquent », et si la distribution observée [16] sur la séquence X des fréquences des $w_k a$ est significativement différente de la distribution estimée par les probabilités $IMM_{k-1}(X_i\ a)$, augmenter l'ordre n'est pas significatif et le modèle IMM_{k-1} est gardé : $\lambda_k(X_{i-1}) = 0$. Enfin, si w_k est moins fréquent, et si ces deux distributions sont significativement différentes, on pondère avec $\lambda_k(X_{i-1})$. Cette combinaison linéaire des probabilités de transitions permet donc de donner un poids plus important aux mots les plus représentés.

Ces modèles IMM sont implémentés dans plusieurs logiciels de détection de gènes : *Glimmer* pour les organismes procaryotes, *Glimmer2* (Delcher *et al.*, 1999, Salzberg *et al.*, 1998), et *EuGène* (Schiex *et al.*, 2000) pour les eucaryotes (la version actuelle de *EuGène* est dédiée à *Arabidopsis Thaliana*).

Une deuxième approche consiste à utiliser les modèles d'arbres de contexte, pour lesquels les probabilités d'apparitions des différentes lettres dépendent des suffixes de longueurs variables précédents (Vert, 2001).

Les semi-chaînes de Markov cachées

[16] Ceci implique d'avoir calculé au préalable tous les comptages $N(w_k a)$ pour tout mot w_k et toute base a de l'alphabet, et ce pour tout ordre k compris entre 0 et un ordre maximal K.

La modélisation par HMM d'une séquence implique que la loi du temps de séjour dans un état u est une loi géométrique de paramètre $1-\pi_e(u,u)$: si on note, $d_u(t)$ la probabilité de rester dans l'état u pendant un temps t, alors,

$$d_u(t) = \pi_e(u,u)^{t-1}(1-\pi_e(u,u))$$

et la longueur moyenne d'une plage de l'état u est $1/(1-\pi_e(u,u))$. Cette contrainte imposée aux longueurs de régions homogènes n'est pas toujours compatible avec l'application biologique considérée notamment pour l'annotation de gènes eucaryotes (distribution non géométrique des introns et des exons). Les modèles de semi-chaînes de Markov cachées $HSMM$ (Hidden Semi-Markov Model) permettent précisément de prendre en compte les lois des longueurs de chaque région. Un modèle *SM1-Mk* est un modèle *M1-Mk* où l'on a remplacé la chaîne de Markov *M1*, $(S_i)_{i\geqslant 0}$ par une semi-chaîne de Markov $SM1$. La semi-chaîne de Markov est définie par une suite de temps d'arrêt $(J_m)_{m\geqslant 1}$ et un processus $(S_i)_{i\geqslant 0}$ tels que $(S_{J_m})_{m\geqslant 1}$ soit une chaîne de Markov de transitions définies par :

– si u est un état non absorbant, alors $\forall v \neq u$,

$$\pi_e(u,v) = \mathbb{P}(S_{J_m} = v | S_{J_{m-1}} = u)$$

avec $\sum_{v\neq u} \pi_e(u,v) = 1$ et $\pi_e(u,u) = 0$;
– si u est un état absorbant, $\pi_e(u,u) = 1$ et $\pi_e(u,v) = 0$, $\forall v \neq u$.

Les temps d'arrêt $(J_m)_{m\geqslant 1}$ sont donc les instants de saut du processus $(S_i)_{i\geqslant 1}$: $J_m = \inf\{i; i > J_{m-1}/S_i \neq S_{J_{m-1}}\}$. On note $(T_m)_{m\geqslant 1}$, la suite des temps de séjour, $T_m = J_m - J_{m-1}$. La loi des temps de séjour est alors définie par $d_u(t) = \mathbb{P}(J_{m+1} - J_m = t \,|\, S_{J_m} = u)$, et on note $\Delta_u(t) = \sum_{t'\geqslant t} d_u(t')$, la probabilité de rester dans l'état u pendant un temps supérieur ou égal à t.

Pour une séquence de longueur n, on définit $M_n = \inf\{m/J_m > n\}$. Avant n, il y a donc $M_n - 1$ sauts, et le dernier instant de saut avant n est J_{M_n-1}. Une des difficultés des $HSMM$ provient du fait que le dernier temps de séjour est partiellement observé [17] : $\sum_{m=1}^{M_n} T_m \geqslant n$, ou encore $J_{M_n} > n$.

La vraisemblance (des données complètes) est évidemment moins facile à manipuler que pour les HMM, et les algorithmes d'identification des régimes cachés plus lourds à mettre en oeuvre. On a dans le modèle *SM1-M1*,

$$f(X,S\,|\,\theta) = \mu_e(S_1)\mu_o(S_1,X_1) \prod_{u,v\in\mathcal{S},v\neq u} \pi_e(u,v)^{N(uv)} \times \prod_{v\in\mathcal{S}} \prod_{t\in\mathbb{N}^*} d_v(t)^{N_t(v)} \Delta_{M_n} \prod_{v\in\mathcal{S}} \prod_{a,b\in\mathcal{A}} \pi_o(v,a,b)^{N(v,ab)}$$

où $N_t(v)$ représente le nombre de fois où est on resté un temps t dans l'état v, et $\Delta_{M_n} = \sum_{t>n-J_{M_n-1}} d_t(u_{M_n})$, u_{M_n} étant le dernier état caché visité.

[17] Dans les applications biologiques utilisant les $HSMM$ comme la détection de gènes, il est d'ailleurs souvent supposé que $\sum_{m=1}^{M_n} T_m = n$, ce qui simplifie grandement les calculs mis en jeu dans les algorithmes, mais introduit un biais dans les résultats.

L'algorithme *EM* a été développé et implémenté au CIRAD dans le logiciel *AMAPmod*[18] par Guédon and Cocozza-Thivent (1990), Guédon (1998), Guédon (1999), dans le cadre de la modélisation de l'architecture des plantes : l'étape M de maximisation est plus difficile à mettre en oeuvre que pour les *HMM*, puisque les lois de temps de séjour doivent être également estimées (de manière paramétrique ou non paramétrique). Nous ne présenterons donc qu'une version simplifiée de l'algorithme de Viterbi, seul algorithme utilisé dans les applications biologiques utilisant des $HSMM$. Nous ne détaillerons que le calcul de $P^* = \max_v z_n(v)$, qui se déduit du calcul, pour tout $i = 1$ à $n-1$, de la probabilité $\tilde{z}_i(v)$, du chemin le plus probable se terminant (exactement) en i dans l'état $S_i = v$:

$$\tilde{z}_i(v) = \max_{s_1,\ldots,s_{i-1}} \mathbb{P}_\theta(X_1,\ldots,X_i, S_1 = s_1,\ldots,S_{i-1} = s_{i-1}, S_i = v, S_{i+1} \neq v)$$

Les $\tilde{z}_i(v)$ se calculent (comme pour les HMM) de manière récursive en distinguant deux cas : on est resté dans le même état v jusqu'à l'instant i, ou on a changé au moins une fois d'état. Pour le deuxième cas, $\tilde{z}_i(v)$ s'écrit en fonction du chemin le plus probable se terminant en $i-t$ dans l'état $u \neq v$ (on reste exactement un temps t dans l'état v), en étudiant toutes les configurations possibles pour u et t. Pour $i = 2,\ldots,n-1$, on a alors

$$\begin{aligned}\tilde{z}_i(v) = \max \Big[& \mu_e(v)\mu_o(v,X_1)d_v(t)\prod_{j=2}^{t}\pi_o(v,X_{j-1},X_j)\ , \\ & \max_{1\leqslant t<i}\max_{u\neq v}\tilde{z}_{i-t}(u)\pi_e(u,v)d_v(t)\prod_{j=1}^{t}\pi_o(v,X_{i-j},X_{i-j+1})\Big] \\ = \max_{1\leqslant t\leqslant i}\Big(& \prod_{j=1}^{t}\pi_o(v,X_{i-j},X_{i-j+1})d_v(t)\max_{u\neq v}\{\tilde{z}_{i-t}(u)\pi_e(u,v)\}\Big)\end{aligned}$$

où par convention $\prod_{j=1}^{0}\pi_o(v,X_{j-1},X_j) = \mu_o(v,X_1)$ et $\tilde{z}_0(v)\pi_e(u,v) = \mu_e(v)$, $\forall v \in \mathcal{S}$. Pour $t = 1$, on initialise l'algorithme par

$$\tilde{z}_1(v) = \mathbb{P}(X_1, S_1 = v, S_2 \neq v) = \mu_e(v)d_j(1)\mu_o(v,X_1),\ \forall v \in \mathcal{S}$$

Pour $t = n$, on obtient

$$z_n(j) = \max_{1\leqslant t\leqslant n}\Big(\prod_{j=1}^{t}\pi_o(v,X_{n-j},X_{n-j+1})\Delta_v(t)\max_{u\neq v}\{\tilde{z}_{n-t}(u)\pi_e(u,v)\}\Big)\ .$$

Pour obtenir le chemin le plus probable S^*, il faut garder en mémoire les trajectoires arrivant dans chacun des états possibles v pour chaque temps de séjour t possible en chaque instant i. Les pointeurs $ptr_i(v)$ seront donc des vecteurs à deux composantes : la première donnera la durée la plus probable t pendant laquelle on est resté dans l'état v, et la deuxième donnera la valeur u de l'état le plus probable visité juste avant l'état v (dans lequel on est resté un temps t).

[18] http://www.cirad.fr/presentation/programmes/amap.html Les algorithmes *EM* et Viterbi sont implémentés dans AMAPmod dans le cadre des *HSMM* (Godin *et al.*, 1997, 1999).

4. Quelques applications biologiques

Nous ne prétendons pas ici faire le tour de toutes les applications biologiques où les *HMM* (et leurs variantes) sont et peuvent être utilisées. Nous nous contenterons de présenter deux applications : la première montre comment on peut décrire simplement, mais efficacement, l'hétérogénéité d'une séquence biologique avec des *HMM*. La deuxième application est un exemple d'emploi de *HMM* très structuré pour décrire un problème biologique précis et complexe, à savoir la détection de gènes.

4.1. Hétérogénéité des bactéries - Transferts de gènes

La première application que nous allons présenter consiste à utiliser directement les *HMM* tels que nous les avons décrits précédemment, pour décrire, sans aucun a priori, l'hétérogénéité d'une séquence biologique. Cette utilisation « à l'aveugle » des HMM a été utilisée par Churchill (1989), Muri (1998), Boys *et al.* (2000) pour segmenter une séquence en régions homogènes, sur la seule base de leur composition locale en oligonucléotides.

Nous allons présenter les résultats d'une telle utilisation des *HMM* sur la bactérie *Bacillus subtilis*. Cette étude, fruit d'une collaboration entre l'Upresa 8071, et l'Unité MIG [19] (Mathématique Informatique et Génome) de l'INRA (Versailles), a permis, d'une part, de mettre en évidence différents niveaux d'hétérogénéité chez *B. subtilis*, et d'autre part, de révéler l'existence de nouveaux transferts de gènes putatifs (Bize *et al.*, 1999, Nicolas, 2000).

Les transferts génétiques horizontaux, opposés aux transferts verticaux de parents à enfants, peuvent avoir lieu entre organismes éloignés phylogénétiquement et sont considérés aujourd'hui comme un mécanisme majeur de l'évolution des génomes microbiens. Ils sont impliqués dans la transmission de facteurs de résistance et de virulence entre espèces microbiennes. On connaît chez les bactéries trois mécanismes d'échange génétique : la transformation (absorption d'un ADN étranger), la transduction (un phage, virus infectant les bactéries, transfère des gènes bactériens d'une cellule hôte à une autre) et la conjugaison (transfert direct de matériel génétique entre deux cellules bactériennes temporairement liées).

L'intégration du matériel génétique échangé peut faire intervenir des éléments génétiques mobiles, comme les transposons, segments d'ADN pouvant se déplacer d'un endroit à un autre à l'intérieur du génome. Les transposons simples contiennent un seul gène codant pour la transposase, enzyme qui catalyse le déplacement du transposon d'un endroit à l'autre du génome. Aux deux extrémités du gène de la transposase se trouvent des répétitions inversées, qui seront rapprochées par la transposase au moment de la transposition. L'enzyme catalysera la coupure et le raccordement de l'ADN nécessaire à l'insertion du transposon sur un certain site. Les transposons complexes comportent, en plus du gène de la transposase, d'autres gènes, comme par exemple des gènes de résistance aux antibiotiques. Ces transposons complexes permettraient à la bactérie de s'adapter à des milieux nouveaux.

[19] Plus précisément, les personnes impliquées dans ce travail sont Pierre Nicolas, Philippe Bessières, François Rodolphe, Mark Hoebeke et Laurent Bize, aujourd'hui au Département BIA de l'INRA de Toulouse.

On peut alors penser que les transferts de gènes correspondent à des segments d'ADN possèdant des propriétés statistiques différentes de celles de l'hôte : une segmentation de la séquence par *HMM* devrait alors permettre de les détecter. Le chromosome complet de *B. subtilis* (environ 4,2 millions de paires de bases que nous avons découpées en fragments de 200 kb) a été segmenté en utilisant le logiciel évolutif $R'HOM$ [20]. Ce logiciel, que nous avons créé pour permettre aux biologistes d'utiliser les *HMM* de manière automatique, regroupe plusieurs algorithmes d'identification (*EM* et méthodes *MCMC*), et produit, entre autres, pour chaque état caché v, une représentation graphique des probabilités estimées des régimes $\hat{P}(S_i = v)$ pour toute position i de la séquence, permettant ainsi une visualisation des régions homogènes détectées. Pour exploiter les résultats, des informations biologiques (comme les annotations au format Genbank de la séquence, ou fournies par la base de données relationnelles MICADO [21]) sont superposées à la segmentation obtenue par $R'HOM$.

Pour *B. subtilis*, il semble qu'un modèle *M1-M2* [22] à $q = 3$ états cachés soit suffisant pour décrire un premier niveau d'hétérogénéité. La segmentation obtenue distingue deux états correspondant aux régions codantes dans les deux sens, et un état qui regroupe les régions intergéniques et/ou, selon les fragments analysés, le codant minoritaire, que nous appellerons codant atypique et où nous chercherons la présence de transferts de gènes. Une augmentation du nombre d'états cachés (4 ou plus) conduit soit à diviser un état en deux en fonction du sens du codant, soit à isoler des classes de gènes particulières, comme des protéines impliquées dans le transport membranaire [23].

La figure 2 (Nicolas, 2000) montre la segmentation obtenue sur les 200000 premières bases de *subtilis* avec 6 états cachés. Cet exemple est remarquable car, outre la séparation des régions codantes en fonction de leur sens (en cyan et magenta) et l'identification précise des régions intergéniques (en noir), il fait apparaître très nettement deux types d'hétérogénéités non liées à des transferts : les ARN structuraux [24], en rouge, à l'intérieur desquels se dessinent très nettement des structures caractérisées par l'état bleu (le motif des ruptures est quasi identique d'un ARN à l'autre), et les protéines ribosomales, en jaune, regroupant les protéines impliquées dans la traduction et la transcription, qui se différencient nettement du contexte chromosomique local (certainement en raison de leur niveau d'expression élevé).

La figure 3 (Nicolas, 2000) présente un cas de transfert suspecté dans la région $4160 - 4200$ *kbp* de *B. subtilis* : il s'agit d'une région pointée comme atypique par $R'HOM$, possèdant des indices biologiques de recombinaison génétique, comme des répétitions aux extrémités, et des longues régions intergéniques, ainsi que la présence de gènes (ou des homologies, si la fonction est

[20] **R**echerche d'**HOM**ogénéité dans une séquence d'ADN. $R'HOM$ est disponible sur le site http ://www.inra.fr/bia/J/AB/genome.html (en cours de perfectionnement et de développement).

[21] http://locus.jouy.inra.fr/micado .

[22] Les modèles d'ordre inférieur donnent un découpage en plages de quelques paires de bases, difficilement exploitables. L'augmentation de l'ordre $k > 2$ du modèle demande d'estimer un nombre trop important de paramètres par rapport au gain obtenu.

[23] Ces protéines sont probablement distinguées par leur hydrophobicité.

[24] Opposés à l'ARN messager, ce sont les ARN stables non traduits en acides aminés comme les ARN de transfert et ribosomaux.

inconnue, recherchées par le programme d'alignement FASTA) impliqués dans des fonctions de résistance et virulence. Signalons qu'un nouveau transposon a été mis en évidence par cette analyse. Notons enfin que 9 prophages (phages intégrés dans le chromosome de l'hôte) sur 10 reportés dans la littérature, ont été détectés dans les régions atypiques, validant ainsi la méthode. Le prophage non détecté, $PBSX$, a une composition très proche du contexte local de *B. subtilis*, ce qui pourrait s'expliquer par une longue évolution dans l'hôte.

