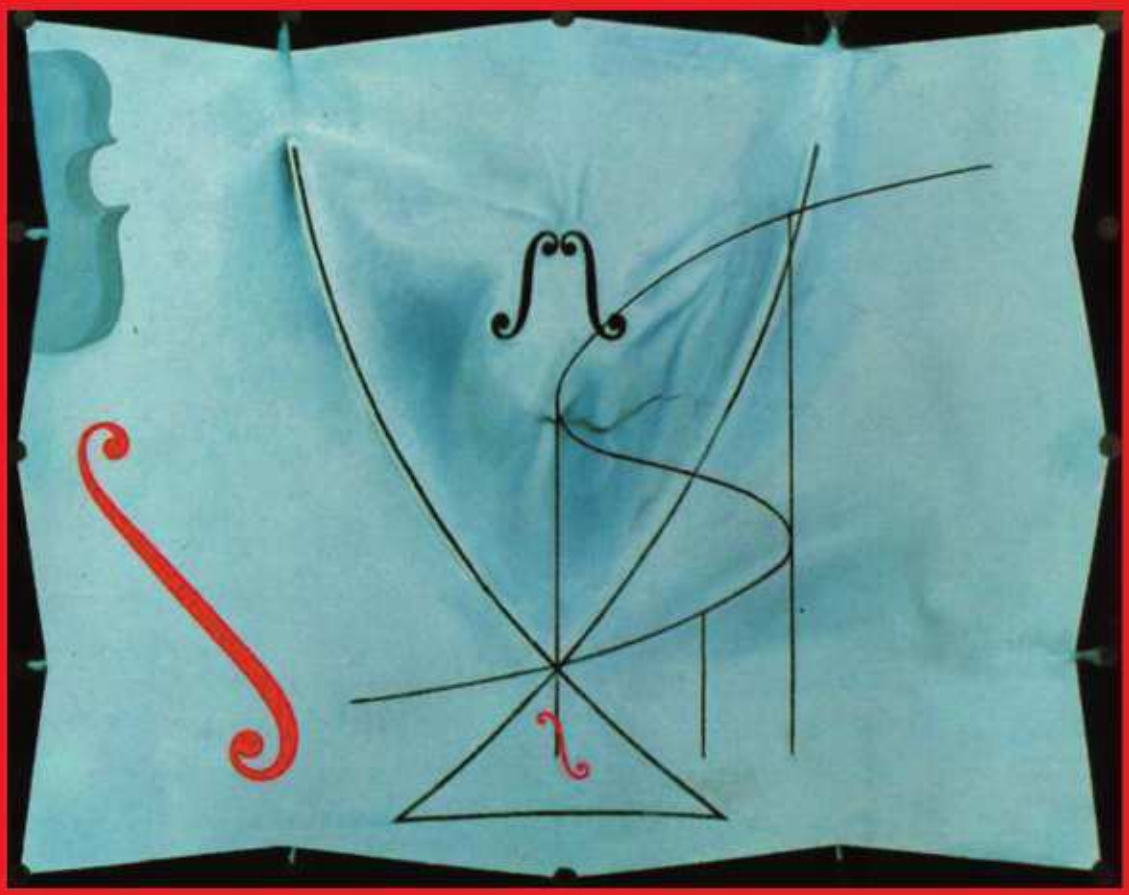
images des mathématiques 2006



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

La couverture de ce fascicule est illustrée par un tableau peint en 1983 par Salvatore DALI (1904-1989).
La « *queue d'aronde* » est la dernière toile de Dali, elle fait référence à l'une des sept « *catastrophes élémentaires* » définies par René Thom.

Réf. : Marc Chaperon, Une hirondelle en hiver,
Pour la Science **338**, déc. 2005, pp 150-151.

2006
IMAGES
DES MATHÉMATIQUES

Images des mathématiques 2006
Directeur de la publication : Michel Lannoo, directeur du département MPPU du CNRS
Rédacteurs en chef : Etienne Ghys, Jacques Istas
CNRS 3, rue Michel-Ange 75794 Paris cedex 16
Fabrication : Louis-Jean n° 660 **ISSN 0994-723 X. Dépôt légal : 269928T. Octobre 2006**

Introduction

Voici Images des Mathématiques 2006. Ce numéro rassemble des articles dont l'ambition est de faire connaître, de manière précise et attrayante, des mathématiques en train de se faire, à des lecteurs scientifiques, en particulier des étudiants en mathématiques. L'exercice est difficile et les auteurs s'y sont soumis avec brio !

Ce numéro est disponible sur le site Web du CNRS à l'adresse :

http://www.math.cnrs.fr/imagesdesmaths/IdM2006.htm

Rappelons que le précédent numéro d'Images des mathématiques 2004 est toujours disponible à l'adresse :

http://www.math.cnrs.fr/imagesdesmaths/IdM2004.htm

Ce numéro comporte 28 articles classés, suivant la tradition du milieu, selon l'ordre alphabétique des auteurs.

Parmi ces articles, un certain nombre reviennent sur l'œuvre de collègues éminents, et qui ont été à l'honneur récemment : H. Cartan à l'occasion de son centenaire, A. Connes avec la médaille d'or du CNRS, J.-P. Serre pour le prix Abel et tout récemment W. Werner, médaille Fields et premier probabiliste à recevoir cette distinction ; ou de grandes figures historiques récemment disparues, fondateurs de nouvelles branches des mathématiques : J.-L. Lions et R. Thom. On trouvera aussi dans ce numéro des articles historiques reliés à des mathématiques actuelles : une note sur L. Bachelier autour du mouvement brownien, et un texte sur H. Poincaré, à l'occasion de sa célèbre conjecture, qui constitue d'ailleurs le sujet d'un autre article.

Ce numéro témoigne de l'unité des mathématiques, de leur ouverture vers d'autres disciplines (citons dans le désordre économie, physique, mécanique céleste, informatique, imagerie, …) ainsi que du dynamisme des mathématiciens français, comme en témoigne la médaille Fields de W. Werner, après celle recue par L. Lafforgue en 2002.

Etienne Ghys et Jacques Istas

Les mathématiques vivent et sont en mouvement. Le cœur est solide et le corps se développe fortement et harmonieusement. C'est bien ce que les mathématiciens doivent comprendre et faire comprendre et c'est pourquoi Jacques Istas et Etienne Ghys ont accepté d'éditer ces « Images des Mathématiques 2006 ».

Le CNRS les remercie, ainsi que tous ceux qui ont contribué à cette réalisation.

Le département MPPU du CNRS.

Sommaire

L'inégalité de Brunn-Minkowski

Franck BARTHE*

Découverte en 1887, l'inégalité de Brunn-Minkowski sur le volume des sommes d'ensembles fut à l'origine d'une théorie toujours en développement qui étudie les propriétés volumiques fines des ensembles convexes. Cette inégalité ensembliste a par la suite eu des formulation fonctionnelles beaucoup plus souples qui ont élargi son domaine d'application, et permis de surprenantes connections avec les équations aux dérivées partielles, la théorie de l'information ou des probabilités.

Introduction

Une présentation simplifiée de la théorie de Brunn-Minkowski consiste à dire qu'elle étudie les relations entre addition des vecteurs et calcul des volumes d'ensembles. Commençons par quelques notations. Pour $\lambda \in \mathbb{R}$ et A un sous-ensemble de $\mathbb{R}^d$, on note $\lambda A = \{\lambda a;\ a \in A\}$. La somme de deux ensembles $A, B \subset \mathbb{R}^d$ est par définition

$$A + B = \{a + b;\ (a,b) \in A \times B\}.$$

L'inégalité de Brunn-Minkowski donne une minoration du volume d'une telle somme.

Théorème 1. *Soient A, B des sous-ensembles compacts non-vides de $\mathbb{R}^d$, alors*

$$\mathrm{Vol}_d(A+B)^{\frac{1}{d}} \geq \mathrm{Vol}_d(A)^{\frac{1}{d}} + \mathrm{Vol}_d(B)^{\frac{1}{d}}. \tag{1}$$

Si A et B sont des convexes homothétiques, il y a égalité. Brunn découvrit ce resultat en 1887 pour A, B convexes en dimension au plus 3. Minkowski démontra ensuite l'inégalité pour des convexes en dimension n et réalisa l'importance de l'énoncé. Il se combinait en effet avec un résultat antérieur de Steiner qui exprimait le volume du t-voisinage d'un convexe compact $A \subset \mathbb{R}^3$ défini pour $t > 0$ par

$$A_t = \{x \in \mathbb{R}^3;\ \exists y \in A,\ |x - y| \leq t\}.$$

On peut remarquer que $A_t = A + tB^3$ où B^d désigne la boule unité pour la norme euclidienne sur $\mathbb{R}^d$. La formule de Steiner assure que pour $t > 0$

$$\mathrm{Vol}_3(A + tB^3) = \mathrm{Vol}_3(A) + tS(A) + 2\pi t^2 W(A) + \frac{4}{3}\pi t^3,$$

* *Franck Barthe, barthe@math.ups-tlse.fr*
Laboratoire de Statistique et Probabilités, CNRS UMR C5583, Université Paul Sabatier, 31062 Toulouse Cedex 09.