4.2. Détection de gènes

Il existe de très nombreux détecteurs de gènes fondés sur des HMM comme *Genscan* (Burge and Karlin, 1997), *Veil* (Henderson *et al.*, 1997), *Genie* (Kulp *et al.*, 1996), *Glimmer* (Salzberg *et al.*, 1998), *Grail* (Xu *et al.*, 1994), *GeneMark.hmm* (Lukashin and Borodovsky, 1998)... et nous ne prétendons pas en faire ici une liste exhaustive, ni décrire le fonctionnement de chacun. Nous souhaitons simplement montrer, à travers quelques exemples, comment, de par leur caractère flexible et adaptatif, les HMM structurés permettent de modéliser des phénomènes biologiques très complexes. Rappelons tout d'abord brièvement (et grossièrement !) la structure d'un gène, qui est différente chez les procaryotes et les eucaryotes. Comme souvent dans le contexte de la détection de gènes, nous avons abusivement désigné par « gène » toute partie d'une séquence d'ADN codant pour une protéine. La synthèse protéique se fait en deux étapes : la transcription et la traduction. Chez les procaryotes (figure 4),

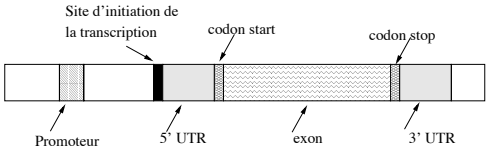


FIG. 4. *Structure d'un gène procaryote (échelle non respectée)*

l'ADN est tout d'abord transcrit en ARNm messager (il s'agit d'une « copie » du gène où la base `t` est remplacée par la lettre `u`) de l'extrémité $5'$ vers l'extrémité $3'$ par l'ARN polymérase. La transcription commence en amont du codon start (initiateur de la traduction) et se termine en aval du codon stop (fin de traduction). Les régions, qui se trouvent aux extrémités des régions codantes, transcriptes en ARNm, mais non traduites s'appellent les UTR. L'ARN polymérase commence la transcription en se liant à l'ADN au niveau du site promoteur (séquence particulière jouant un rôle régulateur [25]), en amont du

[25] Les régions promotrices contiennent plusieurs séquences régulant la transcription, comme les boîtes `tata`, situées généralement à -25 *bp* du début de l'unité de transcription.

site d'initiation de la transcription. La traduction de l'ARNm en acides aminés constituant la protéine, commence au codon start et se termine au codon stop. Chez les eucaryotes (figure 5), le mécanisme d'expression des gènes est

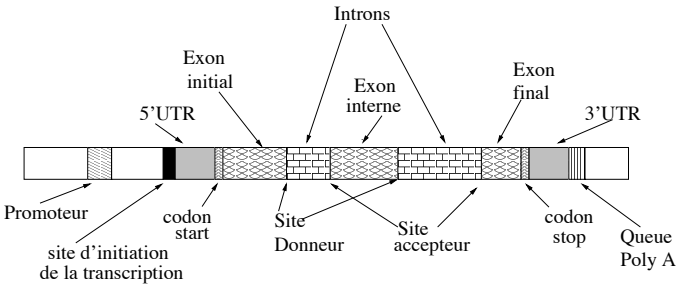


FIG. 5. *Structure d'un gène eucaryote (échelle non respectée)*

plus complexe, car les régions codantes ne sont généralement pas continues, mais composées d'une alternance d'exons et d'introns (le terme exon initial désignera le premier exon rencontré, exon final le dernier, et exon interne tout exon compris entre deux introns). Durant la transcription, les introns et exons sont transcrits en ARN messager à l'intérieur du noyau. L'ARNm sortira du noyau en laissant derrière lui les introns, il s'agit du phénomène (très complexe) d'épissage [26]. Les frontières exons/introns sont signalées par de courtes séquences spécifiques, appelées sites d'épissage. Le début d'un exon est le site accepteur, et la fin d'un exon est le site donneur. Ces séquences spécifiques ne sont pas des motifs strictement conservés d'une séquence à l'autre, mais des séquences contenant des parties (plus ou moins) conservées, à partir desquelles on peut fabriquer ce que l'on appelle une séquence consensus [27]. Les transcrits d'exons sont alors recollés entre eux pour former un ARNm mature (possèdant à l'extrémité 5' une CAP et à l'extrémité 3' une queue poly A, où le nucléotide `a` est répété plusieurs fois). C'est l'ARNm mature, où les bases sont regroupées trois par trois pour former les codons, qui sera alors traduit en acides aminés dans le cytoplasme.

[26] Pour un même transcrit, il peut exister différentes excisions d'introns : on parle alors d'épissage alternatif.

[27] Le site donneur doit contenir le consensus minimal `gt`, et le site accepteur `at`. Chez l'homme, le site donneur comprend les 3 derniers nucléotides de l'exon, et les six premiers de l'intron. La séquence consensus du donneur est `[c/a]AGGT[a/g]AGt`, les crochets indiquant les nucléotides possibles à une position donnée, et les majuscules désignant les nucléotides de fréquence supérieure à 50% ; les deux premiers nucléotides de l'intron `GT` sont fixes d'une séquence à l'autre.

Les *HMM* se révèlent être des outils très performants pour détecter et annoter un gène. La partie cachée du modèle va bien entendu correspondre à la syntaxe et la structure du gène que l'on veut détecter.

Les procaryotes

La structure du modèle est « relativement simple » pour détecter des gènes chez les procaryotes. Au minimum, il faut disposer d'un état modélisant le non codant (NC), d'un état Codant sur le brin direct ($C+$), d'un état codant sur le brin complémentaire ($C-$), d'états modélisant le codon Start dans les sens direct et indirect ($Start+$, $Start-$) et d'états modélisant le codon Stop ($Stop+$, $Stop-$). Bien sûr, certaines transitions entre états seront interdites, il est par exemple impossible de passer directement de l'état NC à l'état $C+$ sans passer par l'état $Start-$. Il est aussi possible de rajouter des informations complémentaires, par exemple, sur le codant. Comme nous l'avons vu dans l'application précédente, il peut exister plusieurs classes de codant : du codant typique (Ct) à la bactérie étudiée, opposé au codant atypique (Ca) contenant, entre autres, des transferts de gènes. C'est ce que fait le programme *GeneMark.hmm* (Lukashin and Borodovsky, 1998), dont l'architecture est représentée dans la figure 6.

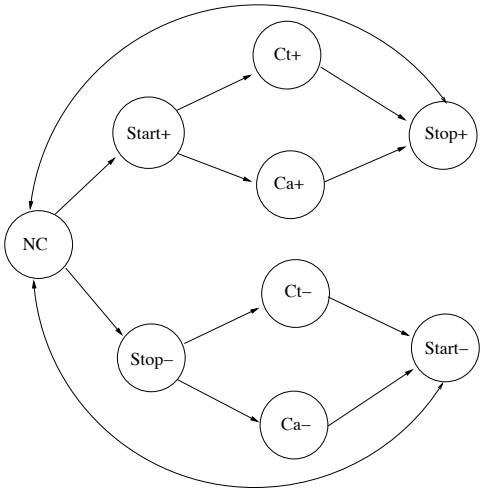


FIG. 6. *Structure du HMM utilisé dans le détecteur de gènes procaryotes GeneMark.hmm.*

Pour détecter des régions codantes, il est nécessaire de prendre en compte la phase, c'est à dire la structure en codons. Une possibilité est de modéliser le codant non plus par un seul état, mais par trois états correspondant chacun à une phase (figure 7). On peut alors modéliser d'éventuelles délétions ou insertions dans le codant en autorisant, par exemple, une transition entre l'état 1 et 3, et l'état 1 vers lui-même. C'est ce que peut faire le logiciel *SHOW*[28] (Nicolas, 2001), parfaitement adapté pour détecter des gènes sur des séquences

[28] **S**tructured **HO**mogeneities **W**atcher, en cours de développement.

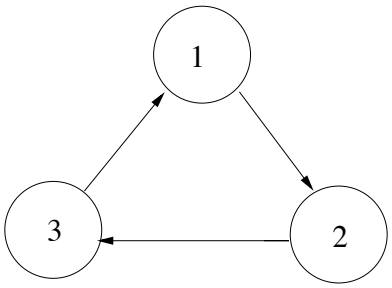


FIG. 7. *Prise en compte de la phase d'une région codante.*

microbiennes de mauvaise qualité (comportant un certain nombre de délétions et d'insertions). De manière générale, *SHOW* est un *HMM* très flexible où l'utilisateur peut lui même structurer le modèle en fonction de son problème biologique (en imposant, ou interdisant des transitions, en choisissant l'ordre de la chaîne de Markov utilisée pour générer les bases dans chaque état...). *SHOW* permet d'ajuster un *HMM* structuré par l'algorithme *EM* et ne nécessite donc pas d'ensemble d'apprentissage pré-segmenté.

Une fois que la structure du modèle est arrêtée, il reste à définir les lois régissant l'apparition des bases dans chaque état. Le codon start est choisi parmi les trois codons `atg`, `gtg`, ou `ttg`, le codon `atg` étant largement majoritaire. Il est possible de modéliser par exemple chaque base du codon Start par un modèle *M0* (un codon Start est alors composé de trois états). On procède évidemment de manière similaire pour les codons stop `taa`, `tag`, `tga`. Pour le codant, les modèles classiquement utilisés sont des chaînes de Markov (inhomogènes pour prendre en compte la phase) d'ordre 5, modélisant la transition d'un codon à un autre. Un modèle *M1* ou *M2* est suffisant pour le non codant. De plus en plus de détecteurs de gènes sont en fait des *HSMM*, appelés encore *GHMM* (*HMM* généralisé), pour autoriser un temps de séjour plus adéquat qu'une distribution géométrique, dans chaque état. *GeneMark.hmm* choisit par exemple de modéliser la longueur des états codants par une loi Gamma tronquée. Comme la plupart des détecteurs de gènes, *GeneMark.hmm* apprend les paramètres du modèle à partir d'un ensemble d'apprentissage de gènes (plus ou moins bien annotés). Tous les paramètres sont alors fixés, et l'algorithme de Viterbi est utilisé pour détecter et annoter la structure d'un nouveau gène pouvant provenir d'une autre bactérie que celle utilisée pour faire l'apprentissage du modèle. Certains paramètres peuvent toutefois être réestimés d'une bactérie à l'autre par une procédure de classification assez lourde à mettre en œuvre, il s'agit des transitions des régions codantes et non codantes (le programme *GeneMark-Genesis*, Hayes and Borodovsky, 1998, réalise cet apprentissage à partir des ORF de la séquence étudiée). Notons que la structure du modèle choisie dans *GeneMark.hmm* ne permet pas de prendre en compte d'éventuels chevauchements de gènes. Une façon de modéliser les chevauchements serait de passer d'une région codante à une autre sans rencontrer de codon start, ce qui revient à autoriser une transition entre un codon stop et une région codante

dans une phase différente (le logiciel *SHOW* permet d'imposer facilement ce genre de contraintes).

Les eucaryotes

Chez les eucaryotes, la structure du modèle devra être beaucoup plus sophistiquée pour espérer détecter et annoter correctement un gène : le cadre de lecture devra être maintenu après le passage par un intron, de nombreux signaux difficiles à modéliser devront être pris en compte comme les sites d'épissage, l'excision des introns peut s'effectuer de manière alternative... Les modèles HMM utilisés pour annoter les gènes eucaryotes sont souvent des HMM « modulaires » où chaque état correspond à un « module » décrivant le mieux possible une structure bien précise du gène. Un module peut lui-même être décrit par un modèle très sophistiqué comme un HMM, un réseau de neurones... L'inter-connection des modules se fait par la chaîne de Markov cachée, les transitions entre états devant être compatibles avec l'ordre des unités fonctionnelles à visiter imposé par la structure du gène. Les différents états (ou modules) à modéliser sont les exons, les introns, les régions intergéniques, les UTR et différents signaux comme les promoteurs, les sites d'épissage, les queues poly-A, les codon start et stop... Une grande difficulté à laquelle sont confrontés les détecteurs de gènes est la grande variabilité existant entre les gènes d'une même espèce : il existe des petits et des grands exons (et introns), les propriétés structurelles ne sont pas les mêmes selon le $\mathsf{g}+\mathsf{c}\%$ de la séquence étudiée... La prise en compte de toute cette information peut bien sûr faire rapidement exploser le nombre d'états cachés utilisés dans le modèle.

Nous allons présenter brièvement quelques aspects de la modélisation utilisée dans les détecteurs de gènes eucaryotes les plus employés. Nous nous centrerons particulièrement sur le logiciel *Genscan* (Burge and Karlin, 1997) dédié à l'annotation des gènes humains. Tous ces logiciels utilisent l'algorithme de Viterbi pour segmenter une séquence à partir d'un modèle préalablement appris sur un ensemble d'apprentissage de gènes annotés (qui sera d'autant plus important que l'information que l'on souhaite incorporer dans le modèle est grande et diversifiée).

L'architecture de *Genscan* ne comporte pas moins de 27 états cachés : région intergénique, promoteur, UTR $5'$, Exon-simple (en l'absence d'introns), Exons initial, interne (trois états pour chacune des phases) et final, UTR $3'$, queue poly-A, Intron (trois pour chaque phase). Les signaux codons start et stop, sites d'épissage sont eux-mêmes des sous-composantes des différents états exons (ce qui diminue considérablement le nombre d'états). Ainsi, un exon-simple comprend le codon start, la région codante, et le stop, un exon initial commence au codon start et se termine par le site donneur, un exon interne démarre avec un site accepteur et se termine par le site donneur, l'exon final débutant par le site accepteur et finissant par le stop (un état intron ne comprend donc aucun site d'épissage). Des états analogues sont bien sûr définis sur le brin complémentaire (excepté pour les régions intergéniques). Remarquons que les trois états introns sont introduits uniquement pour maintenir le cadre de lecture : un intron qui est visité juste après la fin d'un codon est en phase 0, après la première base d'un codon, en phase 1... La phase de l'intron précédant

un exon interne déterminera la position au sein du codon du premier nucléotide de l'exon, et le cadre de lecture sera donc gardé en mémoire.

Les détecteurs de gènes eucaryotes sont essentiellement fondés sur des modèles de semi-chaînes de Markov cachées, la longueur des exons et des introns n'étant clairement pas modélisable par des lois géométriques. De plus, il existe une grande variabilité des longueurs des différentes unités fonctionnelles constituant un gène. Ainsi, on ne choisira pas la même distribution des temps de séjour dans les états exon-simple, initial, interne ou final. La longueur des introns peut également varier considérablement en fonction du $\mathtt{g}+\mathtt{c}\%$ de la séquence. *Genscan* prend en compte cette information en répartissant les séquences (de manière assez arbitraire) en quatre catégories selon leur $\mathtt{g}+\mathtt{c}\%$: I : $< 43\%$ $\mathtt{g}+\mathtt{c}$, II : $43-51\%$ $\mathtt{g}+\mathtt{c}$, III : $51-57\%$ $\mathtt{g}+\mathtt{c}$, IV : $> 57\%$. Par exemple, les introns du groupe I sont en moyennes quatre fois plus grands que ceux du groupe IV. Notons, que pour maintenir le cadre de lecture, la longueur d'un exon interne doit être cohérente avec la phase des deux introns adjacents : si l'intron précédent est en phase 2, et le suivant en phase 1, alors la longueur de l'exon (ici en phase 2) doit être égale à $3t+2$, où t est généré selon la loi du temps séjour dans l'état exon interne.

De nombreuses autres caractéristiques fonctionnelles et structurales du gène varient en fonction du $\mathtt{g+c}\%$, comme le nombre de gènes, le nombre d'exons simples. Ceci suggère d'utiliser des jeux de paramètres différents pour décrire correctement les gènes appartenant aux différents groupes. C'est ce qu'y est précisément fait dans le logiciel *Genscan*, où, par exemple, les transitions entre états sont estimées séparément à partir d'un ensemble d'apprentissage de chaque groupe.

Comme pour les procaryotes, les régions codantes (exon) sont souvent modélisées par des chaînes de Markov inhomogènes d'ordre 5, le non codant (intergénique, introns, UTR) pouvant être modélisé par une *CM* homogène (d'ordre 5 ou moins élevé).

Il est clair que l'estimation préalable de tous ces paramètres (*une CM* d'ordre 5 nécessite l'estimation de 4^6 paramètres, représentant à peu près pas moins de 3000 paramètres indépendants) exige un ensemble d'apprentissage gigantesque. À ce jour, le nombre de gènes humains parfaitement annotés n'étant pas (encore) assez conséquent, beaucoup de séquences utilisées par *Genscan* ne sont en fait que des gènes partiels ou des $cDNA$ [29].

Détection de signaux

Le dernier point que nous allons abordé très succintement concerne la modélisation délicate des signaux biologiques comme les promoteurs, les sites d'épissage... La détection de ces signaux est essentielle pour annoter correctement un gène, par exemple la prédiction exacte des frontières des exons passe obligatoirement par la reconnaissance des sites d'épissage. Des logiciels comme *Veil* ou *Genie* utilise des réseaux de neurones (plus ou moins complexes) pour reconnaître les sites d'épissage. Mais les signaux biologiques sont souvent détectés à partir de profils. Un profil permet de définir ce qui est « conservé » dans une famille de séquences, par exemple un ensemble de sites d'épissages. À partir

[29] Séquences d'ADN obtenues par complémentarité de séquences partielles d'ARN observé dans la cellule.

d'un alignement [30] multiple des séquences de la famille, disons de longueur L, nous désignerons par profil $\mathcal{P}$ l'ensemble des probabilités $p_i(a)$ d'observer la base $a \in \mathcal{A}$ dans la colonne i ($i = 1, \ldots, L$) de l'alignement. Il est alors possible de définir à partir du profil un représentant de la famille, que nous avons appelé auparavant séquence consensus, comme étant la séquence constituée des nucléotides a dont la probabilité $p_i(a)$ est la plus forte en chaque position i.