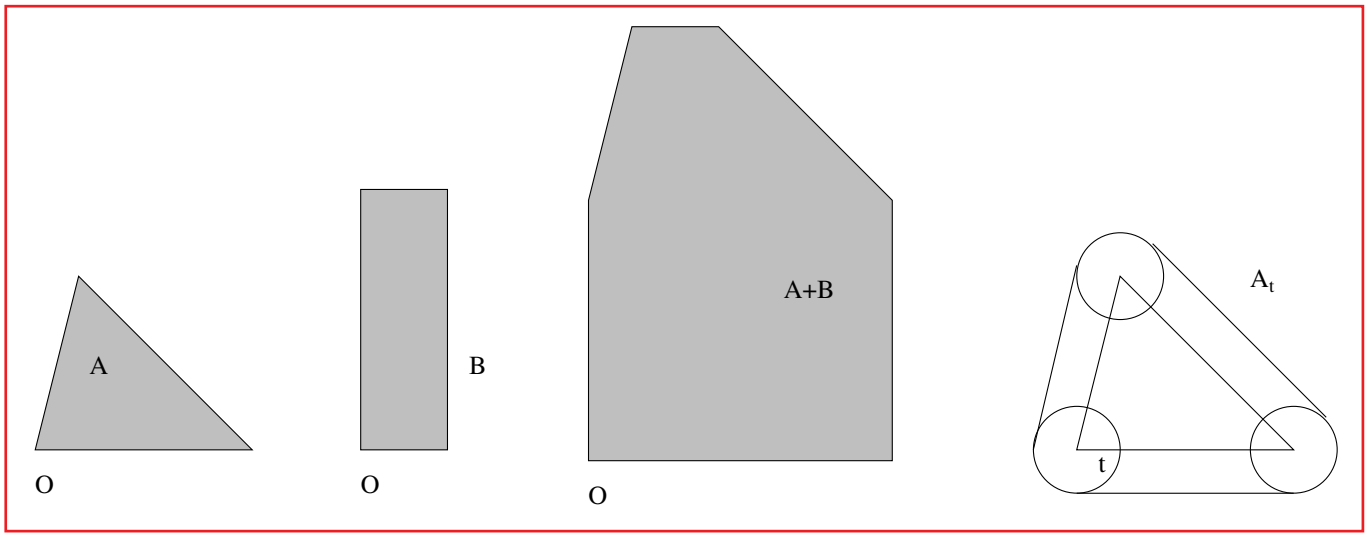


Figure 1

où $S(A)$ est la mesure de surface de A et $W(A)$ est son épaisseur moyenne (la moyenne sur les vecteurs unitaires u de la largeur de la bande minimale orthogonale à u et contenant A). Par le théorème de Brunn-Minkowski,

$$\mathrm{Vol}_3(A + tB^3) \geq \left(\mathrm{Vol}_3(A)^{\frac{1}{3}} + t\mathrm{Vol}_3(B^3)^{\frac{1}{3}}\right)^3$$

avec égalité pour $t = 0$. En comparant les dérivées en zéro des deux termes on obtient

$$S(A) \geq 3\mathrm{Vol}_3(B^3)^{\frac{1}{3}}\mathrm{Vol}_3(A)^{\frac{2}{3}}.$$

Il s'agit de l'inégalité isopérimétrique ; elle signifie que parmi les ensembles A de volume donné, les boules ont une surface minimale. Une autre conséquence directe de l'inégalité de Brunn-Minkowski est que la fonction

$$t \geq 0 \mapsto \mathrm{Vol}_3(A + tB^3)^{\frac{1}{3}}$$

est concave (ici l'hypothèse de la convexité de A est cruciale puisque l'on utilise $\lambda A + (1-\lambda)A = A$ pour $\lambda \in [0,1]$, ceci est faux en général !). La dérivée seconde en $t = 0$ est négative, on peut l'exprimer par la formule de Steiner et obtenir une nouvelle relation géométrique

$$S(A)^2 \geqslant 6\pi W(A)\mathrm{Vol}_3(A).$$

A partir de ces observations Minkowski developpa une théorie dont le point de départ est le suivant : pour $K_1,\ldots,K_m$ des ensembles convexes compacts de $\mathbb{R}^d$ et $\lambda_1,\ldots,\lambda_m \geq 0$, le volume de $\lambda_1 K_1 + \cdots + \lambda_m K_m$ est un polynôme homogène de la forme

$$\mathrm{Vol}_d(\lambda_1 K_1 + \cdots + \lambda_m K_m) = \sum_{i_1,\ldots,i_d=1}^{m} \lambda_{i_1} \ldots \lambda_{i_d} V(K_{i_1},\ldots,K_{i_d}).$$

Ici $V(K_1,\ldots,K_d)$ est par définition le volume mixte de d convexes de $\mathbb{R}^d$. L'étude des propriétés de ces quantités, de leurs interprétations géométriques et des inégalités qui les relient est toujours un domaine actif des mathématiques. Malgré des résultats nombreux et profonds, les volumes mixtes gardent une part de mystère. Citons pour

exemple de problème de Blaschke qui fait intervenir les volumes mixtes les plus simples. Il demande de décrire par un nombre fini d'inégalités l'ensemble des triplets $(\mathrm{Vol}_3(K),\ S(K),\ W(K))$ lorsque K décrit l'ensemble des convexes compacts de $\mathbb{R}^3$. Les relations que nous avons déduites du théorème de Brunn-Minkowski décrivent une partie du bord de cet ensemble, mais une portion de sa frontière est toujours manquante.

Inégalités fonctionnelles

S'il existe de nombreuses preuves géométriques de l'inégalité de Brunn-Minkowski, l'approche la plus féconde est probablement celle qui utilise une version fonctionnelle de l'énoncé :

Théorème 2 [Prékopa-Leindler]. *Soient f,g,h trois fonctions mesurables et positives sur $\mathbb{R}^d$ et soit $\lambda \in [0,1]$. Si pour tout x et tout y dans $\mathbb{R}^d$,*

$$h(\lambda x + (1-\lambda)y) \geq f^{\lambda}(x) g^{1-\lambda}(y),$$

alors

$$\int_{\mathbb{R}^d} h \geq \left(\int_{\mathbb{R}^d} f\right)^{\lambda} \left(\int_{\mathbb{R}^d} g\right)^{1-\lambda}.$$

Si l'on applique cet énoncé aux fonctions $f = \mathbf{1}_C$, $g = \mathbf{1}_D$, $h = \mathbf{1}_{\lambda C + (1-\lambda)D}$ pour des ensembles compacts $C, D \subset \mathbb{R}^d$, on obtient

$$\mathrm{Vol}_d(\lambda C + (1-\lambda)D) \geq \mathrm{Vol}_d(C)^{\lambda} \mathrm{Vol}_d(D)^{1-\lambda}.$$

Cette version multiplicative et sans dimension de l'inégalité de Brunn-Minkowski implique l'énoncé initial si l'on choisit $C = \mathrm{Vol}_d(A)^{-1/d} A$, $D = \mathrm{Vol}_d(B)^{-1/d} B$ et $\lambda = \mathrm{Vol}_d(A)^{1/d} / (\mathrm{Vol}_d(A)^{1/d} + \mathrm{Vol}_d(B)^{1/d})$. Notons que l'inégalité peut se récrire comme

$$\int_{\mathbb{R}^d}^{*} \sup_{\lambda x + (1-\lambda)y = z} f^{\lambda}(x) g^{1-\lambda}(y)\, dz \geq \left(\int_{\mathbb{R}^d} f\right)^{\lambda} \left(\int_{\mathbb{R}^d} g\right)^{1-\lambda},$$

avec une intégrale supérieure qui est une vraie intégrale si le supremum est mesurable. Elle apparaît ainsi comme une forme inverse de l'inégalité de Hölder

$$\int_{\mathbb{R}^d} f^{\lambda}(z) g^{1-\lambda}(z)\, dz \leq \left(\int_{\mathbb{R}^d} f\right)^{\lambda} \left(\int_{\mathbb{R}^d} g\right)^{1-\lambda}.$$

Dans ce qui suit nous présentons l'idée de deux démonstrations utilisant les équations aux dérivées partielles, plus précisément l'équation de Monge-Ampère et l'équation de la chaleur. Nous allons illustrer la première méthode directement pour l'inégalité de Brunn-Minkowski. La difficulté principale de ce résultat réside dans le fait que $A + B$ est un sous-ensemble de $\mathbb{R}^d$ pour lequel on a un paramétrage naturel par $A \times B \subset \mathbb{R}^{2d}$, inutilisable pour étudier le volume. Une maniére de construire un paramétrage par une ensemble de dimension d consiste à sélectionner pour chaque point $a \in A$ un point $\varphi(a) \in B$ en espérant que l'ensemble

$$\{a + \varphi(a);\ a \in A\} \subset A + B$$

a tout de même un grand volume. Une application φ convenable est donnée par un théorème de *transport optimal* de Brenier et McCann qui assure en particulier l'existence d'une fonction $\psi : A \to \mathbb{R}$ telle que son gradient $\nabla\psi$ soit une

application de A dans B, qui transporte la mesure de probabilité uniforme sur A vers la probabilité uniforme sur B. Plus précisément, pour tout borélien C,

$$\frac{\mathrm{Vol}_d(C \cap B)}{\mathrm{Vol}_d(B)} = \frac{\mathrm{Vol}_d((\nabla\psi)^{-1}(C) \cap A)}{\mathrm{Vol}_d(A)}.$$

L'application $\nabla\psi$ préserve donc le volume normalisé. Sous des hypothèses de régularité sur les convexes A, B, $\nabla\psi$ est bijective, différentiable, et cette propriété se traduit localement par le fait que sa différentielle multiplie les volumes par un facteur constant, c'est-à-dire

$$\det(\mathrm{Hess}\,\psi(x)) = \frac{\mathrm{Vol}_d(B)}{\mathrm{Vol}_d(A)}, \quad \forall x \in A.$$

Cet outil permet d'estimer le volume de l'ensemble somme :

$$\mathrm{Vol}_d(A+B) \geq \mathrm{Vol}_d\{a + \nabla\psi(a);\ a \in A\} = \int_A \det(I_d + \mathrm{Hess}\,\psi(x))dx,$$

où I_d est la matrice identité de $\mathbb{R}^d$. Comme le déterminant de la Hessienne de ψ est connu on peut minorer ceci en utilisant l'inégalité vérifiée par toutes les matrices symétriques positives $d \times d$, M et N

$$\det(M+N)^{\frac{1}{d}} \geq \det(M)^{\frac{1}{d}} + \det(N)^{\frac{1}{d}}.$$

Ainsi le transport de Brenier-McCann permet de déduire l'inégalité de Brunn-minkowski d'une inégalité similaire pour les matrices! Cette approche a donné des extensions puissantes des inégalités de Hölder et de Prékopa-Leindler, faisant intervenir cette fois des fonctions de $\mathbb{R}^d$ qui ne dépendent que de certaines directions. Elles donnent des résultats fins d'analyse sur les normes d'applications multilinéaires entre espaces de fonctions L_p, ainsi que des raffinements du théorème de Brunn-Minkowski pour des ensembles plats, dont le volume est nul mais dont la somme peut avoir un grand volume.

Un autre preuve surprenante de l'inégalité de Prékopa-Leindler repose sur une propriété de l'équation de la chaleur remarquée par Borell. Si trois fonctions positives continues et intégrables sur $\mathbb{R}^d$ satisfont

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}$$

alors pour tout $t > 0$, on a

$$P_t h(\lambda x + (1-\lambda)y) \geq (P_t f(x))^\lambda (P_t g(y))^{1-\lambda}$$

où $(P_t)_{t\geq 0}$ est le semigroupe de la chaleur défini comme suit : la fonction $u(t,x) = P_t f(x)$ est solution de l'équation de la chaleur

$$\partial_t u = \frac{1}{2}\Delta u$$

avec condition initiale $u(0,\cdot) = f$. Comme en temps grand on a

$$P_t h(z) \sim \frac{1}{(2\pi t)^{\frac{d}{2}}} \int_{\mathbb{R}^d} h,$$

on obtient l'inégalité de Prékopa-Leindler en faisant tendre le temps t vers l'infini.

Très récemment une inégalité de Brunn-Minkowski pour la mesure Gaussienne à été déduite d'une propriété analogue mais plus subtile de l'équation de la chaleur.

Théorème 3 [Ehrhard-Borell]. *Soit γ_d la mesure gaussienne standard sur $\mathbb{R}^d$,*

$$d\gamma_d(x) = e^{-|x|^2/2}\frac{dx}{(2\pi)^{d/2}}, \quad x \in \mathbb{R}^d.$$

Soient $A, B \subset \mathbb{R}^d$ des ensembles mesurables et $\lambda \in [0,1]$. On définit les demi-espaces $H_A = \{x \in \mathbb{R}^d, x_1 \leqslant a\}$ et $H_B = \{x \in \mathbb{R}^d, x_1 \leqslant b\}$ avec a, b choisis pour avoir $\gamma_d(A) = \gamma_d(H_A)$ et $\gamma_d(B) = \gamma_d(H_B)$. Alors

$$\gamma_d(\lambda A + (1-\lambda)B) \geqslant \gamma_d(\lambda H_A + (1-\lambda)H_B).$$

En d'autres termes, parmi les couples d'ensembles de probabilités données les demi-espaces parallèles ont une somme pondérée de mesure minimale. Ce résultat généralise de nombreuses inégalités très utiles en probabilités, dont l'inégalité de *concentration gaussienne* qui dit qu'au sens de la mesure gaussienne une fonction Lipschitzienne sur $\mathbb{R}^d$ dévie de sa moyenne avec une probabilité très faible.

Quelques perspectives

Pour finir nous citons brièvement d'autres domaines où la théorie de Brunn-Minkowski fait preuve d'une belle vitalité. La théorie de l'information initiée par Shannon fait intervenir l'entropie $S(X) = -\int f \log f$ d'un vecteur aléatoire $X \subset \mathbb{R}^d$ de loi à densité f. Ce nombre représente la quantité d'information de la variable X et admet une interprétation microscopique en terme de volume d'ensembles de suites de valeurs qui sont typiques pour des suites de copies indépendantes de X. L'inégalité de Shannon-Stam assure que si X et Y sont des vecteurs aléatoires à densité sur $\mathbb{R}^d$, indépendants, alors

$$e^{\frac{2}{d}S(X+Y)} \geq e^{\frac{2}{d}S(X)} + e^{\frac{2}{d}S(Y)}.$$

L'interprétation volumique de l'entropie, la ressemblance avec l'inégalité de Brunn-Minkowski, mais aussi l'exposant $2/d$ (au lieu de $1/d$) ont beaucoup intrigué les spécialistes du sujet. Cette enigme a été élucidée récemment par Szarek et Voiculescu. L'idée principale est que l'ensemble des suites typiques pour $X+Y$ ne contient pas toutes les sommes de suites typiques de X et de Y, mais presque toutes (si X et Y sont indépendantes et de même loi, et si $(x_1,\ldots,x_n)$ est une suite de valeurs typique pour X, elle l'est pour Y, mais la somme des deux $(2x_1,\ldots,2x_n)$ semble typique de $2X$ et non de $X+Y$). Il existe une version restreinte de l'inégalité de Brunn-Minkowski, qui minore le volume de sommes incomplètes, avec des exposants $2/d$. D'autres connections avec l'entropie existent.

Certaines équations aux dérivés partielles mettent en jeu des fonctions F définies sur les ensembles convexes de $\mathbb{R}^d$ qui sont α-homogènes et satisfont une inégalité de type Brunn-Minkowski : pour tous convexes compacts A, B

$$F(A+B)^{\frac{1}{\alpha}} \geq F(A)^{\frac{1}{\alpha}} + F(B)^{\frac{1}{\alpha}}.$$

Ceci est vérifié si $F(A)$ est la première valeur propre du Laplacien sur A avec conditions de Dirichlet, ou si $F(A)$ est la capacité électrostatique de A. Les preuves de ces résultats présentent de nombreuses similarités, elle reposent souvent en partie sur l'inégalité de Prékopa-Leindler. Il semble qu'une approche unifiée devrait finir par émerger, peut-être en utilisant les interprétations browniennes des solutions d'équations aux dérivées partielles associées.

Il existe bien d'autres variations sur les inégalités de Brunn-Minkowski, comme des inégalités pour les variétés Riemanniennes, des versions arithmétiques, des correspondances avec la géométrie algébrique. Le lecteur trouvera dans les livres et les articles de synthèse cités en référence des éléments permettant de s'orienter dans une bibliographie passionnante et fournie.

Pour en savoir plus

[Ba] Barthe (F.), Autour de l'inégalité de Brunn-Minkowski, *Ann. Fac. Sci. Toulouse*, **12**(2), 127-178 (2003).

[Bo] Borell (C.), Minkowski sums in Gaussian analysis, Notes de l'école d'hiver « Probabilistic Methods in High Dimension Phenomena », http ://www.lsp.ups-tlse.fr/Proba_Winter_School/ (2005).

[G] Gardner (R. J.), The Brunn-Minkowski inequality, *Bull. Amer. Math. Soc. (N.S.)*, **3**, 355-405 (2002).

[L] Ledoux (M.), The concentration of measure phenomenon, *Mathematical Surveys and Monographs*, volume 89, American Mathematical Society, Providence, RI (2001).

[S] Schneider (R.), Convex bodies : the Brunn-Minkowski theory, *Encyclopedia of Mathematics and its Applications*, volume 44, Cambridge University Press, Cambridge (1993).

[V] Villani (C.), Topics in optimal transportation, *Graduate Studies in Mathematics*, volume 58, American Mathematical Society, Providence, RI (2003).

Un *joli* problème d'Erdös

Michel BENAÏM*

Soit $0 < a < 1$ et $Y^a = \sum_{k=0}^{\infty} a^k \theta_k$ la série aléatoire obtenue en choisissant les $\theta_k \in \{-1,1\}$ au hasard. L'étude de la fonction de répartition $F^a(t) = \mathcal{P}(Y^a \leqslant t)$ révèle quelques connexions étonnantes entre la théorie des chaînes de Markov, l'analyse harmonique, la théorie des nombres algébriques, et la théorie géométrique de la mesure.

Introduction

Si la définition d'un *problème de mathématiques* est relativement consensuelle, celle d'un *joli problème de mathématiques* est nécessairement beaucoup plus subjective. Pourtant je ne doute pas que les lecteurs de cet article s'accorderont avec moi à trouver que le problème que les éditeurs des images 2005 m'ont demandé de présenter ici possède tous les ingrédients d'un *joli* problème de mathématiques :

(i) Il est très facile à décrire ;

(ii) Il n'est toujours pas (complétement) résolu ;

(iii) Son analyse fait appel à des outils empruntés à des domaines particulièrement variés des mathématiques.

Soit $\Omega = \{-1,1\}^{\mathbb{N}}$ et $a \in]0,1[$. Pour toute suite de signes $\theta \in \Omega$, nous notons

$$Y_n^a(\theta) = \sum_{k=0}^{n} a^k \theta_k \quad \text{et} \quad Y^a(\theta) = \sum_{k=0}^{\infty} a^k \theta_k.$$

L'objet qui va nous intéresser ici est la fonction, ou comme disent les probabilistes, *la variable aléatoire* $Y^a : \Omega \mapsto \mathbb{R}$ lorsque les θ_n sont choisis au hasard en jouant à pile ou face. Plus précisément, munissons Ω de la probabilité $\mathcal{P}$ du jeu de pile ou face :

$$\forall \omega \in \Omega, \forall n \in \mathbb{N}, \quad \mathcal{P}(\{\theta \in \Omega : \theta_0 = \omega_0, \ldots, \theta_{n-1} = \omega_{n-1}\}) = \frac{1}{2^n}.$$

La *loi* de Y^a est la probabilité ν^a, image de $\mathcal{P}$ par Y^a :

$$\nu^a(B) = \mathcal{P}(\{\theta \in \Omega : Y^a(\theta) \in B\}),$$

et sa *fonction de répartition* est l'application $F^a : \mathbb{R} \to [0,1]$ définie par

$$F^a(t) = \nu^a(]-\infty, t]) = \mathcal{P}(Y^a \leqslant t).$$

* *Institut de Mathématiques, Rue E-Argand 11, CH-2007 Neuchâtel Suisse.*

L'étude de la fonction F^a (ou de la mesure ν^a) a été initiée dans les années 30 par Wintner et ses collaborateurs et révèle des connexions remarquables entre l'analyse harmonique, la théorie des nombres algébriques, les systèmes dynamiques ou encore la théorie géométrique de la mesure.

Auto-similarité

Soit (X_n) la *suite récurrente aléatoire* définie par

$$X_{n+1} = aX_n + \theta_{n+1}, n \geqslant -1, \tag{1}$$

où les θ_n sont tirés à pile ou face et X_{-1} est une condition initiale arbitraire indépendante des θ_n. Pour tout $n \geqslant 0$,

$$X_n(\theta) = a^{n+1}X_{-1} + \sum_{k=0}^{n} a^k \theta_{n-k},$$

de sorte que les variables X_n et $a^{n+1}X_{-1} + Y_n^a$ ont la même loi car les vecteurs $(\theta_0,\ldots,\theta_n)$ et $(\theta_n,\ldots,\theta_0)$ ont la même loi. Ainsi, si ν_n désigne la loi de X_n et Φ une fonction « test » continue bornée,

$$\int_{\mathbb{R}} \Phi \, d\nu_n = \int_{\Omega} \Phi \circ X_n \, d\mathcal{P} = \int_{\Omega} \Phi \circ (Y_n^a + a^{n+1}X_{-1}) \, d\mathcal{P} \to \int_{\Omega} \Phi \circ Y^a \, d\mathcal{P} = \int_{\mathbb{R}} \Phi \, d\nu^a \tag{2}$$

quand $n \to \infty$. Par ailleurs, d'après (1), la suite (ν_n) vérifie la relation de récurrence

$$\nu_{n+1} = T^a(\nu_n) \tag{3}$$

où T^a est l'operateur qui transforme une mesure ν de fonction de répartion F (*i.e* $F(t) = \nu(]-\infty,t])$) en une mesure $T^a(\nu)$ de fonction de répartition

$$T_a(F)(t) = \frac{1}{2}\left[F\left(\frac{t-1}{a}\right) + F\left(\frac{t+1}{a}\right)\right]. \tag{4}$$

De (2) et (3) on déduit la propriété *d'autosimilarité* suivante.