Ces profils permettent de comparer une séquence avec ce qui a été le plus conservé dans une famille, et de prédire et détecter des signaux particuliers. À partir des probabilités $p_i(a)$ (par exemple estimées sur un ensemble de sites donneurs), on peut calculer la probabilité de générer une séquence $X_D = (X_1, \ldots, X_9)$ candidate à être un site donneur

$$P(X_D \,|\, X_D \text{ est un site donneur}) = \prod_{i=1}^{9} p_i(X_i)$$

que l'on peut comparer à la probabilité d'observer X_D sous le modèle d'indépendance des bases du génome étudié

$$P(X_D \,|\, X_D \text{ « n'est pas un site donneur »}) = \prod_{i=1}^{9} p(X_i) \ .$$

On peut alors évaluer le score d'alignement de X_D contre le profil en calculant

$$S(X_D \,|\, \mathcal{P}) = \sum_{i=1}^{9} \log\big(\frac{p_i(X_i)}{p(X_i)}\big) \ .$$

Ces profils sont utilisés dans *Genscan* pour modéliser les queues poly-A, les sites d'initiations de la traduction, les promoteurs et les sites accepteurs (en remplaçant pour ces derniers l'hypothèse d'indépendance par une dépendance markovienne). Pour les sites donneurs, un modèle plus sophistiqué (Maximal Dependence Decomposition) permet de prendre en compte les dépendances les plus significatives pouvant exister entre différentes positions non adjacentes du signal.

---

[30] L'alignement de deux séquences permet de les comparer en autorisant certaines dissimilarités entre elles. Par exemple, un alignement (global) des deux séquences $X^1 =$ `aattc` et $X^2 =$ `gtc` est

$$\begin{array}{lcl} X^1 & = & \texttt{aattc} \\ X^2 & = & \texttt{g--tc} \end{array}$$

où le symbole - est ce que l'on appelle un « gap » désignant une insertion d'une lettre de X^1. On pourrait représenter un alignement par une succession des trois états mutation M (une identité parfaite, ou match, entre deux bases est une mutation particulière), insertion I (d'une lettre de X^1), insertion D (d'une lettre de X^2), ici $MIIMM$. On peut bien sûr aligner un ensemble de plus de deux séquences, et obtenir ainsi un alignement multiple d'un ensemble de séquences dont la longueur peut être variable. Un alignement multiple, constitué de $L = 8$ colonnes, des 3 séquences `acgctg`, `catgt`, `acagct` est

```
ac--gctg
-catg-t-
aca-gct-
```

La recherche de tels signaux, ou plus généralement de motifs flous (à trou, de longueur variable) peut bien sûr s'effectuer par HMM, plus précisément en utilisant des profils HMM (voir par exemple Durbin *et al.*, 1998). On peut faire correspondre au profil précédent, c'est-à-dire au modèle de distribution des bases de l'alignement multiple (que nous supposerons dans un premier temps sans « gap », ce qui implique que toutes les séquences aient la même longueur L), un HMM à L états de mutations M_i. Les seules transitions autorisées sont celles de M_i à M_{i+1} et chaque base de l'état M_i est générée selon la loi $\pi_o(M_i, a) = p_i(a)$. On peut fabriquer un modèle d'alignement plus raffiné en prenant en compte des insertions (il existe des segments de séquences qui ne correspondent pas au modèle d'alignement) et des délétions (certains segments de l'alignement multiple ne correspondent pas aux bases d'une séquence). On rajoute un état I_i modélisant une insertion après une base correspondant à la $i^{ème}$ colonne de l'alignement multiple. L'état I_i peut être atteint de l'état M_i. On peut alors rester en I_i s'il y a plusieurs insertions successives, ou bien passer directement à M_{i+1}. Dans chaque état d'insertion, on peut supposer que les bases sont générées selon la loi $\pi_o(I_i, a) = p(a)$. Une délétion dans une séquence par rapport à la $i^{\text{ème}}$ colonne de l'alignement est modélisée [31] en rajoutant un état supplémentaire D_i entre les états M_{i-1} et M_{i+1}. Pour prendre en compte plusieurs délétions successives, on crée une transition entre D_i et D_{i+1} (on prend ainsi en compte la position de la délétion par rapport à l'alignement). Les L états D_i ne peuvent bien entendu générer aucune base, et sont donc appelés états silencieux. L'architecture d'un profil HMM pour un alignement avec insertion/délétion est représentée dans la figure 8. Les techniques précédentes d'identification de HMM peuvent bien sûr s'adapter aux profils HMM qui peuvent ainsi s'utiliser dans différents contextes. Si le profil HMM est connu, l'algorithme de Viterbi permet d'aligner une séquence cible contre le profil. Le profil peut bien sûr être estimé par l'algorithme EM. Enfin, on peut effectuer un alignement multiple par profil HMM en alignant chaque séquence séparément au profil [32] pour ensuite en déduire l'alignement multiple.

4.3. Conclusion

Nous n'avons pas ici donné un panorama complet des applications des chaînes de Markov cachées dans l'étude des séquences biologiques ; en particulier nous n'avons pas évoqué leur emploi dans l'alignement de séquences, technique extrêmement employée pour rechercher dans les banques de données des séquences homologues à une séquence fixée. Nous n'avons pas non plus montré leur emploi dans le domaine — voisin de celui des alignements — de la reconstruction d'arbres phylogénétiques.

Néanmoins les applications que nous avons présentées montrent leur puissance en génomique pour tout ce qui concerne la recherche de structures cachées dans une séquence, en particulier pour l'annotation. Le long des génomes se succèdent des gènes séparés de régions intergéniques, ceux ci sont « écrits » sur un brin ou sur l'autre, chez les eucaryotes ils sont eux mêmes une alternance d'exons et d'introns ; de multiples signaux ponctuent la séquence — et l'on est

[31] On pourrait aussi autoriser des transitions entre M_i et M_{i+2}, entre M_i et M_{i+3}... pour modéliser les délétions, mais le nombre de transitions serait beaucoup trop important.
[32] Si le profil est inconnu, on l'estime par *EM*.

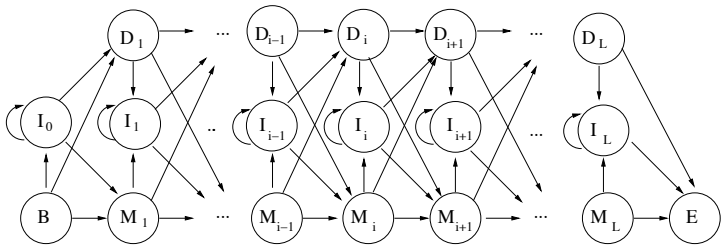


FIG. 8. *Profil HMM pour un alignement avec insertion/délétion. Les états B et E modélisent respectivement le début et la fin de la séquence*

loin de les connaître tous. Les chaînes de Markov cachées sont un outil flexible et adaptatif qui, couplés à d'autres outils tels que la comparaison de séquences, l'exploitation de données d'expression (recueillies entre autres par les « puces » à ARN ou à ADN) permettent d'élaborer des stratégies performantes. Le développement de cet outil (estimation automatique du nombre d'états cachés...) et de ses variantes (semi-markov, ordre variables...) rendra possible la prise en compte des niveaux d'information de plus en plus complexes de la problématique biologique.

L'une des raisons de l'efficacité des méthodes fondées sur les chaînes de Markov cachées réside dans la nature même des données à analyser : on observe une suite de nucléotides dont on sait qu'ils relèvent de phénomènes biologiques divers — intergénique, codant, introns, exons... — et ces états, eux, ne sont pas observables directement. De même que la notion de probabilité est inhérente à la génétique dès que l'on adopte un point de vue « mendélien » sur la transmission aléatoire des allèles, la notion de structures cachées est — sans doute encore pour de nombreuses années — inhérente aux données de séquençage. Aujourd'hui, chaque mois voit arriver dans les banques de données plus de 600 millions de nouvelles lettres (nucléotides) au sein desquelles sont « cachés » quelque 10000 gènes. Extraire la teneur scientifique de cette masse d'information ne peut à l'évidence se faire que si l'on dispose d'un outil informatique performant et capable de s'adapter aux nouvelles questions — outil nécessairement fondé sur une analyse mathématique sérieuse.

5. Références

• Audic, S. and Claverie, J.M. (1998). Self-identification of protein-coding regions in microbial genomes. *Proc. Natl. Acad. Sci. USA*, **95** (17) : 10026-10031.

• Bakry, D., Milhaud, X. and Vandekerkhove, P. (1997). Statistique de chaînes de Markov cachées à espace d'états fini. Le cas non stationnaire. (Statistics of hidden Markov chains, finite state space, nonstationary case). *C. R. Acad. Sci., Paris, Ser. I, Math.* **325, No.2**, 203–206.

• Baldi, P. and Brunak, S. (1998). BIOINFORMATICS. *The machine learning approach*. MIT press, Cambridge.

• Baum, L.E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of finite State Markov Chains. *The Annals of Mathematical Statistics*, **37**, 1554–1563.

• Baum, L., Petrie, T., Soules, G. et Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, **41-1**, 164–171.

• Bize, L., Muri, F., Samson, F., Rodolphe, F., Ehrlich, D., Prum, B. and Bessières, P. (1999). Searching gene transfers on Bacillus subtilis using Hidden Markov models. In : *Recomb'99 Proceedings of the Third Annual International Conference on Computational Molecular Biology*, 43–49.

• Braun, J.V. and Müller, H.G. (1998).Statistical Methods for DNA Sequence Segmentation. Statistical Science, **13**, 2, 142–162.

• Bréhélin, L. and Gascuel, O. (2000). Modèles de Markov cachées et apprentissage de séquences. *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*, Editions Cépaduès, H. Prade, R. Jeansoulin et C. Garbay éditeurs, P. 407–421.

• Boys, R.J., Henderson, D.A. and Wilkinson, D.J. (2000). Detecting homogeneous segments in DNA sequences by using hidden Markov models. *J. R. Stat. Soc., Ser. C, Appl. Stat.*, **49**, No.2, 269–285.

• Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*. Vol.268, pp. 78–94.

• Churchill, G.A. (1989). Stochastic Models for Heterogeneous DNA Sequences. *Bulletin of Mathematical Biology*, **51**, 1, 79–94.

• Churchill, G.A. (1992). Hidden Markov chains and the analysis of genome structure. *Computers chem.*, **16**, 2, 107–115.

• Cowan, R. (1991). Expected frequencies in DNA patterns using Whittle's formula. *J. Appl. Prob.*, **28**, 886–892.

• Delcher, A. L., Harmon, D., Kasif, S., White, O. and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, **27**, 23, 4636–4641.

• Dempster, A.P., Laird, N.M. et Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via EM Algorithm. *J.R.S.S. B.*, **39**, 1–38.

• Douc, R. (2001). *Problèmes statistiques pour des modèles à variables latentes : propriétés asymptotiques de l'estimateur du maximum de vraisemblance*. Thèse, École polytechnique.

• Durbin, R., Eddy, S.R., Krogh, A., and Mitchison, G. (1998). *Biological Sequence Analysis : Probabilistic models of proteins and nucleic acids*. Cambridge University Press, Cambridge.

• Godin, C., Guédon, Y., Costes, E. and Caraglio, Y. (1997). Measuring and anlysing plants with the AMPAmod software. In : Plants to Ecosystems - *Advances in Computational Life Sciences* (Michalewicz, M.T., ed), Vol I, 53–84. CSIRO Publishing.

• Godin, C., Guédon, Y. and Costes, E. (1999). Exploration of a plant architecture database with the AMPAmod software illustrated on an apple tree hybrid family. *Agronomie*, **19**, 163–184.

• Guédon, Y. and Cocozza-Thivent, C. (1990). Explicit state occupancy modelling by hidden semi-Markov models : application of Derin's scheme. *Computer Speech and Language*, **4**, 167–192.
• Guédon, Y. (1998). Hidden semi-Markov chains : A new tool for analyzing nonstationary discrete sequences. In : *Proceedings of the 2nd International Symposium on Semi-Markov Models : Theory and Applications* (Janssen, J. and Limnios, N., eds), Compiègne, France.
• Guédon, Y. (1998). Analyzing nonstationary discrete sequences using hidden semi Markov chains. *Tech. report*.
• Guédon, Y. (1999). Computational methods for dicrete hidden semi-Markov chains. *Appl. Stochastic Models in Business and Industry*, **15**, 195–224.
• Hayes, W. S. and Borodovsky, M.(1998). How to interpret an anonymous bacterial genome : machine learning approach to gene identification. *Genome Research*, **8**, 1154–1171.
• Henderson, J., Salzberg, S. and Fasman, K. (1997). Finding genes in DNA with a Hidden Markov Model. *Journal of Computational Biology*, **4**, 2, 127–141.
• Kulp, D., Haussler, D., Reese, M.G. and Eeckman, F.H. (1996). A generalized hidden Markov model for the recognition of human genes in DNA. In : Proc. Fourth Internet. Conf. Intelligent Systems for Molecular Biology, AAAI Pres, Menlo Park, 134–141.
• Le Gland,F. and Mevel, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Syst.*, **13, No.1**, 63–93.
• Lukashin, A.V. and Borodovsky, M. (1998). GeneMark.hmm : new solutions for gene finding. *Nucleic Acids Research*, Vol.26, No.4, pp. 1107–1115.
• Mercier S. and Daudin J.J. (2000) Exact Distribution for the local score of one i.i.d. random sequence. To appear in *J. Comp. Biol.*
• Mevel, F. (1997). *Statistique asymptotique pour les modèles de Markov cachés*. Thèse, Université Rennes 1.
• Muri, F. (1997). *Comparaison d'algorithmes d'identification de chaînes de Markov cachées et application à la détection de régions homogènes dans les séquences d'ADN*. Thèse, Université René Descartes, Paris V.
• Muri, F. (1998). Modelling Bacterial Genomes Using Hidden Markov Models. In *Compstat'98 Proceedings in Computational Statistics*, R. Payne (Ed.). Physica-Verlag, Berlin.
• Muri, F., Chauveau, D., and Cellier, D. (1998) : "Convergence Assessment in Latent Variable Models : DNA Applications". In Lecture Notes in Statistics *Discretization and MCMC Convergence Assessment* (ed C.P. Robert), Chapter 6, pp. 127–146. Springer-Verlag.
• Nicodème P. (2000) The symbolic package Regexpcount. Presented at GCB00
• Nicodème P. (2001) Fast Approximate Motif Statistics. To appear in *J. Comp. Biol.*
• Nicodème P. , Salvy B., Flajolet Ph. (2001) Motif Statistics. To appear in *Theoretical Computer Science* - Presented at ESA99 (1999)
• Nicolas, P. (2000). Segmentation de génomes bactériens par chaînes de Markov cachées : application à la recherche de transferts de gènes sur les chromosomes de *Bacillus subtilis* et *Escherichia Coli*. Mémoire du DEA, Biomathématiques, Universités Paris VI, Paris VII.
• Nicolas, P. (2001). Approche pour la détection des gènes dans les séquences bactériennes de mauvaise qualité par chaînes de Markov cachées. Communication au colloque IMPG, Détection de gènes, 6 mars 2001.
• Nuel G. (2001) Grandes déviations et chaînes de Markov pour l'étude des mots exceptionnels dans les séquences biologiques. Thèse Université d'Evry

• Propp, J. G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structure and Algorithms*, **9**, 223–252.
• Prum, B., Rodolphe, F. et de Turckheim, É. (1995). Finding words with unexpected frequencies in DNA sequences. *J. Royal Statist. Soc. Ser. B*, **57**, 205–220.
• Qian, W. & Titterington, D.M. (1990). Parameter estimation for hidden Gibbs chains. *Statist. Probab. Lett.*, **10**, 49–58.
• Qian, W. and Titterington, D.M. (1991). Estimation of parameters in hidden Markov models. *Philos. Trans. Roy. Soc. London A*, **337**, 407–428.
• Rabiner, L.A (1989). Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEE*, **77**, 257–286.
• Redner, R.A. & Walker, H.F. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review*, **26**, 2, 195–239.
• Reinert, G., Schbath, S. and Waterman, M.S. (2000) Probabilistic and Statistical Properties of Words : An Overview. *J. Comp. Biol.,* 7 1–46.
• Robert, C.P., Celeux, G. & Diebolt, J. (1993). Bayesian estimation of hidden Markov chains : A stochastic implementation. *Statist. Probab. Lett.*, **16**, 77–83.
• Robert, C.P. (1996). *Méthodes de Monte Carlo par Chaînes de Markov*. Economica.
• Robin, S. and Daudin J.J. (1999) Exact distribution of word occurrences in a random sequence of letters. *J. Appl.Prob* 36, p. 179–193.
• Robin, S. and Daudin, J.J. (2000) Exact distribution of the distances between any occurrences of a set of words. To appear in *Ann. Inst. Statist. Math.*
• Salzberg, S., Delcher, A., Kasif, S. and White, O. (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, **26**, 2, 544–548.
• Schbath, S., Prum, B. et de Turckheim, É (1995). Exceptional motifs in different Markov chain models for a statistical analysis of DNA sequences. *J. Comp. Biol.*, **2**, 417–437.
• Schbath, S. (1995). Compound Poisson approximation of word counts in DNA sequences. *ESAIM : Probability and Statistics*, **1**, 1-16.
Accessible à l'adresse URL http://www.emath.fr/ps/
• Schbath, S. (2000) An overview on the distribution of word counts in Markov chains. *J. Comp. Biol.* 7 193–202.
• Schiex, T., Moisan, A. Duret, L. and Rouzé, P. (2000). Recherche des gènes et des erreurs de séquençage dans les génomes bactériens GC-riches. *Proc. of JOBIM'2000 (Journées Ouvertes Biologie Informatique Mathématiques), Montpellier, 2000.*
• Vandekerkhove, P. (1998). Recuit simulé avec un estimateur séquentiel de l'énergie. (Simulated annealing with a sequential estimator of the energy). *C. R. Acad. Sci., Paris, Ser. I, Math.* **326, No.8**, 1003–1006.
• Vert, J.P. (2001). *Méthodes statistiques pour la modélisation du langage naturel.* Thèse, Paris VI.
• Whittle P. (1955). Some distribution and moment formulae for Markov chains *J. R. Statist. Soc. B*, **17**, 235–242.
• Xu, Y., Einstein, J.R., Mural, R.J., Shah, M. and Uberbacher, E.C. (1994). Constructing gene models from accurately predicted exons : an application of dynamic programming. *Comp. Appl. Biosci.*, **10**, pp. 613–623.

INFORMATIONS

Section 01 du Comité National Compte rendu de la session de printemps 2001 et du concours

La section « Mathématiques et Outils de modélisation » du Comité National a tenu sa session de printemps du 12 au 15 mars 2001. La direction scientifique du CNRS était représentée par Michel Enock Levi, chargé de mission pour les mathématiques. Outre les propositions de promotion des chercheurs et des propositions de délégation-détachement, les points suivants ont fait l'objet de la session : venue de la nouvelle directrice du département SPM (Sciences physiques et mathématiques), rapport avec le département STIC, et documentation.

Une nouvelle directrice au département SPM

Le 1er mars, Élisabeth Giacobino a succédé à Jean-Paul Pouget à la tête du département SPM. Elle a réaffirmé l'importance des mathématiques dans le département. Elle a déclaré souhaiter continuer la politique d'ouverture de Jean-Paul Pouget et en particulier soutenir les liens entre mathématiques et physique.

Création du département STIC (suite)

Francis Jutand, directeur du nouveau département Sciences et Technologies de l'Information et de la Communication, a déclaré avoir reçu le signal de la communauté mathématique quant à son unité, et considère que le département SPM est le cadre principal pour les mathématiques au CNRS. L'analyse de la participation STIC dans les laboratoires de mathématiques est lente, il n'y aura pas de dotation récurrente cette année. Les actions concertées STIC/SPM ont par nature un fonctionnement complexe et feront l'objet d'une étude préalable approfondie, elles viseront en particulier les systèmes d'information, de communication et le traitement de l'information. Il a esquissé le profil de chercheur interdisciplinaire que souhaite recruter ce département, il s'agit de jeunes ayant une compétence reconnue dans un domaine (doctorat de maths par exemple) et une expérience dans les sciences de l'information (post-doc, entreprise). Jean-Luc Sauvageot s'est réjoui des bons rapports existant entre la section et le nouveau département STIC. Il a rappelé l'attente très forte dans la communauté suscitée par la création du département STIC et souhaite que le cloisement STIC/SPM soit invisible, c'est-à-dire qu'il y ait une unité de politique scientifique au CNRS. Une proposition raisonnable en ce sens serait la création d'un poste de chargé de mission pour les mathématiques commun au deux départements. Enfin, la section demande des règles claires sur les concours 01/07

(maths-informatique) notamment pour les concours fléchés et espère que des mathématiciens seront recrutés sur le concours 07, concours doté de 50 postes de chargés de recherche cette année.

Documentation

Comme annoncé lors de la session de printemps, nous avons consacré une après-midi à une réunion sur la documentation. Les participants étaient : P. Bérard, J.-M. Deshouillers, J. Eisenstaedt, M. Enock, B. Helffer, J.-Y. Mérindol, H. Nocton, C. Peskine, C. Sabbah, G. Sureau, B. Teissier, L. Zweig. Nous n'abordons pas dans ce rapport des questions importantes telles que l'archivage, l'accès aux bases de données et nous nous concentrons sur l'augmentation globale du prix de la documentation et les différentes analyses et réponses qu'elle a suscitées.

Tout d'abord, l'état des lieux est alarmant. Le prix des revues est globalement en augmentation de 15% par an depuis quatre ans en tenant compte des fluctuations monétaires ; l'augmentation du prix à la page est presque du même ordre ; de nombreuses bibliothèques ne peuvent faire face et résilient des abonnements. Cette moyenne couvre des réalités très disparates. Pour l'analyse de ces augmentations, il convient de préciser si la revue est académique ou non, c'est-à-dire si son titre est propriété d'une université (ou d'une société savante ou encore d'une association). Les revues académiques sont nettement moins chères que les revues commerciales, et leur augmentation parfois importante (en pourcentage) n'affecte pas trop les budgets. La différence de prix entre revues commerciales et académiques s'explique en partie par des coûts pris en charge par l'institution et peu répercutés, mais surtout par des objectifs de rentabilité. Dans le même laps de temps, les livres ont peu augmenté.

Dans le même temps se développe l'information directe par les serveurs de prépublications, qui est un pis-aller pour faire face aux augmentations. Pour l'instant, le principal serveur mondial de prépublications est le serveur xxx de Los Alamos. Malgré ses miroirs en Europe, ce serveur a une gestion centralisée. En Allemagne, l'institut Max Planck dispose d'un serveur commun pour tous ses laboratoires, qui pourrait constituer les prémices d'un serveur européen de prépublications.

Une autre solution pour réagir passe par la constitution de consortia. Un consortium est une association dont le but est de négocier avec les éditeurs afin d'offrir à ses membres de meilleures conditions financières, sous plusieurs modalités : version papier seule, accès à la toile seul, papier + toile, ou une pondération de tout cela pour chacun des membres. On a deux philosophies en présence :

- Le consortium généraliste COUPERIN. Il regroupe 71 membres (bibliothèques universitaires (BU) et bibliothèques associées) et négocie l'accès à toutes les revues d'un éditeur, toutes spécialités confondues.
- Le consortium thématique du RNBM (Réseau National des Bibliothèques de Mathématiques lié à la cellule Mathdoc). Il représente les bibliothèques de mathématiques pour négocier avec certains éditeurs. Chaque bibliothèque peut choisir les revues qu'elle souhaite.

Rien n'empêche de profiter des conditions de plusieurs consortia, mais on doit prendre garde que les services ne soient pas facturés deux fois. On doit

aussi prendre garde au fait que même s'il est tentant d'acheter beaucoup de revues avec les consortia, on aboutit souvent à une augmentation globale du coût de la documentation.

Dans les deux cas, les contrats existent pour le court terme (entre un et quatre ans), et on ignore absolument en quels termes les prochains contrats pourraient être négociés, cela dépend en particulier de la politique des grands éditeurs, et de l'attitude des universités américaines, dont certaines sont réticentes à ces achats de tout un catalogue. Pour l'instant, l'achat électronique est conditionné aux achats papiers, mais on s'attend à ce qu'il devienne un marché propre. Une inquiétude à laquelle n'ont pas répondu les éditeurs pour les versions électroniques des revues est l'archivage ; à l'heure actuelle, une bibliothèque qui achète une revue électronique pour les quatre années qui viennent n'a aucun engagement sur le futur, c'est-à-dire à la possibilité d'accéder à ces données dans cinq années.

Ces considérations financières ne doivent pas faire oublier que la question principale est la qualité scientifique des revues. De ce point de vue, l'achat de toutes les revues d'un éditeur empêche les bibliothèques concernées de se désabonner parce qu'une revue n'est plus intéressante. De plus, les petites revues, dont le rôle est prépondérant, auront du mal à suivre le mouvement de l'électronisation, si elles restent isolées. En effet, tout laisse à penser qu'à l'heure des fusions des grands éditeurs, les petites revues auront du mal à maintenir leur visibilité, notamment sur la toile.

En conclusion, la question de la documentation dépasse dans une certaine mesure le cadre des mathématiques et est un problème posé à toutes les universités. Le futur à court terme est peu visible, notamment sur la place du marché électronique des revues. Par rapport aux autres disciplines, les mathématiques ont deux atouts, en premier lieu la forte structuration de sa communauté et en second lieu le nombre et la place des revues académiques qui constituent les revues les plus importantes quant à leur impact. Parce que l'électronique va devenir un marché propre, les termes dans lesquels seront signés les prochains contrats sont fondamentaux pour l'avenir de la documentation.

Concours

Le concours s'est déroulé du 24 au 27 avril avec le nombre de postes indiqué à la session d'automne, c'est-à-dire 17 postes de chargés de recherches et 6 de directeurs de recherches. À ces concours se sont présentés respectivement 202 et 64 candidats ; pour les chargés de recherches, il s'agit d'une augmentation de 60% par rapport à l'année précédente s'expliquant partiellement par les deux postes ouverts sur les STIC.

La section a constaté le très haut niveau de ces deux concours et le fait que les écoles doctorales forment un réservoir important de jeunes chercheurs de talent.

La section 01 du Comité National du CNRS tient à rendre hommage à la mémoire de Philippe Benilan, professeur à l'Université de Franche-Comté à Besançon, décédé le 17 février 2001. Philippe Benilan a apporté une contribution remarquable à la théorie des semi-groupes et au développement de l'école française des équations d'évolution non linéaires.

Du bon usage des Comptes rendus de l'Académie des Sciences

Philippe G. Ciarlet et Bernard Malgrange

À quoi servent aujourd'hui les Comptes rendus ? Quelle est la politique éditoriale de cette publication ? Comment fonctionne le système d'acceptation ou de refus des projets de Note ? Ces questions nous sont fréquemment posées, par les auteurs potentiels, par les experts consultés pour évaluer les projets de Note, par les collègues siègeant dans les Commissions de Spécialistes, au CNU ou au Comité National. Les quelques lignes qui suivent ont pour seul but d'indiquer rapidement comment nous voyons les choses et de décrire les principes qui guident notre politique éditoriale.

Notre priorité est la qualité : à cet égard, nous ne faisons que nous en tenir à la « définition » qui figure en bonne place dans chaque numéro : « Une Note aux Comptes rendus est la première relation brève d'une découverte importante ou d'un résultat significatif. »

Deux objections, en fait liées, sur cette définition sont souvent formulées :
– À quoi sert une publication « brève » d'un résultat qui, de toute façon, sera publié en détail ultérieurement ?
– À quoi sert une telle « annonce », puisque, ou bien l'article est prêt et rien n'empêche de le diffuser immédiatement sur la Toile avant sa publication dans un journal, ou bien l'article n'est « pas vraiment » prêt, auquel cas il est prématuré d'en annoncer les résultats ?
La première objection n'en est pas vraiment une. Nous pensons simplement que nous n'avons pas à nous substituer à tel ou tel auteur pour juger de la pertinence d'annoncer des résultats qui, bien souvent, ne seront pas publiés avant deux ou trois ans, délai courant dans les principales revues. De toute façon, le souci d'informer les autres spécialistes ou le souci de priorité nous paraissent des motivations éminemment respectables ! Si certains auteurs ne s'embarrassent pas de tels soucis, libre à eux !
La deuxième objection mérite que l'on s'y arrête davantage. Nous tenons tout d'abord à rappeler très clairement que, contrairement au passé, les Comptes rendus constituent aujourd'hui une revue avec rapporteurs : pour qu'une Note soit publiée, il est nécessaire qu'elle ait été expertisée et qu'elle ait fait l'objet d'au moins un rapport favorable. Ce point fondamental est dorénavant précisé en toutes lettres dans chaque numéro des Comptes rendus !

Qui sont donc les rapporteurs ? Comment sont-ils désignés ? Un rapporteur peut-être le « Présentateur » lui-même, un expert sollicité directement par le Présentateur, ou un expert désigné par le « Comité de Lecture » placé sous notre responsabilité. Cette troisième possibilité peut éventuellement compléter la première ou la seconde. Le Comité de Lecture est constitué par un petit groupe de Confrères, dont les compétences recouvrent en principe toutes les « rubriques » publiées.
Il résulte de cette procédure qu'une Note publiée, ayant fait l'objet d'une expertise, doit figurer dans les dossiers de candidature au même titre que les

publications dans les revues « avec rapporteurs » traditionnelles. Ce n'est évidemment pas le cas des « publications » faites sans aucun contrôle sur la Toile !

Bien sûr, le sérieux de l'expertise entraîne des délais inévitables que nous nous efforçons de limiter autant que faire se peut, grâce à la remarquable efficacité du Service des Comptes rendus de l'Académie placé sous la bienveillante houlette de Mireille Monteil, et grâce à l'Éditeur, Elsevier-France, qui s'est en effet engagé à respecter un délai de six semaines entre la remise du « bon à tirer » d'un numéro et sa sortie effective. La majorité des rapporteurs jouent le jeu, permettant ainsi de concilier rapidité de publication et sérieux de l'expertise. Il faut bien comprendre que les délais trop longs que l'on observe parfois, et qui exaspèrent à juste titre les auteurs, sont dus presque toujours à des rapporteurs négligents qu'un nombre impressionnant de relances du secrétariat des Comptes rendus laissent apparemment de marbre ! Dans quelques rares cas, les délais sont aussi dus à des auteurs qui, inexplicablement, tardent à corriger leurs épreuves !

Sur le principe d'une publication où les preuves sont seulement esquissées, voire absentes, nous remarquerons simplement qu'un résumé bien fait nous paraît constituer dans bien des cas une information aussi utile qu'un article complet, que ce soit pour le rapporteur d'un dossier de candidature ou pour le mathématicien qui désire se tenir au courant d'un sujet. De ce point de vue, les résumés de thèses sont extrêmement utiles, et nous encourageons vivement leur publication dans les Comptes rendus.
Mais des mathématiciens chevronnés seraient également bien avisés de donner dans une même Note une présentation synthétique d'un travail en cours, plutôt que de rédiger une série de Notes, souvent très « ponctuelles » et parfois très arides, sur un même sujet, série que, par principe, nous sommes rapidement enclins à interrompre.

« Last but not least », nous nous efforçons depuis plusieurs années d'encourager la publication de Notes par des collègues étrangers. Nos efforts commencent à porter leurs fruits, puisque le tiers environ des Notes actuellement publiées provient de l'étranger. Mais c'est une proportion encore trop faible, et nous demandons à tous nos collègues de faire savoir, lors de leurs déplacements, que les Comptes rendus sont une revue internationale !

Il va de soi que tous les commentaires sur ces quelques remarques seront les bienvenus.

Interview de Jean-Yves Mérindol

François Digne

La Gazette a souhaité t'interviewer en tant que président et mathématicien.

Je dois avant tout préciser la situation de l'Université Louis Pasteur. Les disciplines représentées sont toutes celles des sciences exactes et naturelles, celles de la santé, les sciences économiques et de gestion, la psychologie, la géographie et les sciences de l'éducation. Nous sommes donc à dominante sciences et santé, mais avec un secteur de sciences humaines et sociales très présent. Bien que de taille moyenne par le nombre d'étudiants, actuellement un peu plus de 16 800, c'est un établissement qui a une activité de recherche très importante, avec plus de 80 unités de recherche, pour la plupart associées à des organismes nationaux (INSERM, INRA, CNRS). Nous avons d'autres particularités comme l'existence au sein de l'université, d'un planétarium, de plusieurs grands musées scientifiques, de beaucoup de contrats de recherche, d'activités de valorisation particulièrement développées. Tout ceci entraîne l'emploi de plusieurs centaines de personnes sur ressources propres qui effectuent ces activités. Les questions de personnel sont particulièrement importantes. Le budget de l'université dépasse 750 MF/an et, ce qui permet de mieux comparer avec d'autres structures, est autour de 1 300 MF/an si on tient compte des salaires des personnels fonctionnaires. C'est l'un des plus gros budgets d'université en France, comparable à ceux de Paris Sud ou de Paris VI. Enfin nous délivrons autour de 300 doctorats par an.