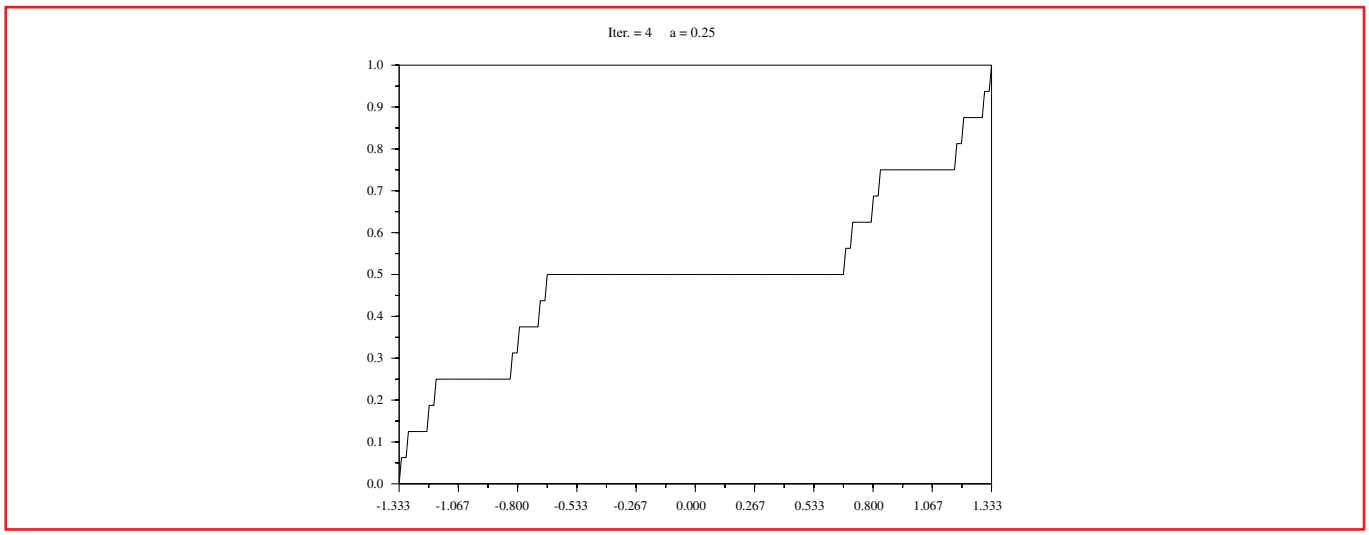


Figure 1 - *Graphe de F_4 pour $a = 0{,}25$.*

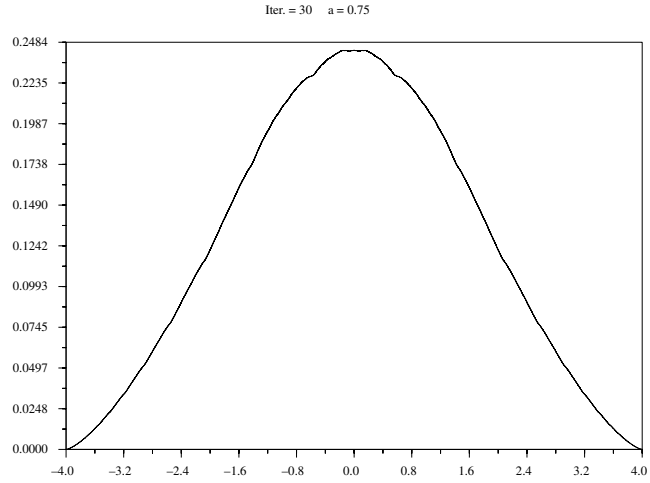


Figure 2 - *Graphe de F_{30} pour $a = 0{,}75$.*

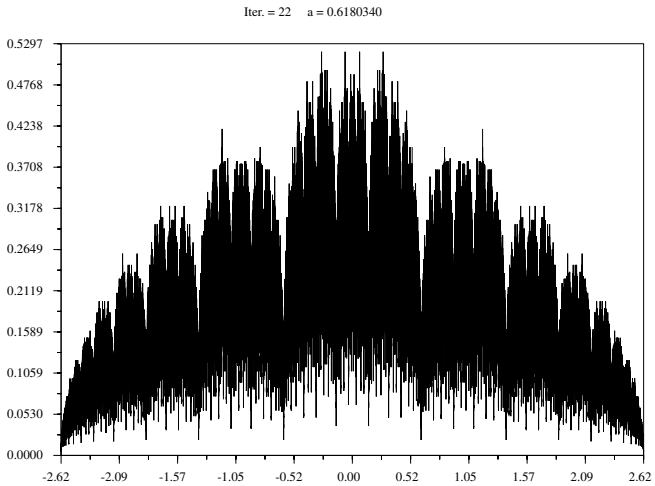


Figure 3 - *Graphe de f_{22} pour $a = 0{,}6180340$.*

Proposition 1. *ν^a est l'unique probabilité solution de $\nu = T^a(\nu)$. C'est la probabilité invariante de (1).*

Simulations

La fonction de répartition F_n de ν_n vérifie $F_n = T_a(F_{n-1})$. Si, par ailleurs, ν_{-1} admet une densité, alors ν_n admet une densité f_n et $f_n = \frac{1}{a} T_a(f_{n-1})$. L'itération de ces formules est à la base des programmes informatiques réalisés par Michel Delasnerie pour représenter le graphe des fonctions F_n et f_n. Les figures données ici sont obtenues pour différentes valeurs de a et différentes valeurs de n lorsque X_{-1} suit une loi uniforme sur $[0,1]$. Pour en comprendre la diversité il faut lire les sections suivantes.

Loi des types purs

Rappelons qu'une mesure ν sur $\mathbb{R}$ de fonction de répartition $F(t) = \nu(]-\infty,t])$ est dite *absolument continue*, si $F(t) = \int_{-\infty}^{t} f(u)du$ où f est une fonction positive intégrable (appelée densité) ; et *singulière* si il existe un ensemble $N \subset \mathbb{R}$ de mesure de Lebesgue nulle et tel que $\nu(\mathbb{R} \setminus N) = 0$. En général, une mesure ν sur $\mathbb{R}$ peut toujours s'écrire de manière unique sous la forme $\nu = \nu_{ac} + \nu_s$ où ν_{ac} est absolument continue et ν_s est singulière (éventuellement nulle).

Dans les années 30, le mathématicien Wintner et ses collaborateurs ont cherché à décrire le type de ν^a. Jessen et Wintner démontrent en 1935 le résultat suivant.

Proposition 2 [Jessen et Wintner, 1935]. *La fonction F_a est continue. De plus, la mesure ν^a est absolument continue* **ou bien** *continue singulière.*

La seconde partie de cette proposition est une conséquence directe de la proposition 1. En effet, l'opérateur T^a laissant invariant les mesures absolument continues et les mesures singulières, les parties ν^a_{ac} et ν^a_s sont des points fixes de T^a, et par unicité, une de ces partie doit être nulle.

Les escaliers du diable : $a < \frac{1}{2}$

Pour $a < \frac{1}{2}$, $Y_n(\theta)$ prend au plus 2^{n+1} valeurs distinctes et le reste $r_n = Y^a(\theta) - Y^a_n(\theta)$ a un module $\leqslant \frac{a^{n+1}}{a-1}$. La mesure ν^a est donc concentrée sur une réunion d'intervalles dont la longueur totale $l_n = 2^{n+1}a^{n+1}/(1-a)$. tend vers zéro quand $n \to \infty$. Ainsi,

Proposition 3 [Kershner et Wintner, 1935]. *Pour $a < \frac{1}{2}$, ν^a est continue singulière. Plus précisément, le support de ν^a est un Cantor de mesure de Lebesgue nulle.*

La fonction F^a est donc un *escalier du diable* (voir l'encadré 1 et la figure 1).

Encadré 1

Escalier du diable

Un escalier du diable est une fonction continue croissante mais presque partout constante. Ce type de fonction a été découvert en 1885 par Ludwig Scheefer élève de Georg Cantor.

Les théorèmes d'Erdös et Solomyak : $a \geqslant \frac{1}{2}$

Pour $a = \frac{1}{2}$, il est facile de vérifier (en utilisant par exemple la proposition 1) que $\nu_{1/2}$ est la loi uniforme sur $I_{1/2}$. Pour $a \geqslant \frac{1}{2}$ le support de ν^a coincïde avec l'intervalle $I_a = \left[-\frac{1}{1-a}, \frac{1}{1-a}\right]$ et F^a est donc une fonction continue strictement croissante sur I^a. Par ailleurs, Wintner en 1935 prouve que F^a est dérivable de classe C^{k-1} pour $a = 2^{-1/k}$ et $k \geqslant 2$. Ces résultats pourraient laisser supposer que F^a est absolument continue pour $a \geqslant \frac{1}{2}$, mais Erdös en 1939 prouve le surprenant résultat suivant.

Théorème 1 [Erdös, 1939]. Soit $1 < g < 2$ un nombre de Pisot (voir encadré 2) et $a = 1/g$. Alors ν^a est singulière.

La figure 3 « illustre » ce théorème lorsque a est l'inverse du nombre d'or.