La diversité des questions à aborder, l'ampleur de l'établissement nous conduisent à une organisation institutionnelle qui s'écarte un peu de la moyenne des universités en France. Il y a par exemple de nombreux vice-présidents qui sont élus sur ma proposition, ce qui me permet de travailler avec une équipe, en laissant une très forte délégation aux vice-présidents. Il y a, en plus du vice-président en charge de la recherche et de celui en charge des formations, un vice-président chargé des relations avec les entreprises, un vice-président chargé des locaux et des moyens, un vice-président chargé des relations internationales et un vice-président en charge de la vie étudiante (qui est un étudiant). Nous nous réunissons chaque semaine pendant toute une matinée afin de passer en revue toutes les décisions à prendre.

Je suis avant tout un chef d'orchestre, favorisant la définition d'une stratégie globale de l'université.

Quel est ton avis sur la situation de l'enseignement universitaire ?

L'effort de développement la formation professionnelle (IUP, DESS, autres formations) me semble maintenant un point acquis. La formation continue, en particulier « diplomante », commence à se développer. Les mathématiciens sont un peu en retard sur ces deux sujets, en particulier sur le second. Même si je comprends bien que certaines sous-spécialités de recherche ne s'y prêtent pas directement, c'est quand même un point qui devrait les mobiliser.

Pour la formation initiale (DEUG sciences), chacun connaît les baisses d'effectif, assez considérables en physique et en chimie. L'organisation des DEUG sciences est à revoir, afin d'aller vers une simplification de la première année et une présentation plus vivante des diverses matières. C'est en cours dans plusieurs établissements, mais peut-être pas toujours avec l'ambition souhaitable.

D'autre part il faut absolument se préoccuper de la mise en œuvre des technologies de l'information et de la communication dans l'enseignement. L'introduction de ces technologies va aboutir vraisemblablement à une nouvelle forme de la compétition entre universités, ce qui est assez nouveau pour l'enseignement. L'enseignement de premier et second cycle sera bientôt jugé par les étudiants et leurs familles en tenant compte de l'existence et de la qualité des enseignements conçus grâce aux TIC dans les cursus proposés. Cela fournira des éléments de comparaison précis de l'enseignement entre établissements au niveau national et international alors que jusque-là la comparaison portait principalement sur la qualité de la recherche, ce qui la limitait aux initiés et aux troisième et second cycles. On peut se retrouver avec toutes les combinatoires possibles pour le rapport enseignement-recherche, et en particulier avec des universités très bonnes en recherche mais où les enseignants se sont peu investis dans les nouvelles technologies et dans la formation à distance, à côté d'universités peu visibles en recherche mais s'impliquant énormément dans ces nouvelles formes d'enseignement. Il est nécessaire, partout, d'avoir une stratégie, des moyens et d'arriver à former ou à attirer des talents sur ce sujet. Il y a évidemment des réticences et certains enseignants pensent qu'il s'agit d'un simple effet de mode. Je ne le crois pas. Certaines disciplines sont d'ailleurs déjà bien avancées, en particulier la médecine, les sciences de la vie et plus généralement les disciplines qui ont besoin traditionnellement dans leur enseignement de support « image », films ou photographies. Les mathématiques ne me semblent pas souvent en première ligne, ce qui va à terme leur poser quelques problèmes.

On a déjà connu une situation voisine avec l'informatique, vue comme discipline de service. Les mathématiciens ne l'ont pas vraiment intégré à leur enseignement alors que d'autres disciplines scientifiques, plus expérimentales, l'ont fait avec beaucoup d'ampleur.

Le dernier facteur qui va faire évoluer l'enseignement est l'ouverture internationale. Il est indispensable que nos étudiants prennent l'habitude d'aller passer au moins un semestre de leurs études à l'étranger. C'est déjà réalisé pour les études de commerce, et en voie de réalisation pour les études d'ingénieur ou les autres enseignements professionnalisés. On ne peut pas se satisfaire d'une organisation à deux vitesses, avec des formations qui resteraient franco-françaises. Il en va de l'attractivité de ces formations, et aussi de l'influence que les jeunes diplômés auront dans leur vie professionnelle et sociale. Les chercheurs ont une pratique professionnelle internationalisée depuis longtemps. Mais les étudiants en sciences restent trop formés dans un seul pays, ce qui ne peut pas leur permettre de jouer un rôle suffisant dans un monde ou la mobilité internationale devient indispensable, au moins à un certain niveau de responsabilité. Je considère comme un défi de la première importance que d'arriver à éviter que les responsables internationaux, capables de comprendre plusieurs cultures, ne

soient pas seulement formés dans des écoles de finance, de commerce ou de sciences politiques.

Comme président d'université et mathématicien, comment vois-tu les différences entre les cultures propres à chaque discipline ?

Les communautés scientifiques sont très différentes et communiquent peu entre elles. Chacune pense que son mode d'organisation est celui de toute la science : cette myopie existe même à des niveaux de haute responsabilité. L'organisation des mathématiciens est très différente de celles des autres communautés, ce qui est souvent justifié. Mais ceci peut parfois se révéler un handicap et il faut alors en avoir conscience pour éviter que ce handicap ne se transforme en véritable défaut. Ce risque est amplifié chez les mathématiciens qui sont souvent plus isolés que les autres scientifiques. Les physiciens développent de plus en plus de liens avec la biologie, les biologistes avec l'informatique et on pourrait multiplier les exemples. Et ces liens ne sont pas de simples prestations de service ; chaque discipline y apporte ses compétences et sa culture, et retire de ce contact des nouveaux problèmes, souvent profonds, et de nouvelles méthodes. Dans un gros laboratoire de chimie, il doit y avoir des compétences en physique, en informatique, en sciences pour l'ingénieur et en biologie puisque le travail sur la matière, qui ne se ramène pas aux divisions habituelles entre disciplines, amène naturellement à travailler en recherchant d'autres compétences scientifiques. En math il y a bien sûr des chercheurs qui s'intéressent ou collaborent avec d'autres disciplines mais ce n'est pas une obligation de la discipline elle-même. Un chercheur peut faire une carrière extrêmement brillante sans jamais avoir été obligé de discuter avec un physicien, un chimiste ou un économiste. Les contacts les plus fréquents, par exemple avec les physiciens théoriciens, sont bien entendu fructueux, avec parfois des réussites spectaculaires, mais ne conduisent pas à une même ouverture que les exemples cités plus haut. Et l'enseignement en formation initiale ne corrige pas significativement ces isolements. La communauté mathématique est plus refermée sur elle-même que les autres communautés scientifiques.

Cela n'a pas de conséquences graves immédiates, à cause de la vitalité de la recherche mathématique dans la plupart de ses spécialités (mais ceci n'est jamais éternel), à cause du bon niveau scientifique des mathématiques en France, et — plus prosaïquement — parce que les besoins financiers des laboratoires de mathématiques ne sont pas à la même échelle que ceux des disciplines expérimentales. C'est cependant embarrassant en termes sociologiques ou politiques, par exemple dans les débats nationaux où les progrès dûs aux maths passent souvent inaperçus alors que ceux apportés par d'autres disciplines sont bien plus valorisés. Ceci va aussi poser un problème pour les postes et les arbitrages sur les emplois. Le nombre de mathématiciens a beaucoup cru depuis une vingtaine d'années à cause de l'augmentation importante des effectifs étudiants, toutes sciences confondues. Beaucoup d'établissements, universités ou organismes de recherche, ont des priorités scientifiques hors des mathématiques, et les arbitrages nécessaires pour les redéploiements peuvent affecter les mathématiques là où les mathématiciens n'ont pas su s'ouvrir au dialogue avec les autres. Sans faire de pronostic, je ne trouve pas la situation actuelle idéale, même si les maths sont aussi protégées par le bon niveau mondial de l'école française ce qui

est bien connu des autres scientifiques. Mais cette « protection » ne joue pas forcément dans chaque établissement.

Un autre point de différence entre les maths et les autres communautés est que nous ne sommes pas, au moins pour les mathématiques dites pures, dans un univers de « laboratoire » avec des expériences, des machines, des moyens lourds, tout ce qui fait que les directeurs de laboratoires doivent aussi être des managers de la science, cherchant en permanence des financements importants, négociant des contrats ou valorisant des brevets. Les matheux ne sont souvent pas concernés, sauf certaines branches, par un certain nombre de débats de l'université ou des organismes de recherche, ce qui les amène à être discrets dans certaines instances.

Une autre différence porte sur l'avenir des doctorants. Plusieurs facteurs jouent : l'organisation de la recherche en math n'est pas segmentée et répartie en tâches spécialisées, beaucoup de doctorants mathématiciens sont agrégés, les débouché académiques sont encore assez importants et se font sans périodes post-doctorales obligatoires. Tout ceci est plutôt positif, mais entraine un manque de mobilité internationale, et aussi un suivi encore un peu lointain de l'avenir professionnel des docteurs.

Enfin, je souhaite souligner l'existence des IREM. C'est un outil rare, aucune autre discipline n'a rien créé de comparable, qui devrait permettre aux mathématiciens d'être très actifs dans les questions d'enseignement. Cependant l'essentiel de la communauté mathématique les a un peu perdu de vue. Ce serait dommage de perdre ces compétences car il n'y a pas d'autre discipline où la relation entre enseignants du second degré et universitaires soit aussi forte. Ces instituts sont en train de surmonter, avec encore quelques difficultés, l'un de leurs défauts institutionnels, à savoir leur ancienne organisation nationale en prise directe avec certaines directions du ministère. Ceci les avait parfois coupé de leur université et de ses mathématiciens. C'est un outil utile qui doit retrouver sa place dans le monde universitaire.

Comment vois-tu la question de la documentation électronique ?

C'est un important point d'actualité. Les mathématiciens sont devant un paradoxe. Ils ont été à l'avant-garde pour tout ce qui est documentation, abonnements et archivages électroniques. Je pense en particulier au rôle de la cellule math-doc à Grenoble, soutenue par le CNRS et l'Université Joseph Fourier, et à diverses autres initiatives. Les physiciens théoriciens les ont devancés pour diffuser les préprints électroniques mais la communauté mathématique les a rapidement rejoint, d'abord dans certains sous champs disciplinaires puis de façon très large. Nous sommes l'une des rares disciplines qui a besoin d'une diffusion électronique des preprints, à cause de la très longue période entre la sortie des résultats et de la publication dans une revue avec comité de lecture. Mais les autres sciences ont les mêmes besoins que nous pour l'accès en ligne aux revues électroniques des éditeurs. Elles ont même des besoins supérieurs aux nôtres sur certains points (bases de données, recherche bibliographique...).

Il faut savoir que pour les éditeurs, même ceux qui sont très importants pour les mathématiciens, nos revues ne représentent qu'une toute petite partie du volume des abonnements. Les éditeurs souhaitent avoir comme interlocuteurs

les représentants du plus grand nombre de lecteurs. Ceci conduit chaque établissement, université et organisme, à négocier des marchés pour l'ensemble de ses besoins, toutes disciplines confondues. Il faut que les mathématiciens soient présents là où les décisions vont être prises pour faire entendre leurs demandes spécifiques, tout en s'inscrivant dans les décisions qui intéressent tous les scientifiques, et bientôt les étudiants. Il existe maintenant un consortium entre universités, Couperin, qui se charge de négociations collectives. Il rassemble la quasi totalité des universités ayant un secteur scientifique, et j'en assume actuellement la présidence. Nous cherchons à aboutir à un accord avec le CNRS, l'INSERM et les autres organismes de recherche afin d'éviter d'avoir des politiques séparées, contradictoires et coûteuses. Nous avons maintenant un accord de principe des directions de ces organismes pour arriver à une définition conjointe de notre politique et de nos organisations. Reste à mettre ceci en musique et j'espère que nous allons y arriver d'ici peu. Il reste évidemment à savoir ne pas oublier les spécificités de chaque champ scientifique. Notre discipline doit pouvoir accéder à des données qui peuvent être très anciennes. Ceci impose deux choses :

– reprendre sous forme électronique les anciens ouvrages ou des anciens numéros des revues, avec toutes les méthodes d'indexation que l'on peut souhaiter afin de faciliter l'accès par hypertexte à ces documents,

– prêter une attention particulière aux questions d'archivage et à la pérennité de celui-ci.

Il n'y a que dans le secteur des sciences humaines et sociales que la situation soit comparable, mais son retard sur ces questions, au moins en France, donne une responsabilité particulière aux mathématiciens. Des décisions importantes sont en train de se prendre en France et au niveau européen et il est important que les mathématiciens français soient présents, en lien avec les décisions de leurs établissements.

Penses-tu que les mathématiciens sont bien représentés dans les instances dirigeantes des universités ?

Je pense qu'on a encore une image trompeuse des mathématiciens qui auraient peu d'activités collectives, un peu trop influencée par l'image d'Épinal du savant distrait et incompréhensible. Pour ne prendre qu'un exemple, qui n'a pas de valeur statistique, beaucoup d'élèves de Grothendieck ont eu des activités collectives de premier plan dans l'administration de l'enseignement et de la recherche, citons par exemple Demazure, Giraud ou Verdier. Il y a eu plusieurs présidents mathématiciens ayant eu de l'influence bien au délà de leur université : je pense à Kahane (Paris-Sud), Payan (Joseph Fourier-Grenoble), Poitou (ENS-Paris), ou plus récemment Ekeland (Paris Dauphine) et Méla (Paris Villetaneuse). Ce dernier est maintenant chef de la mission scientifique universitaire auprès du ministère.

Actuellement il y a trois présidents mathématiciens, ce qui est peut-être un peu en dessous de notre représentativité arithmétique, mais d'autres disciplines me semblent beaucoup plus absentes à ce niveau de responsabilité. Je rencontre souvent des mathématiciens, soit parce qu'ils sont vice-présidents de leur université, soit parce qu'ils sont chargés de missions particulières sur l'enseignement, la recherche ou d'autres activités.

J'ajoute que j'avais au départ une idée fausse, à savoir que les responsables de gros laboratoires expérimentaux seraient plus naturellement conduit à devenir président d'université. En fait cela ne joue qu'à la marge. Il y a aussi bien par exemple des littéraires purs parmi les présidents, même dans des universités pluridisciplinaires. D'autres facteurs professionnels et personnels entrent en jeu, en particulier les responsabilités collectives assumées à d'autres moments et d'autres endroits. Par ailleurs il y a une sur-représentation des mathématiciens dans les équipes de direction des IUFM. Cela vient probablement de l'investissement qu'ont toujours eu les matheux dans les préparations aux concours de recrutement du secondaire, au développement de la didactique des mathématiques qui est plutôt mieux organisée que la didactique d'autres disciplines, et à la place des IREM.

Que penses-tu des relations avec le CNRS ?

J'ai été responsable de la commission recherche de la conférence des présidents d'université entre décembre 1998 et décembre 2000, ce qui m'a amené à discuter avec les différents organismes de recherche et avec le ministère. D'abord un point important : la recherche en France est loin de se limiter aux universités et au CNRS comme les mathématiciens ont parfois tendance à le croire. En fait ce n'est ni le seul ni le plus riche organisme de recherche. Il y en a d'autres très puissants comme le CNES, le CEA, l'INRIA, l'INSERM, l'INRA, qui développent des compétences importantes, y compris en mathématiques. Il y a eu aussi le CNET mais il a éclaté après la privatisation des télécommunications. Il y avait au CNET une qualité de recherche bien au dessus de celle du CNRS, et ceci impliquait plusieurs mathématiciens.

Le CNRS a développé pendant près de 20 ans, dans l'après-guerre, de grosses structures (laboratoires propres) où les matheux étaient quasiment absents, à l'exception de certains mathématiciens appliqués. Un tournant décisif a été pris dans les années 60 avec la création d'équipes associées ce qui a permis le développement d'équipes universitaires. Ceci a été très utile pour la qualité de la recherche en France. Cette politique a même été determinante pendant 15 à 20 ans. Mais le CNRS n'a pas voulu ou pas pu suivre le développement universitaire qui a repris à partir des années 85, pas plus qu'il n'avait su bien accompagner la création de nouvelles universités au début des années 70, ce qui traduit un essoufflement de son rôle de catalyseur. On l'a vu au moment de la création des universités nouvelles. Il n'y a pas eu de décision stratégique ni du CNRS ni des autres organismes de recherche pour s'y investir massivement. La décision a plutôt été d'attendre et de voir venir. Alors que ces universités ont recruté beaucoup d'enseignants-chercheurs, elles n'ont que très peu de laboratoires avec l'étiquette CNRS. Je pense que cela a été une erreur grave des organismes de recherche de ne pas s'investir pleinement dans ces établissements alors qu'ils avaient la capacité d'accompagner ce développement et même de contribuer à leur pilotage en recherche, en s'appuyant sur les nouveaux emplois que pouvait apporter le partenaire universitaire. Le CNRS a eu envie de se concentrer sur des centres d'excellence, ce qui est un argument compréhensible mais qui a été appliqué avec un manque de vision prospective. Cette myopie des organismes de recherche va être difficile à rattraper car ces universités n'ont plus le taux de croissance qu'elles ont connu au départ. Les autres universités

scientifiques non plus. On est maintenant devant le problème démographique du renouvellement des chercheurs et des enseignants-chercheurs. Il doit être abordé globalement par les université et les organismes de recherche.

La situation est particulièrement critique pour les maths à cause de plusieurs facteurs. J'ai déjà évoqué la discrétion et le manque de visibilité des mathématiques. Pour employer un euphémisme, la raison d'être de la recherche en maths n'est pas bien comprise dans le grand public, ce qui n'incite pas les gouvernements successifs à afficher les maths comme une priorité. Le moteur des créations de postes liés aux effectifs étudiants ne tourne plus au même régime qu'il y a 10 ans. La répartition des emplois entre les universités et les organismes conduit, en mathématiques, à confier à ces premières la responsabilité principale pour ce renouvellement, ce qui nécessite d'aborder la relation avec le CNRS suivant des modes différents de ce qui se fait ailleurs, comme en physique ou en chimie.

Autre sujet, les négociations des contrats de plan état-région ont mis en évidence l'importance, et la difficulté, des relations des universités et des autorités locales avec le CNRS, et les autres organismes. Ce sont ces contrats qui permettent les gros investissements, par exemple les constructions pour l'enseignement et la recherche, les équipements lourds de recherche, les réseaux informatiques... Cependant, le CNRS a une organisation nationale et politiquement centralisée autour des départements scientifiques. Le délégué régional n'a pas le pouvoir d'arbitrage entre projets différents, ce qui ralentit la prise de décision et empêche le CNRS de peser avec efficacité sur les choix qui se préparent en région. Les universités sont dans une situation plus simple et peuvent afficher des priorités scientifiques. Le CNRS est surtout handicapé quand il y a des problèmes de déplacements de crédits d'un projet soutenu par un département scientifique vers un projet soutenu par un autre, ou quand il s'agit du financement de projets transversaux (je pense au domaine de l'environnement par exemple). Les universités se retrouvent à être les seules à faire le travail global de prospective, le CNRS le faisant plutôt par discipline. Si cela est rarement embarrassant pour les maths, encore que ceci conduit souvent à l'écarter des financements d'équipements, c'est souvent gênant ailleurs et particulièrement dans certains domaines transversaux. Il reste d'autres questions à améliorer dans les relations entre CNRS et Universités, en particulier pour beaucoup d'aspects pratiques de la vie des UMR. En fait, on ne sait pas très bien si ces UMR sont des unités universitaires associées au CNRS (c'était le cas des anciennes URA) ou l'inverse (cas de plusieurs anciennes UPR). La théorie intermédiaire, l'ancienne direction du CNRS parlait de « copropriété », ce que j'ai toujours trouvé très ambigu, conduit à mal régler diverses questions matérielles portant sur la gestion des crédits, l'organisation de la valorisation, les personnels. J'espère que ces sujets, certes un peu secondaires, que les anciens responsables de cet organisme ne semblaient pas pressé d'aborder au fond, pourront maintenant faire l'objet de discussions sérieuses.

Quel rôle et quelle influence ont le fait d'être mathématicien sur l'exercice de la présidence d'une université ?

À mon avis, pas d'influence directe. Le travail d'un président est d'organiser la stratégie de l'établissement et de négocier avec les différents partenaires. Il

y a un petit avantage dès que l'on doit négocier les budgets : le fait de ne pas être issu d'une discipline coûteuse permet d'échapper à certaines critiques, ce qui donne un peu de sérénité et simplifie un peu la tâche du président. Mais ce n'est vrai que pour les équipements lourds en recherche. Pour l'essentiel, rien ne distingue les présidents mathématiciens de leurs collègues. Il ne me semble pas qu'il existe un style particulier dans l'exercice de la présidence par un mathématicien.

CARNET

Philippe Benilan (6 octobre 1940 - 17 février 2001)

Abdallah Al Amrami

Philippe Bénilan [1] est décédé le samedi 17 février 2001. Je voudrais à cette occasion douloureuse, évoquer un ou deux souvenirs à propos de Benilan, l'enseignant de mathématiques.
C'était en 1968-1969, à l'unique (à l'époque) Faculté des Sciences au Maroc (à Rabat). J'étais en MP1 (maths-physique, 1ère année). Une soixantaine d'étudiants dans cette filière (pour tout le pays [2]), dans un amphi flambant neuf [3], du nom d'Al Khawarizmi. Bénilan nous faisait cours, ainsi que quelques séances de travaux dirigés, en analyse. Jeune, beau, cravaté-costumé parfois [4], c'était un véritable ouvrier au tableau. Ouvrier voulant dire faisant véritable œuvre.

C'était de l'analyse bien fine. Les limites étaient justifiées, ε à l'appui. Les démonstrations naissaient, se développaient et arrivaient à terme, devant nous, poussière blanche sur tableau noir. C'était fascinant. Les étudiants n'étaient pas toujours patients, surtout à la fin de l'heure, les ε ne se découpant pas parfois bien facilement. Mais le sentiment de se trouver devant un professionnel-artiste n'échappait à personne.

Les TD avec lui, c'était le prolongement naturel du cours. C'était consistant, riche et vivant. Qu'un étudiant passe au tableau ou non, Bénilan était à l'œuvre, toujours. Mais cette fois, contrairement au cours, il nous faisait participer par ses questions permanentes, à ses démonstrations.

Il n'hésitait pas à fumer, devant nous. Ça ne nous gênait pas... on le sentait près de nous. C'est avec Coca et cigarettes sur la table, qu'il m'a fait passer l'oral en juin, au sous-sol de la Faculté, sur la convergence uniforme de fonctions continues... à pareil mois, au Maroc, il fait chaud. Nature ! berce-le fraîchement.

---

[1] Professeur à l'Université de Franche-Comté.
[2] On n'était plus que 15 environ en MP2. Malheureusement, sans Benilan comme prof.
[3] La faculté, elle, date des années 1950.
[4] Pour des raisons administratives occasionnelles, sans doute.

COURRIER DES LECTEURS

Mathématiques marocaines et francophonie

Mohamed Akkar (*Université Bordeaux I*)

« Aujourd'hui le Maroc est le pays où la langue française est non seulement aimée et respectée en tant qu'invitée mais elle fonde et participe au paysage culturel d'une société trilingue (arabe, berbère, français). »
(Tahar Benjelloun, Prix Goncourt 1987)

La francophonie en mathématiques ce n'est pas uniquement enseigner en français ; cette performance, pour les pays africains, est un héritage colonial et il est souhaitable qu'elle ne soit pas seulement un acte de paternalisme (les français, en général, à défaut de coloniser aiment bien « paterner » : penser à la grand-messe des sommets Afrique-France). Elle doit être un échange et un partage permanent du savoir mathématique (dans les revues, les séminaires, les colloques, les ouvrages, les compétitions et jeux...). Comme le dit mon collègue et ami Saliou Touré (ex-ministre de l'Enseignement Supérieur de Côte d'Ivoire) actuellement professeur de mathématiques à la Faculté des Sciences d'Abidjan : « La francophonie, c'est aussi une certaine conception de l'humanisme lié au développement, c'est l'affirmation des valeurs de la solidarité et du partage. »

Il y a quelque temps un collègue m'a contacté pour me demander d'écrire quelques lignes sur les mathématiques marocaines vues sous l'angle de la francophonie. J'avoue avoir été sensible au fait d'être choisi pour parler de ce sujet que je connais bien au titre de franco-marocain francophone et d'enseignant-chercheur en mathématiques. Après avoir bien réfléchi à la question je me suis rendu compte que je pouvais soit écrire un texte purement descriptif sur les activités des différents départements de mathématiques du Royaume (ce qui, à mon avis, n'a qu'un intérêt limité et n'apprendrait pas grand chose au lecteur quand il saura qu'on enseigne au Maroc, en français, par exemple au niveau bac + 3, l'intégrale de Lebesgue, le calcul différentiel dans les Banach et la théorie des corps commutatifs en Maths Pures, que les proba-stat., l'analyse numérique et la recherche opérationnelle font partie du cursus des Maths Appli. et qu'après la maîtrise on peut faire un DEA ou un DESS, qu'il existe des séminaires et que des thèses sont rédigées et soutenues en français) soit dire des vérités, pas toujours partagées par les autres, sur la coopération scientifique entre le Maroc et la France ou de façon plus générale entre l'Afrique et l'Hexagone. Je me suis donc abstenu. Mais dernièrement j'ai feuilleté le n°86 de la *Gazette des mathématiciens* (octobre 2000) et j'ai alors découvert le dossier Francophonie introduit par J.-J. Risler et surtout constaté que ce dernier (et la SMF) accordait de l'intérêt au sujet et posait un problème de fond : « faut-il encourager les mathématiciens francophones à publier en français ? » Comme dans ce dossier, les contributions africaines ne répondaient pas à la question je me suis dit que c'est l'occasion de compléter les articles de mes collègues algériens et tunisiens par quelques réflexions sur cet important problème. J.-J. Risler s'interroge aussi sur les publications mathématiques

en français. On peut ajouter une troisième question plus spécifique aux publications et ouvrages en français dans le domaine de l'enseignement. Enfin, disons que quelques collègues se demandent tout simplement : « pourquoi la francophonie en mathématiques ? Ne peut-on pas, ajoutent-ils, transmettre une culture dans une autre langue ? »

1. La francophonie en général.

N'étant pas spécialiste de cette grande question, qu'on me permette juste d'avancer quelques idées simples et peut-être naïves et quelques constatations. Pour moi (eu égard à mes origines) l'Afrique est le centre de gravité historique et géographique de la francophonie : en l'an 2000, il y a moins de francophones en France, en Europe et en Amérique réunies qu'ailleurs. Certes la langue française appartient à ceux qui la parlent et surtout ceux qui l'utilisent, mais il revient à la France (et peut-être aussi à la Belgique, au Québec...) de faire en sorte que cette langue continue à bien se porter, aussi bien au Maghreb qu'en Afrique subsaharienne. Pour cela la coopération culturelle et surtout scientifique de la France avec les pays francophones africains doit être soutenue et doit se développer indépendamment d'intérêts mondialistes, européens ou commerciaux : on constate aujourd'hui qu'une bonne partie de l'aide scientifique anciennement accordée à l'Afrique francophone s'est réorientée vers des pays anglophones, vers des pays de l'Est ou vers des marchés potentiels d'Amérique du Sud. Qu'on décide d'intensifier la coopération française avec le Ghana ou la Pologne, le Kenya ou la Hongrie ou encore l'Arabie Saoudite, le Brésil ou la Tchécoslovaquie, il doit y avoir des raisons légitimes à cela et que je ne discute pas, mais je ne peux comprendre que cela se fasse aux dépens du centre de gravité de la francophonie !

Depuis quelques années la France fait payer le prix fort à ceux qui veulent étudier dans les établissements français en Afrique. Maintenant, seuls les enfants des privilégiés marocains fréquentent ce qu'on appelle les établissements de la Mission culturelle française. De même, on voit beaucoup moins d'étudiants, enfants des couches populaires, venir étudier dans les Universités françaises car, pour obtenir un visa d'études, les parents doivent attester (par un document bancaire) de l'envoi mensuel de l'équivalent du salaire d'un professeur de lycée. Comme il n'y a plus de bourse, même pour les plus brillants, la francophonie sera destinée à une catégorie sociale plus qu'aisée. Un autre phénomène qui aggrave la situation est le suivant : pour apprendre le français et la culture française au Maroc (et ailleurs) il faut un environnement linguistique (revues, films, théâtre, expositions, conférences...). Ceci était proposé dans le temps par les centres culturels français financés par le Ministère des Affaires Étrangères. Tout cela est bien terminé ; au Maroc par exemple l'action culturelle française doit être auto-financée par les recettes (droits d'entrée et d'études). Le résultat est simple : la majorité des élèves (de parents non fortunés) écoutent le français par l'intermédiaire des séries télévisuelles américaines.

2. Les mathématiques francophones.

Revenons aux sciences et surtout aux mathématiques. Dans les colloques internationaux les Africains francophones exposent, presque toujours, en français (je me rappelle avoir fait un scandale une fois, à l'Université de Leeds, quand j'ai fait une conférence en français alors que mon séjour était financé par le British Council de Rabat !). Ce qui est loin d'être le cas des français. Un souvenir (même lointain) : au congrès international des mathématiciens de Helsinki un collègue français avait fait un exposé sur l'action des IREM... en anglais... et à la fin de sa prestation j'ai cru bon d'intervenir, en français, pour donner mon point de vue sur le rôle des IREM lors de la réforme des programmes scolaires et voilà qu'un autre collègue français bienveillant a éprouvé le besoin de traduire en anglais mes propos. Tout cela pour dire, que de

mon point de vue, les premiers détracteurs de la francophonie sont... les français eux-mêmes. Il y a 3 ans je crois, j'ai été invité, en tant que président du jury de l'agrégation marocaine de mathématiques, à une réunion d'information, au Lycée Janson à Paris, sur cette fameuse troisième épreuve d'oral sur la modélisation scientifique. Pour illustrer ce qu'on attend des agrégatifs lors de cette épreuve la Présidente du Jury de l'époque a distribué à tous les responsables des prépas-agrég de l'hexagone un texte mathématique (modèle) en anglais et une bibliographie d'ouvrages (3 en français et 27 en anglais). Dans tous les colloques internationaux à l'étranger et même en France la plupart des conférenciers français exposent en anglais. La réponse habituelle est la suivante : c'est pour se faire comprendre par le maximum de chercheurs et ainsi faire connaître davantage ses travaux. Il se pose donc un problème : pourquoi la francophonie en mathématiques ? Ce qui est important : est-ce de toucher le maximum d'auditeurs ou de faire savoir que les mathématiques francophones se portent bien, sont de bonne qualité et que pour y accéder un petit effort est à faire ? S'il y a une culture mathématique francophone, s'il y a un état d'esprit et une façon de faire des mathématiques à la française il faut exposer et rédiger en français.

Je ne nie pas le fait que pour mieux vendre un livre ou pour diffuser plus largement une théorie, l'utilisation de l'anglais est plus avantageuse. Il est clair aussi qu'en recherche mathématique le français ne peut résister à l'actuelle hégémonie anglo-américaine, mais il faut rester vigilant ; le français ne peut se mesurer à l'anglais, mais ce qu'il faut espérer c'est qu'il ne décroche pas de la production scientifique. Le français a eu son heure, comme le grec (3ème siècle av. J.C. – 4ème siècle ap. J.C.), le latin (12ème siècle – 16ème) et l'arabe (du 8ème siècle au 13ème). Aujourd'hui c'est l'anglais. Il ne faut pas s'inquiéter de cette prédominance, par contre il ne faut pas l'encourager en présentant des thèses rédigées sous forme d'articles de revues en anglais ou en exposant en anglais dans un séminaire hebdomadaire courant. Le grave danger qui guette la francophonie mathématique c'est que le français cesse aussi (il n'est déjà plus langue de communication scientifique) d'être une langue d'usage ! Ceci voudrait dire par exemple que dans un domaine de recherche nouveau qui vient de démarrer dans un pays anglophone les chercheurs francophones se mettent à utiliser l'anglais pour leur travail personnel, pour leur usage interne sans faire l'effort de réfléchir, de chercher, de conjecturer, d'élaborer leur savoir dans la langue française. L'avenir de la francophonie mathématique est moins mis en cause dans les congrès que dans les laboratoires des universités et du CNRS. Si on se laisse aller, le français mathématique ne serait plus une langue vivante et encore moins une langue douée de vitalité et de pouvoir de création et d'invention. D'où la nécessaire solidarité de la communauté mathématique francophone dont la langue commune est un trait d'union et un facteur de rapprochement et de dialogue et inversement la francophonie est un outil d'enrichissement et de renforcement de la culture et de la langue françaises.