Encadré 2

Nombre de Pisot

Un nombre de Pisot g est un entier algébrique, c'est-à-dire la racine d'un polynôme unitaire à coefficients dans $\mathbb{Z}$, dont les conjugués (i.e les autres racines) $g_2,\ldots,g_m$ sont de module < 1. Par exemple, le nombre d'or $g = (\sqrt{5}+1)/2$ est un nombre de Pisot car c'est la racine du polynôme $X^2 - X - 1$ dont le conjugué $(1-\sqrt{5})/2$ a un module < 1.
De même les racines positives $\theta_1 \sim 1.32471$ de $X^3 - X - 1$ et $\theta_2 \sim 1.3802777$ de $X^4 - X^3 - 1$ sont des nombres de Pisot (ce sont d'ailleurs les seuls nombres de Pisot contenus dans $]1,\sqrt{2}[$).
Une propriété remarquable des nombres de Pisot est que leurs puissances s'approchent exponentiellement vite des entiers. En effet, l'expression $g^n + g_2^n + \ldots + g_m^n$ est un polynôme symétrique des racines et peut donc s'écrire comme un polynôme à coefficients dans $\mathbb{Z}$, des fonctions symétriques élémentaires. Ainsi $g^n + g_2^n + \ldots + g_m^n \in \mathbb{Z}$ et

$$dist(g^n,\mathbb{Z}) \leqslant \rho^n \qquad (5)$$

pour un certain $0 < \rho < 1$.

Encadré 3

La preuve du théorème d'Erdös

Soit

$$\Phi_t(Y^a) = \mathcal{E}(e^{itY^a}) = \int_{\mathbb{R}} e^{itx}\nu^a(dx) = \prod_{k=0}^{\infty}\frac{1}{2}(e^{ita^k} + e^{-ita^k})) = \prod_{k=0}^{\infty}\cos(ta^k)$$

la fonction caractéristique de Y^a (i.e la transformée de Fourier de ν^a). La preuve du théorème d'Erdös consiste à montrer que $\Phi_t(Y^a)$ ne converge pas vers 0 quand $t \to \infty$, car d'après le lemme de Riemann-Lebesgue, la transformée de Fourier d'une mesure absolument continue tend vers zéro en l'infini. Pour $a = 1/g$ et $t(n) = \pi g^n$,

$$\Phi_{t(n)}(Y^a) = \prod_{k=0}^{\infty}\cos(g^{n-k}\pi) = \prod_{k=0}^{n}\cos(g^k\pi)C, \quad \text{où } C = \prod_{k=1}^{\infty}\cos(a^k\pi).$$

D'après (5) $||\cos(g^k\pi)| - 1| \leqslant \pi\rho^k$. D'où

$$\lim_{n\to\infty}|\Phi_{t(n)}(Y^a)| = |C|\prod_{k=0}^{\infty}|\cos(g^k\pi)| > 0.$$

Le problème d'Erdös et le théorème de Solomyak

Soit $\mathcal{S} \subset]\frac{1}{2},1[$ l'ensemble des valeurs de a pour lesquelles ν^a est singulière. Le théorème d'Erdös pose le problème de la description de $\mathcal{S}$. En 1940 Erdös prouve que $a \notin \mathcal{S}$ pour presque tout $a \in]a^*,1[$ pour un certain a^* proche de 1. Kahane et Salem en 1958 obtiennent un critère assurant que ν^a possède une densité L^2. En 1962

Garsia prouve que $a \notin \mathcal{S}$ si a est l'inverse d'un entier algébrique (voir encadré 2) dont les conjugués ont un module > 1 et tel que le terme constant du polynome unitaire soit 2 ou -2. Un tel polynome étant par exemple $x^{n+p} - x^n - 2$ avec $p, n \geqslant 1$ et $\max p, n \geqslant 2$. Dans le même travail, Garsia conjecture que S est de mesure nulle, mais il faudra attendre 1995 pour que cette conjecture soit prouvée par Solomyak.

Théorème 2 [Solomyak, 1995]. *$\mathcal{S}$ est de mesure de Lebesgue nulle.*

A ce jour, on ne sait toujours pas si $\mathcal{S}$ contient des points qui ne sont pas des inverses de nombres de Pisot.

Encadré 4

La preuve du théorème de Solomyak

En 1996, Y. Peres et B. Solomyak ont proposé une preuve remarquablement courte du théorème de Solomyak dont voici l'idée générale. Soit

$$D^a(x) = \liminf_{r\to 0} \frac{1}{2r}\nu^a(]x-r,x+r[).$$

Une application du théorème de recouvrement de Vitali montre que $a \notin \mathcal{S}$ si et seulement si $D^a(x) < \infty$ pour ν^a presque tout x. Il suffit donc de prouver que $s = \int_\alpha^\beta \int_{\mathbb{R}} D^a(x)\nu^a(dx)da < \infty$ pour en déduire que $\lambda(S \cap]\alpha,\beta[) = 0$ où λ désigne la mesure de Lebesgue. Par le lemme de Fatou, le théorème de la mesure image et le théorème de Fubini,

$$s \leqslant \liminf_{r\to 0} \frac{1}{2r}\int_\alpha^\beta \int_{\mathbb{R}} \nu^a(]x-r,x+r[\nu^a(dx)da = \liminf_{r\to 0} \frac{1}{2r}\int_\alpha^\beta \mathcal{P}\otimes\mathcal{P}((\theta,\tilde\theta)\in\Omega^2 : |Y^a(\theta)-Y^a(\tilde\theta)|<r)da$$

$$= \liminf_{r\to 0} \frac{1}{2r}\int_{\Omega^2} \lambda(a\in]\alpha,\beta[: |Y^a(\theta)-Y^a(\tilde\theta)|<r)\, d\mathrm{P}(\theta)d\mathrm{P}(\tilde\theta).$$

Le terme $|Y^a(\theta) - Y^a(\tilde\theta)|$ peut s'écrire sous la forme $2a^k|g(a)|$ où $k = |\theta \wedge \tilde\theta| = \min\{n : \theta_n \neq \tilde\theta_n\}$ et g est de la forme

$$g(a) = 1 + \sum_{n\geqslant 1} b_n a^n \qquad (6)$$

avec $b_n \in \{-1,0,1\}$.

Supposons que pour un certain $\delta > 0$, toute fonction de la forme (6) vérifie la propriété, dite de δ-transversalité, suivante : $\forall a \in]\alpha,\beta[|g(a)| \leqslant \delta \Rightarrow g'(a) \leqslant -\delta$. Alors $\lambda(a\in]\alpha,\beta[: |g(a)| \leqslant r) \leqslant 2\delta^{-1}r$. Cette inégalité est tautologie pour $2r \geqslant \delta$ et une conséquence de la δ-transversalité pour $2r < \delta$. Ainsi, $\lambda(a\in]\alpha,\beta[: |Y^a(\theta) - Y^a(\tilde\theta)| < r) \leqslant r(\delta\alpha^{\theta\wedge\tilde\theta})^{-1}$. D'où

$$s \leqslant (2\delta)^{-1}\int_{\Omega^2} (\alpha^{\theta\wedge\tilde\theta})^{-1} d\mathcal{P}(\theta)d\mathcal{P}(\tilde\theta) = (2\delta)^{-1}\sum_k \alpha^{-k}\mathcal{P}\otimes\mathcal{P}(\theta\wedge\tilde\theta = k) = (2\delta)^{-1}\sum_k (2\alpha)^{-k} < \infty$$

car $2\alpha < 1$ et, donc $\lambda(S\cap]\alpha,\beta[) = 0$.

L'étape suivante de la preuve de Peres et Solomyak consiste à prouver la propriété de δ-transversalité sur l'intervalle $[2^{-1},2^{-2/3}]$, puis, par un argument similaire, à couvrir l'intevalle $[2^{-2/3},2^{-1/2}]$. Pour conclure, il suffit de remarquer que, ν_a etant la convolution de ν_{a^2} avec une autre mesure, l'absolue continuité de ν_{a^2} entraine celle de ν_a. Ce qui permet de couvrir tout l'intervalle $[2^{-1},1[$ à partir de $[2^{-1},2^{-1/2}]$.

Pour en savoir plus

Ce texte est inspiré du chapitre 1 du livre de

BENAÏM (M.), EL KAROUI (N.), *Promenade Aléatoire*, éditions de l'école polytechnique, ISBN 2-7302-1168-3 (2004).

Pour plus de détails, de nombreux développements et d'autres références le lecteur intéressé pourra consulter l'article de revue de

PERES (Y.), SCHLAG (W.), SOLOMYAK (B.), « Sixty years of Bernoulli convolutions, » in *fractal geometry and stochastics II*, **46**, 39-65, Prog. Prob. Birkhauser, Basel, (2000).

Remerciements

Je remercie Michel Delasnerie qui a réalisé les programmes à l'origine des figures ainsi que Francois Ledrappier et Christophe Leuridan pour d'utiles commentaires.

La preuve de la conjecture de Poincaré d'après G. Perelman

L. BESSIÈRES *, G. BESSON *, M. BOILEAU **

Comment une conjecture, *a priori* purement topologique, résiste 100 ans aux topologues pour se livrer aux géomètres. Le programme lancé par Richard Hamilton en 1982 et mené à son terme par Grigori Perelman en 2003, repose sur le flot de la courbure de Ricci, une équation d'évolution qui tend à homogénéiser la métrique.

Introduction

La topologie des surfaces est bien comprise dès la fin du 19-ième siècle. Toute surface orientée et sans bord peut être décrite topologiquement comme le bord d'un bretzel. Le nombre de trous du bretzel est aussi le nombre maximal de courbes disjointes tracées sur la surface sans la séparer : ce nombre, appelé genre, suffit à classer les surfaces. Du point de vue topologique la surface la plus simple est donc la sphère $\mathbf{S}^2 \subset \mathbf{R}^3$ qui est le bord de la boule unité et sur laquelle toute courbe fermée sépare.

Une étude similaire pour les hypersurfaces (ou variétés) sans bord de dimension supérieure n'a vraiment commencé qu'avec le mémoire de Henri Poincaré sur l'Analysis Situs en 1895 ([Poin]), qui marque la naissance de la topologie algébrique moderne.