3. Publications en français.

J'en arrive à la question de J.-J. Risler : faut-il publier en français ? Faut-il encourager les jeunes chercheurs (qui ont besoin de faire connaître leurs travaux) à publier en français ? En réalité la question qui se pose, surtout aux jeunes chercheurs mathématiciens francophones, est différente, bien qu'étroitement liée à la première : a-t-on intérêt, pour être publié, à rédiger sa production en français ? Si je réponds oui à la première question, ma réponse à la seconde, par expérience sera non. En réalité je suis assez gêné de dire clairement ce que je pense car la situation des publications dans les revues francophones, surtout françaises est assez ambiguë (il y a, me semble-t-il, un peu de copinage et quelques clans... j'espère me tromper). En tout cas, les africains en général et les maghrébins en particulier

ont beaucoup de difficultés à y publier. J'encourageais souvent, il y a quelques années, de jeunes chercheurs marocains méritants d'envoyer leurs articles à Paris, à Grenoble... mais plus maintenant car, au Maroc , pour soutenir une thèse il faut 2 ou 3 publications dans des revues à comité de lecture international, et s'il fallait attendre la bonne humeur de certains rapporteurs bien hexagonaux, la carrière mathématique des jeunes lions de l'Atlas serait bien retardée pour ne pas dire bloquée. Avant d'attendre le pavé qui va éblouir les rédacteurs des revues françaises je conseille aux jeunes chercheurs africains d'envoyer leurs textes à des revues francophones en dehors du carré hexagonal (au Québec, en Belgique...) ou tout simplement de rédiger leur texte en anglais et de le publier aux USA (attention mon ami Michel Mendès France qui est en plus anglophone, m'a appris qu'il y a une revue américaine qui ne publie que les articles rédigés en Américain !) ou en Inde ou encore au Japon. Je leur conseille aussi de s'adresser à des revues polonaises, hongroises. Ceux qui veulent quand même conserver leur texte en français devraient s'adresser aux revues italiennes, espagnoles ou roumaines (elles ont l'air d'être francophiles). À propos de revue indienne permettez moi de vous conter l'histoire suivante : un de mes anciens élèves avait publié un article, à mon agréable surprise, en anglais dans une revue française ! Quelques années plus tard, un brillant chercheur marocain trouvait un résultat remarquable sur le même sujet mais qui allait beaucoup plus loin aussi bien sur le plan des méthodes et des techniques de démonstration que sur le plan de la puissance et de l'élégance du résultat. Au lieu d'envoyer le travail à un Rendiconti ou à une Hungarica j'ai poussé mon jeune poulain à bien le rédiger en français et à l'envoyer à la même revue franco-française. La réponse était la suivante (je n'en ferai aucun commentaire, ce n'est pas l'objet ici) : l'article est correct, seulement il est trop spécialisé pour le large public que nous souhaitons intéresser. Par dépit je l'ai envoyé à une revue indienne (car entre autres choses le dit article généralisait un théorème d'un chercheur indien publié par cette même revue). Il fut accepté sans aucune correction... à condition qu'il soit rédigé... en anglais. Je dois raconter une autre mésaventure : j'ai envoyé un article il y a 3 ans (malgré ce que j'ai dit plus haut) d'un chercheur marocain très fin au rédacteur français d'une revue. Quelques jours après j'ai reçu mon enveloppe non ouverte dans une belle enveloppe avec un en-tête illustre et le petit mot non moins illustre suivant : nous avons tellement d'articles en attente qu'il s'avère impossible d'examiner votre travail. J'avais vraiment envie d'écrire à cet honorable rédacteur pour lui dire que malgré ses dons de Madame Soleil à lire à l'intérieur des enveloppes, il s'était trompé, pour la simple raison que mon enveloppe ne soumettait aucun article mathématique mais contenait plutôt un (autre) don en dollars africains destiné à l'association de la qualité de la science.

Pourquoi je réponds oui à la question de J.-J. Risler : il faut encourager, malgré tout, les jeunes à publier en français ou plutôt les aider à publier en français. Tout simplement parce que cela rendra justice à la production scientifique des chercheurs francophones qui est de qualité, surtout en mathématiques. Et cela contribuera au rééquilibrage des publications mathématiques devant le raz de marée anglo-saxon et l'impérialisme d'une langue et d'un pays, surtout que la situation risque d'empirer avec l'informatique (il faudra un jour se pencher sur la francophonie et Internet). Pour me faire mentir, j'attends des rédacteurs de revues françaises d'encourager les jeunes chercheurs africains à publier chez eux. Qu'on ne se méprenne pas sur mes intentions, je ne demande pas de priorité francophone mais une attention particulière aux articles de tous ces jeunes chercheurs élevés dans la culture et la science française pour les aider à progresser et à s'affirmer dans la

langue qu'ils aiment et qu'ils veulent partager scientifiquement et culturellement. Comme le dit J.-J. Risler dans l'introduction au dossier Francophonie il faut « soutenir les collègues étrangers qui combattent pour la francophonie, ou au moins leur marquer un peu d'attention ». Les africains sont de fervents et sincères défenseurs de la francophonie, qu'on leur accorde alors la place qu'ils méritent en tant qu'acteurs actifs. Il est tout à fait anormal de constater qu'ils sont peu représentés au sein des instances internationales de la francophonie, pour jouer des rôles scientifiques (et non politiques) et il est également curieux que malgré les potentialités de recherche mathématique en Afrique il y a peu ou presque pas d'équipes africaines qui bénéficient de subventions francophones pour organiser des colloques ou des écoles de recherche ou encore des compétitions mathématiques pour les jeunes ! Pour illustrer cette situation permettez moi, encore une fois de vous conter ce qui s'est passé à l'ouverture du colloque EM2000 (l'enseignement des mathématiques dans les pays francophones au XXè siècle...) qui a eu lieu à Grenoble du 15 au 17 juillet 2000 : il y a eu plein de discours intéressants des organisateurs, de l'ICMI, de francophones non africains... ; il y a eu beaucoup d'annonces de réunions mathématiques importantes dans le monde, au Japon, en Australie et je ne sais plus où encore. Mais rien n'a été dit ni par un africain ni sur l'Afrique, bien qu'il y avait une quarantaine d'africains francophones sur les presque 200 participants (Algérie, Côte d'Ivoire, Maroc, Mauritanie, Tunisie...). Devant cette absence intolérable j'ai pris la parole en tant qu'africain en disant que j'avais une annonce importante à faire : « Vu que la Fifa vient de refuser au Maroc et à l'Afrique du Sud l'organisation de la prochaine coupe du monde de football, les mathématiciens africains ont décidé d'organiser un safari mathématique en 2002 qui aura lieu quelque part en Afrique et vous y êtes tous invités. »

4. Conclusion.

Je voudrais conclure en disant quelques mots sur les ouvrages de mathématiques. La situation est également désastreuse (au grand désavantage du français) en ce qui concerne les publications et ouvrages pour l'enseignement. Seulement, je suis convaincu que si la lutte avec l'anglais est perdue d'avance dans le domaine de la recherche, il n'en serait pas de même si les mathématiciens et les autorités francophones faisaient un effort substantiel pour encourager la publication d'ouvrages d'enseignement de qualité chez des éditeurs francophones pour les diffuser ou/et les offrir aux bibliothèques des départements de mathématiques africains. Il existe certes quelques ouvrages de DEUG ou de niveau supérieur écrits par des mathématiciens francophones de talent (Choquet, Dixmier, Godement...), mais il n'y en a pas beaucoup. Et dans ce domaine il reste beaucoup de choses à faire (dans la rubrique Livres de la *Gazette des mathématiciens* (n°86 toujours) par exemple, 14 livres sont analysés (je n'ose pas dire reviewés) mais 13 sont en anglais et un seul en français (merci à la série *Panoramas et Synthèses*)). Et c'est possible d'écrire en français de bons livres pour l'enseignement qui seraient analogues à ceux de Rudin ou de Lang. Si on est incapable de bien écrire, il n'y a qu'à traduire ; à ce propos je veux dire encore deux petites choses : l'excellent livre d'Analyse fonctionnelle de Rudin a été traduit en français par mon compatriote Abdellatif Abouhazim qui enseignait il y a quelques années à la Faculté des Sciences de Kénitra, et le projet de traduction de livres pour la bibliothèque Scopios va dans le bon sens. Mais il faut écrire des livres en français, chers collègues francophones !

Cohomologie, stabilisation et changement de base.
Jean-Pierre Labesse Avec deux appendices *Changement de base pour les représentations cohomologiques de certains groupes unitaires* par Laurent Clozel et Jean-Pierre Labesse, *Ensembles croisés et algèbre simpliciale* par Lawrence Breen
Astérisque **257** (1999), 167 pages. ISSN : 0303-1179. 200 FF

The title doesn't say so, but this book is about understanding the trace formula — the twisted Arthur-Selberg trace formula, to be more precise. The space $\mathcal{A}_d(G)$ of discrete automorphic forms on a connected reductive algebraic group G over $\mathbb{Q}$ is an enormous object, but its structure can in principle be understood completely in terms of a single, admittedly complicated, invariant: its trace. A smooth compactly supported function f on the adele group $G(\mathbb{A})$ of G (a *test function*) naturally defines an operator $r(f)$ on $\mathcal{A}_d(G)$, and it is known (thanks to work of W. Müller) that $r(f)$ is an operator of trace class. The functional $f \mapsto \operatorname{tr} r(f)$ is a distribution on $G(\mathbb{A})$ invariant (under conjugation by $G(\mathbb{A})$) that determines the representation of $G(\mathbb{A})$ on $\mathcal{A}_d(G)$ up to isomorphism. More precisely, $\mathcal{A}_d(G)$ can be written as an infinite direct sum

$$\mathcal{A}_d(G) = \bigoplus_\pi m(\pi)\pi, \tag{1}$$

where π runs over irreducible admissible representations of $G(\mathbb{A})$ (with a few adjustments at the archimedean place) and $m(\pi)$ are non-negative integers. Then we have the equality of distributions

$$\operatorname{tr} r = \sum_\pi m(\pi) \operatorname{tr} \pi \tag{2}$$

where $f \mapsto \operatorname{tr} \pi(f)$ is the distribution character of the representation π, the fundamental invariant in harmonic analysis on the locally compact group $G(\mathbb{A})$.

Formula (2) is called the *spectral expansion* of $\operatorname{tr} r$. In the representation theory of a finite group H, the characters of irreducible representations of H form a basis for the space of invariant functions, and a second basis is formed by the characteristic functions of conjugacy classes. In the theory of admissible representations of $G(\mathbb{A})$, the characters are replaced by the distribution characters, which are distributions rather than functions; the characteristic functions of conjugacy classes are replaced by orbital integrals:

$$\Phi_{G(\mathbb{A})}(\gamma, f) = \int_{I_\gamma \backslash G(\mathbb{A})} f(x^{-1}\gamma x) d\dot{x}. \tag{3}$$

Here I_γ is the centralizer in $G(\mathbb{A})$ of γ and $d\dot{x}$ is a quotient measure. (Thus $\Phi_{G(\mathbb{A})}$ depends on choices of measures, but so does the trace.) The *geometric expansion* of $\operatorname{tr} r$ is then an expression of the form

$$\operatorname{tr} r = \sum_\gamma \Phi_{G(\mathbb{A})}(\gamma, \bullet) \tag{4}$$

where γ runs over conjugacy classes in $G(\mathbb{Q})$. The trace formula is then the identity of the spectral and geometric expansions of $\operatorname{tr} r$. If G were momentarily replaced by

the additive group, the dictionary between characters and conjugacy classes (which are just elements) is provided by Fourier analysis, and the trace formula reduces to the Poisson summation formula. Thus the trace formula is a non-abelian analogue of the Poisson summation formula.

Unfortunately, as anyone with even the slightest familiarity with the trace formula will already have noticed, such a trace formula exists only when G is anisotropic; equivalently, when $G(\mathbb{Q})\backslash G(\mathbb{A})$ is compact. This excludes most of the interesting cases, so we have to try again. The reasoning that yields the identity between the right-hand sides of (2) and (4) runs into serious problems of convergence when $G(\mathbb{Q})\backslash G(\mathbb{A})$ is non-compact. For $SL(2)$ and certain other groups of rank 1 these problems were first addressed, and solved, by Selberg; the general case was successfully resolved by Arthur, in a series of articles spanning nearly 20 years. In the correct formula, both expansions must be supplemented by analytically elaborate terms (weighted characters on the spectral side, weighted orbital integrals on the geometric side). Labesse's book is *not* about these difficulties, and I will say nothing more about them. Rather, Labesse is concerned with a problem that arises once one attempts to apply the trace formula to obtain useful information.

Taken alone, the geometric expansion sheds no light on the spectral expansion, the matter of genuine interest. In most applications, the trace formula is used to compare spectra of two distinct, but related groups G and G'. This is achieved by relating the geometric expansions for the two groups. This may be possible when there is a map (the *norm*, in the cases considered by Labesse) from rational conjugacy classes in G' to rational conjugacy classes in G, and a second map (*transfer*) of test functions on G' to test functions on G. The first classic example is the Jacquet-Langlands correspondence between automorphic forms on $G = GL(2)$ and on an inner form G' (the multiplicative group of a quaternion algebra), proved precisely by comparing geometric expansions on the two groups.

The second classic example is the base change of Labesse's title. Let E/F be a cyclic extension of number fields of prime degree, θ a generator of $Gal(E/F)$, and let $G = GL(n)_E$, $G' = GL(n)_F$, viewed as groups over $\mathbb{Q}$ by restriction of scalars. Base change is a map from automorphic forms (more correctly, automorphic representations) on G' to automorphic forms on G. The map $\gamma \mapsto x^{-1}\gamma\theta(x)$, for $x \in G(\mathbb{A})$, defines an equivalence relation called twisted conjugacy on elements of $G(\mathbb{A})$, that can also be expressed in terms of conjugacy on the *disconnected* group $Gal(E/F) \ltimes G$. In this case there is a norm map from twisted conjugacy classes in G to conjugacy classes in G', as well as a *twisted* transfer from test functions on G to test functions on G', and base change is obtained by comparing the geometric expansions of the trace formula on G' and the twisted trace formula on G.

Attempts to apply these techniques to other groups immediately run into the problem that, in general, norm maps are only defined on conjugacy classes over algebraically closed fields. Let F be either $\mathbb{Q}$ or a local field, and G a connected reductive group over F. The elements $\gamma, \gamma' \in G(F)$ are *stably conjugate* if they become conjugate over the algebraic closure of F. Conjugacy and stable conjugacy coincide for inner forms of $GL(n)$, but are distinct for other groups. In joint work with Langlands in the 1970s, Labesse discovered that the distributions appearing in the geometric and spectral expansions of the trace formula for $SL(2)$ are not stable; i.e., are not invariant under stable conjugacy. This phenomenon is the rule. Stabilization of the trace formula refers to the expectation that the geometric and spectral expansions can be written as sums of stable distributions, not on G alone but on a finite set of related groups called *endoscopic groups*.

The difference between conjugacy and stable conjugacy can be measured in terms

of Galois cohomology. This approach, initiated by Langlands, developed systematically by Kottwitz, and extended to twisted conjugacy by Kottwitz and Shelstad, is initially expressed in terms of non-abelian cohomology, but, using basic results of Kneser, they found that the relevant obstructions lie (for the most part) in abelian groups. Generalizing Tate-Nakayama duality for tori, they found expressions for these obstruction groups in terms of the Langlands dual groups. The resulting theory is efficient but difficult to grasp, because (1) there are different kinds of obstruction groups (for local and global fields, and for the comparison between the two), whose definitions increase in complexity as generality is increased, and (2) to apply the theory to the trace formula one has to undo the dualization; the constructions that appear natural in the setting of the Langlands dual are no longer so back on the group.

The principal innovation of Labesse's book is the development of new cohomological methods that apply directly to the group, without passing through dualization. The fundamental notion, introduced on the second page of the first chapter, is what Labesse calls a "crossed set". This consists of a triple (X, A, B), where X is a set, A is a group acting on X, and B is a set fibered in groups over X and admitting an action by A and a map to $A \times X$. The whole satisfies a list of axioms at first sight far from intuitive, and this is hardly surprising: as L. Breen explains in an extended appendix, the categorical structure naturally related to a crossed set is that of a (certain kind of) 2-groupoid.

Nevertheless, Labesse manages to define Galois cohomology with coefficients in a crossed set. Certain kinds of complexes of reductive groups over local and global fields define crossed sets; this is the case, for example, for the inclusion $I = I_\gamma \mapsto G$ considered above, at least when γ is a semisimple element of G. In this way, when (say) F is a local field, Labesse can define Galois cohomology sets and abelianized Galois cohomology groups $H^i(F, I\backslash G)$ (for small i) and $H^i_{ab}(F, I\backslash G)$, respectively, with excellent functorial properties. For instance, there is a long exact sequence (of pointed sets)

$$1 \to I(F) \to G(F) \to H^0(F, I\backslash G) \xrightarrow{\partial} H^1(F, I) \to H^1(F, G) \to \dots.$$

The difference between conjugacy and stable conjugacy is then given by the image of the map ∂. This is the beginning of the local and adelic theory of stabilization. The stable and unstable orbital integrals are defined as integrals over the set $H^0(F, I\backslash G)$, or its adelic counterpart, which naturally can be represented as a disjoint union of sets of the form $I'(F)\backslash G(F)$, where the I' are certain inner twists of I.

The invariants appearing in the global theory of stabilization arise when F is replaced in the above diagrams by $\mathbb{A}/\mathbb{Q}$, essentially the cone on the global Galois cohomology and the product over completions of the local Galois cohomologies. This provides two sets of long exact sequences for the abelianized cohomology groups: one for the morphism of groups $I \mapsto G$, the other for adelic localization of Galois cohomology. Comparing these exact sequences, Labesse defines the *endoscopic character group* $\mathfrak{K}(I, G; F)_1$ relative to I as the group of characters of the cokernel of the map $H^0_{ab}(\mathbb{A}, G) \to H^0_{ab}(\mathbb{A}/\mathbb{Q}, I\backslash G)$.

The endoscopic character groups, which are finite in the cases of interest, are defined by means of dualization in Kottwitz' (pre)stabilization of the elliptic part of the standard trace formula, and in the Kottwitz-Shelstad (pre)stabilization of the strongly regular elliptic terms in the twisted trace formula. Labesse's formalism, provides a natural (pre)stabilization of all elliptic terms in the twisted trace formula,

at least in the cases relevant to cyclic base change. More precisely, via a natural sequence of maps

$$H^0(\mathbb{A}, I\backslash G) \to H^0(\mathbb{A}/\mathbb{Q}, I\backslash G) \to H^0_{ab}(\mathbb{A}/\mathbb{Q}, I\backslash G)$$

an endoscopic character defines a function on $H^0(\mathbb{A}, I\backslash G)$ that can be inserted into the orbital integral. In this way, Labesse obtains a natural expression for the sum of elliptic terms as a sum of κ-orbital integrals, where κ runs through the group of endoscopic characters; this is what is meant by prestabilization.

The elliptic part of the trace formula corresponds to the terms in the geometric expansion that have the naive form indicated in formula (4). For certain test functions f one actually obtains equalities between the naive spectral and geometric expansions:

$$\sum_{\pi} m(\pi) \operatorname{tr} \pi(f) = \sum_{\gamma} \Phi_{G(\mathbb{A})}(\gamma, f), \tag{5}$$

where now γ runs over elliptic conjugacy classes. Such a formula holds in general for anisotropic G, of course, but also when f satisfies local hypotheses at two places; one then obtains the *simple* trace formula, twisted or not. Under an additional local hypothesis, all κ-orbital integrals of f vanish for $\kappa \neq 1$, and only the stable term remains. Without further ado, the geometric expansion is thus in a state to be compared with the trace formula of another group.

The local hypotheses are somewhat restrictive, but they suffice for a variety of applications, notably to local questions. More immediately, they allow Labesse to prove the existence of cyclic base change, and transfer between inner forms, in a variety of situations in which they were not previously known. As an application of his methods, Labesse proves the existence of cyclic base change of cuspidal automorphic representations locally Steinberg at an appropriate set of finite places (any two finite places will do if the group is semisimple and simply connected). In a final application, this theorem is applied to prove the non-triviality of cuspidal cohomology of S-arithmetic groups, generalizing a theorem of Borel, Labessse, and Schwermer.

The book concludes with two appendices. Breen's appendix was already mentioned; it reinterprets Labesse's cohomological formalism in terms of simplicial sets. The other appendix, due to Clozel and Labesse, completes and corrects some errors in Clozel's proof of the existence of base change and descent of cohomological automorphic representations of unitary groups of division algebras over CM fields. This is the principal ingredient in Clozel's theorem associating compatible systems of ℓ-adic representations to (certain) cohomological automorphic representations of $GL(n)$. Clozel's theorem, in turn, is used in an essential way in recent work on the local Langlands conjecture.

Arthur has recently obtained the full stabilization of both sides of the (untwisted) trace formula, assuming a series of conjectures in local harmonic analysis. These conjectures, generally known as the "fundamental lemma," have thus become the central question for everyone concerned with the trace formula. Indeed, Waldspurger has already reduced the problem of (endoscopic) transfer to the fundamental lemma. Much of Labesse's book is taken up with questions of local harmonic analysis, deriving the cases of transfer necessary for base change from Waldspurger's result (the relevant fundamental lemma was already known for cyclic base change, though the published proofs are incomplete; Labesse's book also includes a complete proof). It is clear to the reviewer that Labesse's formalism will be the language of choice in extending

Arthur's results to the twisted trace formula. For this reason alone[1], Labesse's book is required reading for anyone with an interest in the future of automorphic forms.

Michael Harris, Université Denis Diderot (Paris 7)

Leçons de mathématiques d'aujourd'hui
Éditées par Éric Charpentier et Nicolas Nikolski
Cassini, 1998. 384 p. ISBN 2-84225-007-9. 98 F

Les « leçons de mathématiques d'aujourd'hui », publiées par les Éditions Cassini, rassemblent en un ouvrage douze « Leçons » données à Bordeaux entre 1993 et 1997. L'École doctorale de mathématiques de l'Université de Bordeaux a en effet pris l'excellente initiative d'inviter des conférenciers à présenter un thème de recherche d'une manière accessible aux étudiants, et de faire rédiger ces conférences par l'un des auditeurs, en suivant au plus près le discours parlé. Le fruit de ces efforts orchestrés par Éric Charpentier et Nicolas Nikolski est un livre original, où les travaux contemporains dévoilent toute leur variété. Il est exceptionnel de rencontrer un ouvrage qui bouscule ces barrières qui trop souvent cloisonnent la recherche, tout en dépassant de bien loin le niveau des conférences de salon. Parcourons donc ensemble les thèmes abordés dans ces douze leçons.

La conférence de Jean Pierre Kahane commence par le théorème de Pythagore, replacé dans son contexte historique, et diverses façons de l'établir, dont le plus simple consiste à remarquer qu'on crée trois triangles semblables en abaissant la hauteur issue de l'angle droit d'un triangle rectangle. L'itération de cette opération conduit alors à la courbe de Polya, qui remplit tout un triangle rectangle, et à l'analyse multifractale. La courbe de Polya fournit une solution du « problème du voyageur de commerce » pour la variation quadratique, ce qui nous conduit naturellement au mouvement brownien. Enfin, la dynamique des mesures harmoniques, encore mal comprise, est reliée aux dendrites de la minéralogie, formées par exemple par le dépot du sulfate de cuivre sur la cathode. Une discussion très instructive conclut la conférence.

Pierre Cartier développe dans sa conférence un cadre rigoureux pour l'intégrale de Feynman. Il expose tout d'abord les difficultés liées à l'absence de mesures invariantes en dimension infinie (rappelons nous que la compacité locale est une condition nécessaire à l'existence de la mesure de Haar), puis examine les solutions apportées par Wiener pour formaliser le mouvement brownien, puis par Feynman pour formaliser la mécanique quantique. Les chaînes de Markov, présentées ici comme des graphes orientés, sont l'outil probabiliste « élémentaire » qui joue le rôle majeur. Le délicat problème de normalisation, proposé aux mathématiciens par Feynman, est analysé.

Le calcul des serpents, c'est-à-dire du nombre de types topologiques des fonctions réelles ayant un nombre fini de points critiques avec des valeurs critiques toutes différentes, sert de point de départ à Vladimir Arnold. L'une des figures de sa conférence sert également de couverture au livre. Il montre que ce calcul d'apparence élémentaire, qui fournit une suite de nombres (K_n) commençant par 1, 1, 1, 2, 5, 16, 61... est lié *via* les fonctions génératrices aux développements non triviaux des fonctions classiques (comme la fonction tangente) en série, et d'autre part *via* les discriminants à la théorie de Morse et à la théorie des catastrophes. Cette suite est ensuite reliée à la combinatoire des groupes de Coxeter, c'est-à-dire des groupes finis engendrés par des réflexions dans un espace euclidien.

La conférence de Don Zagier concerne la cohomologie de $SL_2(\mathbf{Z})$. Il s'agit là d'un sujet très ardu pour un non-spécialiste (comme par exemple l'auteur de ces lignes).

---

[1] Attention: "*For this reason alone*" veut dire : cette raison est suffisante; je ne veux pas dire "c'est la seule raison..." !

Cependant D. Zagier sait communiquer des idées simples, comme de voir la fonction *sinh* comme un cas limite des fonctions θ de Jacobi, ou de faire le lien entre les formules d'addition des cotangentes et les relations des périodes dans la théorie des formes modulaires. Des applications arithmétiques sont esquissées ; on devine derrière ces applications de vertigineux calculs. L'un des résultats les plus surprenants relie le nombre de solutions d'une équation fonctionnelle avec le spectre du Laplacien agissant sur le quotient du demi-plan de Poincaré par $SL_2(\mathbf{Z})$.

Haïm Brézis commence par étudier les fonctions à valeurs dans le cercle unité qui minimisent la fonctionnelle d'énergie de Dirichlet dans le disque, et remarque que le degré topologique de telles fonctions est nécessairement nul. Il s'intéresse ensuite aux fonctions de degré strictement positif, pour lesquelles cette fonctionnelle est infinie, et ajoute à la fonctionelle de Dirichlet une intégrale dépendant d'un paramètre ε qui permet de définir l'énergie de Ginzburg-Landau, laquelle provient de la théorie de la supraconductivité. Il explique alors comment s'inspirer d'un problème tridimensionnel rencontré dans un problème des cristaux liquides pour minimiser l'énergie de Ginzburg-Landau, puis fait tendre ε vers 0 pour obtenir des solutions « de superfluidité » de degré positif à valeurs dans le disque du problème de minimisation de la fonctionnelle de Dirichlet. On a donc obtenu ces fonctions par une méthode de pénalisation et renormalisation. Les singularités de ces fonctions sont l'expression mathématique des lignes de vortex de l'hélium liquide superfluide en rotation.

Dans sa conférence, Bernard Malgrange définit d'abord le polynôme de Bernstein-Sato d'une fonction f de n variables holomorphe au voisinage de 0 telle que $f(0) = 0$. Ce polynôme est issu d'une identité différentielle reliant f^s et f^{s+1}. Ses zéros interviennent dans le prolongement analytique de l'intégrale de f^s quand celui-ci est obtenu d'une façon similaire à celui de la fonction Γ, et ils font apparaître des pôles de ce prolongement. Ils apparaissent similairement dans le développement asymptotique des intégrales oscillantes du type « Fresnel » lorsque la phase (« stationnaire ») a des points critiques, en exposants des termes du développement. La partie la plus profonde de la conférence consiste à relier les zéros du polynôme de Bernstein-Sato de f à des invariants topologiques liés à f. Il y est établi que les valeurs propres de la monodromie de f en 0 sont au facteur $(-2i\pi)$ près les exponentielles des zéros du polynôme de Bernstein-Sato de f ; pour expliquer cela, il faut d'abord introduire l'homologie singulière, et la fibration localement (mais non globalement) triviale de Milnor. Le théorème de monodromie apparaît alors comme une conséquence de la rationalité des zéros des pôlynomes de Berstein-Sato.

Le point de départ de John Coates est le problème des nombres congruents : quels sont les nombres entiers qui sont l'aire d'un triangle rectangle à côtés rationnels ? Ce problème très ancien n'est pas encore pleinement résolu ; on conjecture que tout entier sans facteur carré congru à 5, 6 ou 7 modulo 8 est congruent. Fermat a montré que 1 n'est pas congruent, en se ramenant à l'équation $x^4 - y^4 = z^2$, et en établissant par sa méthode de descente infinie que celle-ci n'a pas de solution non triviale. En fait, 5 est le plus petit des nombres congruents. Un calcul simple montre ensuite que D est congruent si et seulement s'il existe un point rationnel d'ordonnée non nulle sur la cubique non singulière d'équation $y^2 = x^3 - D^2x$. Le lien est ainsi fait avec le groupe des points rationnels de la cubique, qui est abélien de type fini, et donc avec les fonctions elliptiques, puis avec les coefficients des formes modulaires. Notons que la loi de groupe suffit à montrer que si D est congruent, il existe une infinité de triangles solutions. John Coates introduit ensuite les points de p^n- divisions (qui généralisent les « points d'inflexion »), la cohomologie de la courbe elliptique et les « fonctions L » elliptiques. La conférence se termine par la difficulté qui arrêtait A. Wiles dans la démonstration, alors incomplète, du « dernier théorème » de Fermat. Cet obstacle a depuis été contourné par A. Wiles qui a établi ce théorème, mais à ma

connaissance l'obstruction initiale n'a toujours pas été surmontée.

Nous revenons à l'analyse avec la conférence d'Yves Meyer, qui remarque d'abord qu'en remplaçant les polynômes de degré n par des rapports de tels polynômes, on obtient des approximations beaucoup plus précises de la fonction $|x|$ en doublant seulement le nombre des coefficients à déterminer. Cependant, l'ensemble des approximants est alors une variété et l'approximation n'est plus linéaire. Le théorème de Peller explique ce résultat en montrant que les fonctions « bien approximables » par les fractions rationelles sont exactement celles qui appartiennent à une intersection d'espaces de Besov, ce qui autorise la présence de « pointes » bien que l'approximation soit bonne. Les ondelettes fournissent le bon point de vue sur ce théorème ; en effet les fonctions de Peller sont exactement celles dont la suite des coefficients sur une base d'ondelettes, réordonnée de façon décroissante, est à décroissance rapide. Cette méthode d'approximation non linéaire, qui consiste à ordonner par les amplitudes et non par les fréquences, est de fait beaucoup mieux adaptée aux ondelettes qu'aux séries de Fourier ; un résultat récent de T. Körner montre par exemple que le « théorème de Carleson » devient faux quand on réordonne la série de Fourier par coefficients décroissants. Le point de vue « ondelettes » conduit à une extension du théorème de Peller à la dimension n. De nombreuses applications industrielles de l'approximation non linéaire sont évoquées (traitement du signal, *monitoring* des centrales nucléaires...), et la conférence est émaillée de remarques lumineuses sur les problèmes posés par la reconnaissance de la parole ou des formes, et le rôle important joué par l'analyse fonctionnelle dans leurs solutions.

Henry Helson présente un panorama essentiellement historique de l'analyse harmonique de 1945 à 1965. Cette période où suivant ses termes, les séries de Fourier devinrent analyse harmonique, a été mathématiquement si riche que ce panorama historique nous fait rencontrer de nombreux théorèmes importants. Parmi ceux-ci figure le théorème de Szegö, suivi du théorème de Helson-Szegö, qui détermine pour quelles mesures μ sur le cercle l'opérateur de conjugaison est borné dans $L^2(\mu)$. Ce résultat classique a été utilisé récemment par N. Kalton et C. Le Merdy pour montrer l'existence d'un opérateur à puisances bornées sur l'espace de Hilbert qui n'est pas semblable à un opérateur de petite norme.

Les réseaux électriques étudiés dans la conférence d'Yves Colin de Verdière sont des graphes finis, où l'on distingue les sommets terminaux, ou bornes, et où les arètes ont une certaine conductance. L'application L qui fait correspondre aux potentiels aux N bornes l'intensité du courant qui en sort apparaît alors comme un opérateur symétrique d'un espace de dimension $(N-1)$ dans son dual. Cette application décrit donc la *réponse* du réseau aux stimuli électriques. Un premier théorème énonce la caractérisation des applications L qui sont la réponse d'un réseau plan dont les fils ne se coupent pas, donc d'un réseau planaire. L'application qui à un réseau associe sa réponse n'est pas injective, et un second théorème affirme que deux réseaux planaires qui ont même réponse peuvent être déduits l'un de l'autre par des transformations électriques simples, ce qui permet de montrer que l'équivalence électrique des réseaux planaires est un problème algorithmiquement décidable. La démonstration du second théorème repose sur des arguments simples mais subtils de géométrie plane.

La conférence de Frédéric Pham commence avec des éléments d'optique géométrique et quelques photographies de caustiques, qui le conduisent naturellement à parler d'enveloppes. Il nous montre comment les « enveloppes au but » observées peuvent être relevées en « enveloppes à la source » lisses dont elles sont les projections, ce qui ouvre la voie vers la théorie des catastrophes de Thom. Il revient ensuite à l'optique avec le principe de Huygens et les intégrales de Fresnel, dont la partie principale pour les grandes fréquences est donnée par le théorème de la phase stationnaire (un grand classique de l'oral de l'Agrégation...). Pour connaître la « vraie » fonction

d'onde et non son développement asymptotique aux grandes fréquences, il faut alors utiliser la variété lagrangienne, l'analyse complexe et les feuillets de Riemann, ce qui conduit jusqu'à la *résurgence* au sens défini par Écalle.

Pierre-Louis Lions commence par établir les équations d'Euler et de Navier-Stokes pour les fluides compressibles « barotropes », c'est-à-dire tels qu'il existe une relation simple entre la pression et la densité. Cette relation est non linéaire, ainsi que les équations qu'il démontre en utilisant la conservation de la masse et de la quantité de mouvement ; dans celle d'Euler, on suppose que le fluide est non visqueux. Les problèmes analytiques considérables (« légère difficulté », dans l'article original d'Euler) que posent la résolution de ces équations sont largement ouverts. Pierre-Louis Lions présente les obstructions (explosion des solutions en temps fini, oscillations persistantes dans le cas de Navier-Stokes, perte d'unicité), et les méthodes parfois inattendues qui donnent des éléments de solution, du moins dans le cas plus simple de la dimension 1 : critères entropiques, compacité faible, analyse harmonique et espaces de Hardy, phénomène de compensation « à la Stein-Fefferman ». Il expose également les liens entre ces questions théoriques et des problèmes très concrets comme celui du décollage d'Ariane V, et insiste sur l'aide que les mathématiciens peuvent et doivent apporter aux ingénieurs, bien avant d'avoir « tout compris » aux équations d'Euler.

Ce livre atteint parfaitement son but : faire apprécier à un public aussi large que possible, sans trop de technique mais sans complaisance, les « mathématiques d'aujourd'hui ». Les Éditions Cassini ont efficacement contribué au succès du projet : la présentation du livre est attrayante, et son format en rend le transport et la lecture facile. Je recommande très vivement aux collègues qui souhaitent élargir et approfondir leur point de vue de lire ces « Leçons ». Pourvu qu'ils soient d'humeur un peu studieuse, ils y trouveront le plus enrichissant et le plus agréable des devoirs de vacances.

— *Gilles Godefroy*