En 1904, dans le cinquième et dernier complément à l'Analysis Situs ([Poin]), Poincaré construit un exemple qui montre qu'en dimension trois on ne peut pas caractériser la sphère unité $\mathbf{S}^3 \subset \mathbf{R}^4$ par la propriété que toute surface plongée la sépare. Il faut faire appel à des notions topologiques plus fines. Pour distinguer l'espace tridimensionnel qu'il a construit de la sphère $\mathbf{S}^3$ Poincaré utilise le *groupe fondamental*. C'est un invariant algébrique qu'il a introduit dans son premier mémoire sur l'Analysis Situs et qui prend en compte les chemins fermés (lacets) dans l'espace considéré qui ne peuvent pas être rétractés sur un point par une déformation continue (on dit alors que le lacet est *essentiel*). Si aucun lacet n'est essentiel, le groupe fondamental est trivial et l'espace est dit *simplement connexe*[1]. C'est le cas de toutes les sphères $\mathbf{S}^n$ de dimension $n \geqslant 2$. Poincaré montre que son exemple n'est pas simplement connexe.

* *Université de Grenoble, Institut Fourier UMR CNRS 5582, 100 rue des maths, BP74, 38402 Saint Martin d'Hères Cedex (France), G.Besson@ujf-grenoble.fr (France), Laurent.Bessieres@ujf-grenoble.fr*

** *Université de Toulouse 3, Paul Sabatier, Laboratoire Emile Picard, UMR CNRS 5580, 118 route de Narbonne, 31062 Toulouse Cedex 4 (France), boileau@picard.ups-tlse.fr*

[1] Il s'agit de la terminologie actuelle. Poincaré utilisait le mot *simplement connexe* pour désigner une sphère.

A la fin de son article il pose la question suivante, désormais célèbre : « *Est-il possible que le groupe fondamental d'une variété V de dimension* 3 *se réduise à la substitution identique, et que pourtant V ne soit pas la sphère*[2] ? ».

L'affirmation qu'une variété simplement connexe de dimension 3 est la sphère $\mathbf{S}^3$ est connue sous le nom de *Conjecture de Poincaré*. Cette conjecture s'est révélée être un problème extrêmement difficile. Elle trouve une généralisation naturelle dans la *Conjecture de Géométrisation* formulée par William Thurston dans les années 70 pour décrire toutes les variétés de dimension 3.

Thurston ([Thu]) conjecture que huit géométries homogènes suffisent pour décrire les briques élémentaires permettant de construire toutes les variétés de dimension 3 (voir aussi ([Sco]). La conjecture de géométrisation a ainsi replacé la géométrie différentielle au cœur de l'étude des variétés de dimension 3.

Au début des années quatre-vingts Richard Hamilton a lancé un nouveau programme pour démontrer la conjecture de géométrisation et, en particulier, la conjecture de Poincaré. Son approche est basée sur le flot de Ricci : il s'agit d'analyser et de contrôler les solutions d'une équation différentielle, liée à la courbure (le flot de Ricci), sur l'espace des métriques riemanniennes de la variété considérée. Lorsque le flot évolue la métrique s'homogénéise, mais l'étalement de la courbure n'est pas uniforme : en temps fini elle peut s'accumuler et devenir infinie (nous dirons « exploser ») en certains points de la variété. Ce sont ces phénomènes, appelés singularités du flot de Ricci, que Hamilton n'a pas réussi à contrôler.

Récemment, Grigori Perelman a défini pour le flot de Ricci une quantité monotone, appelée *entropie*, qui lui a permis de décrire la manière dont apparaissent les singularités et de les classer. Avec cela il a réussi le tour de force de mener à son terme le programme de Hamilton. Il construit à partir du flot de Ricci un *flot avec chirurgie* permettant de se débarrasser des singularités. Nous décrivons maintenant la preuve de la conjecture de Poincaré proposée par Perelman.

Le flot associé à la courbure de Ricci

On cherche un procédé évolutif qui produise une métrique riemannienne privilégiée sur une variété différentielle M donnée (voir encadré 1). Le souhait est que celle-ci soit une métrique de courbure de Ricci constante ; une telle métrique est dite d'Einstein.

Le flot associé à la courbure de Ricci (voir encadré 1) est une équation différentielle sur l'espace (de dimension infinie) $\mathcal{M}$ des métriques riemanniennes sur la variété M. L'idéal serait que cette équation différentielle ordinaire soit donnée par l'opposé du gradient d'une fonction (de sorte que les trajectoires convergent vers les minima) ; le candidat naturel pour une telle fonction est celle que les physiciens appellent la fonctionnelle de Hilbert-Einstein ; il s'agit d'une intégrale de courbure (la courbure scalaire) dont les points critiques sont les métriques dites d'Einstein, c'est-à-dire les métriques telles que $\mathrm{Ric}_g = \lambda g$. Malheureusement un calcul simple montre que son gradient donne naissance à une équation qui n'admet pas de solutions en général. Par contre une modification de cette équation stérile convient. On appelle flot de Ricci, une famille $g(t)$ de métriques riemanniennes sur M, définie sur un intervalle $[0, T[$ et qui vérifie l'équation d'évolution suivante (voir encadré 2) :

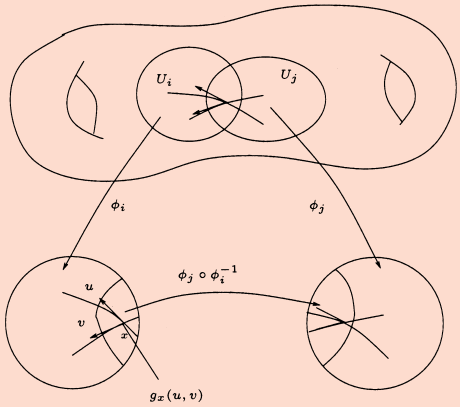
$$\frac{\partial g}{\partial t} = -2\mathrm{Ric}_{g(t)}\,. \tag{1}$$

Encadré 1

Définition 1

Une variété différentielle de dimension n est un espace localement modelé sur l'espace euclidien standard $\mathbf{R}^n$ *: chaque point a un voisinage ouvert homéomorphe à* $\mathbf{R}^n$ *qu'on appelle une carte et le passage d'une carte à une autre s'opère par un difféomorphisme de classe* C^∞*. En dimension 3 la structure différentielle est unique à difféomorphisme près alors qu'on peut munir* $\mathbf{R}^4$ *d'une infinité de structures différentielles deux-à-deux non difféomorphes. Dans la suite toutes les variétés seront de classe* C^∞ *et orientables.*

[2] Poincaré utilise ici le mot *simplement connexe.*



Pour faire de la géométrie dans M (par exemple calculer la longueur d'une courbe, des distances, des volumes, etc.) on a besoin d'une structure supplémentaire, appelée ***métrique riemannienne*** *et notée g : c'est la donnée en chaque point x des cartes d'un produit scalaire $g_i(x)$, variant de façon C^∞ avec x et compatibles avec les changements de cartes. Les courbures associées à une métrique mesurent l'écart infinitésimal à la métrique standard de $\mathbf{R}^n$. Elles se calculent par des expressions polynomiales des coefficients de g, ∂g et $\partial^2 g$.*

Par exemple, à chaque 2-plan P de l'espace tangent T_xM, on associe la courbure sectionnelle $K(P)$, qu'on peut définir comme suit. On appelle $C(r)$ le cercle de centre x, de rayon r tangent à P. Alors sa longueur satisfait la formule

$$\ell(C(r)) = 2\pi r(1 - \frac{K(P)}{6}r^2 + o(r^2)),$$

et $K(P)$ mesure le défaut au périmètre euclidien. La courbure de Ricci (on dit aussi le tenseur de Ricci) est, en chaque point x de M, une forme bilinéaire symétrique sur T_xM (pas nécessairement définie positive). Sa valeur dans une direction $v \in T_xM$, qu'on note $\mathrm{Ric}_g(v,v)_x$, se calcule en prenant la somme des courbures sectionnelles des 2-plans engendrés par v et e_i où e_i parcourt une base orthonormée de l'orthogonal de v dans T_xM. Elle mesure un défaut dans l'aire de petites sphères. La ***courbure scalaire*** *$R(x)$ est une fonction sur M définie en chaque point x comme la trace de la courbure de Ricci $\mathrm{Ric}_g(\cdot,\cdot)_x$ par rapport au produit scalaire g_x, c'est-à-dire la somme de ses valeurs propres. Elle mesure un défaut d'« euclidianité » dans le volume des petites boules.*

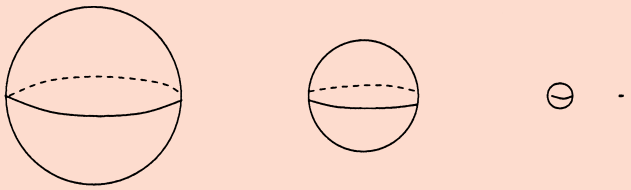
Le point de vue des équations différentielles, évoqué ci-dessus, est très difficile à mettre en œuvre sur l'espace de dimension infinie $\mathcal{M}$. On lui préfère l'approche plus efficace qui consiste à écrire (1) en paramétrant M (par des coordonnées locales). Alors, (1) devient une équation aux dérivées partielles parabolique du type *réaction-diffusion*.

Encadré 2

Exemple 2

Sur la sphère ronde, le flot de Ricci a pour solution $g(t) = (1 - 2\lambda t)g_0$ sur l'intervalle $[0, \frac{1}{2\lambda}[$, si la courbure de Ricci de la métrique initiale est $\mathrm{Ric}_{g_0} = \lambda g_0$ avec $\lambda > 0$.

Plus généralement le flot évolue par homothétie si la métrique initiale est d'Einstein. On peut résumer l'évolution en disant que les métriques de courbure négative enflent et celles de courbure positive se contractent.



Les travaux de R. Hamilton

Hamilton ([Ham-3]) a démontré l'existence d'une solution en temps petit pour toute donnée initiale lisse (voir [DeT] pour une preuve simple). De plus on peut prolonger le flot tant que les courbures sectionnelles (voir encadré 1) restent bornées en valeur absolue. Pour contrôler les courbures, on écrit leurs équations d'évolution et on utilise des principes du maximum.

Principes du maximum sur les courbures

Commençons par la courbure scalaire, qui évolue selon l'équation parabolique :

$$\frac{\partial R}{\partial t} = \Delta R + 2|\mathrm{Ric}|^2 \tag{2}$$

où toutes les quantités dépendent de $g(t)$. En un point x minimisant la courbure scalaire de $g(t)$, le laplacien ΔR est positif ou nul et donc $\frac{\partial R}{\partial t} \geqslant 0$. Heuristiquement, on peut penser que le minimum sur M de la courbure scalaire de $g(t)$, qu'on note $R_{min}(t)$, croît avec t. Un principe du maximum permet de montrer cela de manière rigoureuse. Mieux, si $R_{min}(0) > 0$, on peut montrer que $R_{min}(t)$ tend vers $+\infty$ en temps fini (si le flot existe). Dans ce cas, on est sûr que le maximum des courbures sectionnelles tend vers $+\infty$ en temps fini. Pour avoir plus d'informations, on utilise l'équation d'évolution du tenseur de Ricci, qui est de la forme :

$$\frac{\partial \mathrm{Ric}}{\partial t} = \Delta \mathrm{Ric} + Q(\mathrm{Ric}) \tag{3}$$

où Q est une expression quadratique. Un principe du maximum vectoriel montre que si $\mathrm{Ric}_{g_0} \geqslant 0$ alors $\mathrm{Ric}_{g(t)} \geqslant 0$. Si de plus $\mathrm{Ric}_{g_0} > 0$, c'est vrai pour tout t et on a alors en tout point le pincement

$$\frac{|\mathrm{Ric} - \frac{R}{3} g|}{R} \leqslant \frac{\alpha}{R^\beta}, \tag{4}$$

où α et β sont des constantes strictement positives. Ceci signifie que lorsque $R(x,t) \to +\infty$, l'écart relatif en x de $\mathrm{Ric}_{g(t)}$ à sa moyenne $\frac{R}{3} g(t)$ tend vers 0. A l'aide d'un contrôle du gradient de la courbure scalaire, Hamilton montre que, sous l'hypothèse de stricte positivité de la courbure de Ricci, celle-ci explose en temps fini partout à la même vitesse. Alors, quitte à la renormaliser pour la rendre de volume constant, $g(t)$ converge vers une métrique de courbure sectionnelle constante strictement positive, d'où le

Théorème 3. *Si M est une variété riemannienne fermée possédant une métrique de courbure de Ricci strictement positive, alors M peut être munie d'une métrique de courbure sectionnelle constante strictement positive.*

Remarque 4. *En particulier M est le quotient de la sphère $\mathbf{S}^3$ par un groupe fini d'isométries. Une telle variété est dite* ***sphérique****, comme par exemple l'espace des droites de $\mathbf{R}^4$, qui est le quotient de la sphère $\mathbf{S}^3$ par l'antipodie ; c'est l'espace projectif, noté $\mathbf{P}^3(\mathbf{R})$. C'est le théorème fondateur de toute la théorie et le premier pas vers la conjecture de Poincaré.*

La situation est radicalement différente si la courbure de Ricci n'est pas strictement positive. Le résultat le plus général est que, *pour toute donnée initiale, le flot existe sur un intervalle maximal* $[0, T[$ *et que si $T < \infty$ le maximum des courbures sectionnelles au temps t tend vers $+\infty$ quand $t \to T$.* Dans le dernier cas, on dit que T est un temps *singulier*. En général la courbure explose sur une partie seulement de la variété, on dit que le flot rencontre une singularité. Cependant une variante des résultats précédents, le théorème dit de « pincement » de Hamilton-Ivey,

montre que la partie négative de la courbure devient négligeable comparée à la courbure scalaire. En particulier, la courbure scalaire contrôle toutes les courbures. Disons maintenant quelques mots sur l'étude des singularités.

Etude des singularités : la technique du zoom

Cette technique, classique en analyse, a été mise en œuvre dans ce contexte par Hamilton dans ([Ham-sing]) et ([Ham-comp]). Un zoom consiste à dilater la métrique et ralentir l'écoulement du temps pour avoir une nouvelle solution du flot. On va considérer des suites de zooms et tenter de passer à la limite (voir encadré 3). Si on peut montrer l'existence de flots limites et les classifier, on obtient des modèles pour les singularités. L'existence du flot limite a été un des points de blocage du programme d'Hamilton. Cette question vient d'être complètement résolue par Perelman.

En général on considère une suite de zooms sur des points (x_k,t_k) telle que $Q_k := R(x_k,t_k) \to +\infty$ et maximise la courbure scalaire sur $M \times [0,t_k]$. Alors la suite de dilatations paraboliques $g_k(t)$ en (x_k,t_k) est de courbure bornée sur des intervalles $[-t_k Q_k,0]$ convergeant vers $]-\infty,0]$. Une condition adéquate pour assurer la convergence de la suite (ou d'une sous-suite) $(M,g_k(t),x_k)$ – en un sens que nous ne préciserons pas – vers un flot $(M_\infty,g_\infty(t),x_\infty)$ est une minoration du volume de la boule unité centrée en x_k (pour la métrique $g_k(0)$) par une constante strictement positive indépendante de k. Le premier apport frappant de Perelman ([Per1]) est d'établir que cette minoration est toujours satisfaite si l'explosion de la courbure a lieu en temps fini. Par construction, la limite obtenue est un flot sur $]-\infty,0]$ de courbure bornée et non nulle. De plus, le théorème de pincement de Hamilton-Ivey permet de montrer que la courbure sectionnelle est positive ou nulle sur le flot limite.

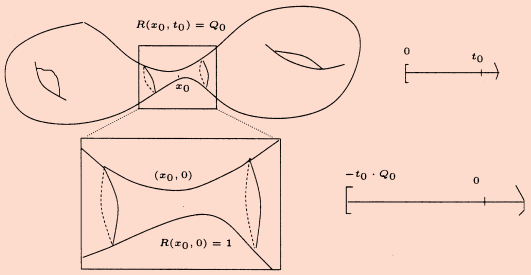
Encadré 3

Dilatation parabolique

L'idée du zoom est formalisée par la notion de dilatation parabolique. *Etant donné un flot de Ricci $g(t)$ sur $M \times [0,T[$, un point x_0 et un temps t_0, c'est la solution du flot donnée par la formule*

$$g_0(t) = Q_0 \cdot g\left(t_0 + \frac{t}{Q_0}\right)$$

où $Q_0 = R(x_0,t_0)$. Elle est définie sur $[-t_0 Q_0, (T-t_0)Q_0[$. C'est une renormalisation telle que $R_{g_0}(x_0,0)$ soit égale à 1.



Le flot avec chirurgie de G. Perelman

Un résultat majeur du premier article de Perelman ([Per1]) est le théorème des voisinages canoniques, qui décrit la métrique $g(t)$ aux points de grande courbure scalaire. Si la courbure est grande partout, on en déduit la classification de M. Dans ce cas on dit que le flot s'éteint. Sinon l'idée (qui remonte à Hamilton ([Ham-iso])) est de se débarrasser des morceaux de M de grande courbure en coupant la variété le long de sphères $\mathbf{S}^2$ et en rebouchant les trous par des boules B^3. Bien entendu, il faut faire cela en contrôlant la topologie et la géométrie. Puis on relance le flot sur la nouvelle variété, peut-être non connexe, et on itère le processus. Eventuellement, certaines composantes connexes disparaissent lors des chirurgies, on dit aussi qu'elles s'éteignent. Perelman démontre dans [P2] qu'on peut poursuivre ce flot avec chirurgie indéfiniment, pour toute donnée initiale convenablement normalisée. Sur chaque intervalle de temps fini, on n'opère qu'un nombre fini de chirurgies. Si le flot s'éteint complètement en temps fini,

on peut classer toutes ses composantes connexes et donc la variété de départ. C'est ce qui se passe dans la preuve de la conjecture de Poincaré. La classification en temps long est plus difficile et nous ne l'aborderons pas ici.

Les voisinages canoniques

Le théorème des voisinages canoniques affirme essentiellement qu'aux points de grande courbure scalaire d'un flot de Ricci, la géométrie est canonique, c'est-à-dire presque isométrique à un nombre fini de modèles simples. Pour ne pas inclure trop de paramètres, par dilatation de la métrique initiale, on peut supposer que le flot vit sur [0,1] au moins et demander que les boules unités de la métrique initiale soient presque euclidiennes. On dira la donnée initiale normalisée. Alors,

Théorème 5. *Pour tout $\varepsilon > 0$ assez petit, il existe une constante universelle $r = r(\varepsilon) > 0$ avec la propriété suivante. Soit $(M, g(t))$ un flot de Ricci de donnée initiale normalisée, $x \in M$ et $t \geqslant 1$ tel que $R(x,t) \geqslant r^{-2}$. Alors x possède un voisinage, ε-presque-isométrique, après une dilatation de facteur $\sqrt{R(x,t)}$, à un des modèles suivants :*

i) Un cylindre $\mathbf{S}^2 \times]-1/\varepsilon, 1/\varepsilon[$, avec la métrique canonique produit, de courbure scalaire 1. On appelle ce voisinage une ε-gorge.

ii) Une boule B^3 ou le complémentaire d'une boule dans l'espace projectif, c'est-à-dire $\mathbf{P}^3(\mathbf{R}) - \overline{B^3}$, munie d'une métrique de courbure strictement positive qui est proche, en dehors d'un compact, d'un cylindre sphérique comme ci-dessus. On appelle un tel voisinage un ε-capuchon.

iii) Une variété fermée de courbure sectionnelle strictement positive.

On dira que $g(t)$ satisfait l'hypothèse des voisinages canoniques à l'échelle r (voir fig. 1).

Par deux variétés ε-presque-isométriques nous entendons deux variétés difféomorphes et dont les métriques riemanniennes sont ε-proches ainsi que leurs dérivées d'ordre $\leqslant 1/\epsilon$. En particulier, les courbures sur ces voisinages sont comparables à la courbure scalaire $R(x,t)$. La taille des voisinages correspondant à i) et ii) est comparable à $R(x,t)^{-1/2} \times 2/\varepsilon$. De plus les oscillations spatiales et temporelles de la courbure scalaire sont contrôlées par des constantes universelles.

Remarque 6. *Dans le dernier cas, par connexité, tout M est contenue dans le voisinage et d'après le théorème 1, M est difféomorphe à une variété sphérique.*

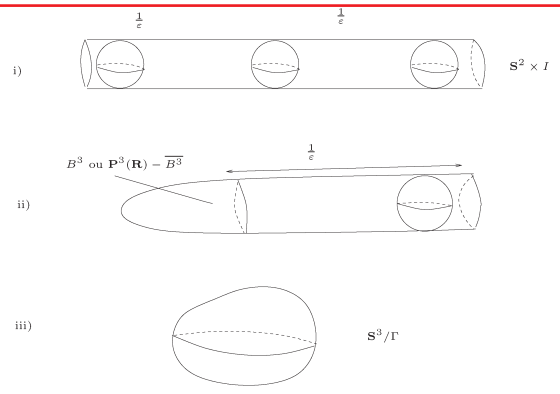


Figure 1

Description du premier temps singulier

Dans ce paragraphe, on suppose fixés $\varepsilon > 0$ et une échelle $r > 0$ pour laquelle $g(t)$ satisfait l'hypothèse des voisinages canoniques. On décrit la métrique $g(t)$ lorsque $t \to T < \infty$ et T est un temps singulier. Notons Ω l'ensemble des points où la courbure scalaire reste bornée, c'est-à-dire

$$\Omega = \{x \in M,\ R(x,\cdot) \leqslant c(x) < +\infty\}.$$

Par hypothèse, il existe $x \in M$ tel que $R(x,t) \to +\infty$ donc Ω est strictement plus petit que M.

Ω est vide : le flot s'éteint

Alors on peut montrer que M est une variété sphérique, un produit $\mathbf{S}^2 \times \mathbf{S}^1$ ou une somme connexe de projectifs notée $\mathbf{P}^3(\mathbf{R})\#\mathbf{P}^3(\mathbf{R})$. En effet, si la courbure explose partout, on peut trouver un temps t_0 proche du temps singulier T tel que $(M, g(t_0))$ est recouverte par un nombre fini de voisinages canoniques. S'il y a un voisinage de type iii) M est difféomorphe à une variété sphérique. Sinon on met bout à bout des gorges jusqu'à ce qu'elles se referment en un $\mathbf{S}^2 \times \mathbf{S}^1$ ou bien on les bouche par des capuchons pour obtenir $\mathbf{S}^3$, $\mathbf{P}^3(\mathbf{R})$ ou $\mathbf{P}^3(\mathbf{R})\#\mathbf{P}^3(\mathbf{R})$.

Remarque 7. *Dans le cas où M est simplement connexe, on obtient qu'elle est difféomorphe à* $\mathbf{S}^3$.

Ω n'est pas vide

Les contrôles sur la courbure permettent de montrer que Ω est un ouvert sur lequel la métrique $g(t)$ converge vers une métrique régulière $g(T)$. Essentiellement, $g(T)$ satisfait l'hypothèse des voisinages canoniques par passage à la limite. Pour comprendre la structure de Ω, on se donne une échelle de courbure $\rho < r$ et on définit l'ensemble

$$\Omega_\rho = \{x \in \Omega\,;\ R(x,T) \leqslant \rho^{-2}\}.$$

L'ensemble $\Omega \backslash \Omega_\rho$ est recouvert par des gorges et des capuchons. L'examen des différentes combinaisons montre que tout point $x \in \Omega \backslash \Omega_\rho$ est dans un des ensembles suivants :

i) *un ε-tube* : un cylindre $\mathbf{S}^2 \times I$ union d'un nombre fini de gorges et dont le bord est dans Ω_ρ.

ii) Une *ε-pointe*, c'est-à-dire une réunion d'une infinité de gorges, difféomorphe à $\mathbf{S}^2 \times \mathbf{R}^+$. Le bout $\mathbf{S}^2 \times \{0\}$ est dans Ω_ρ, tandis qu'à l'autre bout la courbure scalaire tend vers $+\infty$.

iii) Une réunion d'un nombre fini de gorges fermée par un capuchon reliée par le bord à Ω_ρ.

iv) Des composantes connexes disjointes de Ω_ρ : des *doubles pointes* difféomorphes à $\mathbf{S}^2 \times \mathbf{R}$, réunion d'une infinité de gorges, et des *capuchons pointés* difféomorphes à $\mathbf{R}^3$, réunion d'une infinité de gorges fermée par un capuchon (voir fig. 2).

Remarque 8. *Si Ω_ρ est vide, on montre comme dans le cas où Ω est vide que M est difféomorphe à* $\mathbf{S}^2 \times \mathbf{S}^1$, $\mathbf{P}^3(\mathbf{R})\#\mathbf{P}^3(\mathbf{R})$ *ou à une variété sphérique. On dit encore que le flot s'éteint, même si la courbure n'explose pas partout.*

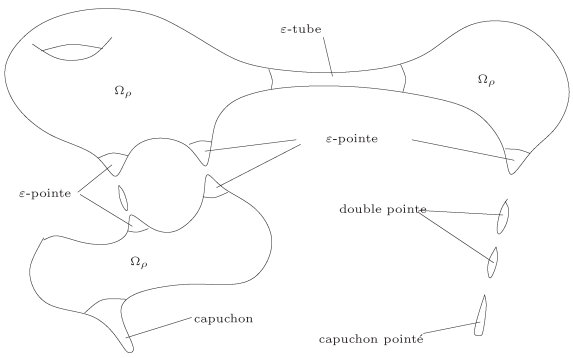


Figure 2

La chirurgie

On opère sur cet ensemble Ω la chirurgie décrite comme suit :

1°) on élimine toutes les composantes connexes de Ω disjointes de Ω_ρ,

2°) on tronque les pointes (reliées à Ω_ρ) et on les bouche par un capuchon difféomorphe à une boule (voir fig. 3).

On obtient ainsi une nouvelle variété différentielle, éventuellement non connexe, que nous noterons M_1. En tout temps $t < T$ proche de T, $(M\backslash\Omega_\rho, g(t))$ est recouverte par des voisinages canoniques. On vérifie ainsi que M est la somme des différentes composantes connexes de M_1 et éventuellement d'un nombre fini d'anses $\mathbf{S}^2 \times \mathbf{S}^1$ et d'espaces projectifs $\mathbf{P}^3(\mathbf{R})$.

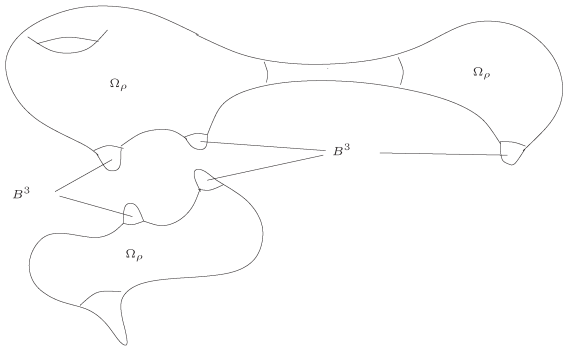


Figure 3

Cette chirurgie doit se pratiquer de manière métrique, c'est-à-dire en contrôlant précisément les recollements effectués. Pour cela, on choisit de tronquer les pointes au milieu d'une δ-gorge, pour un paramètre $0 < \delta \ll \varepsilon$. On définit ainsi une chirurgie avec paramètre (r,δ), où r est l'échelle de courbure (dépendant de ε) à partir de laquelle on a des voisinages canoniques et le paramètre ρ est fixé en posant $\rho = \delta r$. On munit alors M_1 d'une métrique riemannienne bien choisie $g_1(T)$ qui devient la donnée initiale de l'équation (1) et on relance le flot simultanément sur les composantes connexes de M_1.

Remarque 9. *Si l'ensemble Ω_ρ est vide, la chirurgie ci-dessus a un sens. Mais dans ce cas $M_1 = \emptyset$ et le flot s'éteint.*

Le flot avec chirurgie

Le tour de force de Perelman dans [P2] est d'avoir réussi à itérer indéfiniment la construction ci-dessus : fixons un $\varepsilon > 0$,

Définition 10. *Soit $r(t)$, $\delta(t)$ des fonctions strictement positives décroissantes sur $[0, +\infty[$. On appelle flot avec chirurgie la donnée*

i) d'une suite discrète $(t_k)_{0\leqslant k\leqslant N\leqslant\infty}$ de $[0, +\infty[$, strictement croissante et pour chaque entier k,

ii) d'une variété compacte M_k, pouvant être non connexe ou vide,

iii) d'un flot de Ricci $g_k(t)$ sur $M_k \times [t_k, t_{k+1}[$, singulier en t_{k+1}, satisfaisant l'hypothèse des voisinages canoniques à l'échelle $r(t)$,

tels que $(M_{k+1}, g_{k+1}(t_{k+1}))$ est obtenu de $(M_k, g_k(t))$ par une chirurgie de paramètres (r,δ) au temps t_{k+1}.

On dit qu'une variété riemannienne (M,g_0) est normalisée si les courbures sectionnelles sont bornées en valeur absolue par 1 et si le volume de toute boule unité est au moins la moitié du volume euclidien. Perelman démontre qu'il existe des fonctions strictement décroissantes $(r(t),\delta(t))$ universelles, telle que le flot avec chirurgie existe sur $[0,\infty[$ pour toute donnée initiale (M,g_0) normalisée.

En particulier, il n'y a qu'un nombre fini de chirurgies sur chaque intervalle fini. Si M_k est la variété obtenue après le k-ième temps singulier, en tenant compte des composantes qui s'éteignent, M s'obtient comme la somme connexe des composantes connexes de M_k avec un certain nombre de copies de $\mathbf{S}^2\times\mathbf{S}^1$ et de quotients finis de $\mathbf{S}^3$. Si M_k est vide le flot s'éteint et M est difféomorphe à une somme connexe d'un nombre fini de $\mathbf{S}^2\times\mathbf{S}^1$ et de quotients finis de $\mathbf{S}^3$. En particulier si M est simplement connexe elle est difféomorphe à $\mathbf{S}^3$.

La conjecture de Poincaré

L'existence d'un flot avec chirurgie en temps infini étant établie, la preuve de la conjecture de Poincaré consiste à montrer qu'il s'éteint en temps fini sur une sphère d'homotopie, c'est-à-dire sur une variété simplement connexe. D'après ce qui précède celle-ci est alors difféomorphe à $\mathbf{S}^3$. Nous donnons maintenant plus de détails.

On se donne une variété M_0 compacte, simplement connexe que l'on suppose irréductible (voir encadré 4). On la munit d'une métrique normalisée g_0. Pour cette donnée initiale, on construit un flot avec chirurgie $(M_k,g_k(t))$ défini sur $[0,\infty)$. On sait que chaque M_k (tant qu'il est non vide) contient une composante M_k^1, difféomorphe à M_0 et que les autres composantes sont des sphères. On peut donc considérer la restriction du flot avec chirurgie à cette seule composante. Pour montrer qu'il s'éteint en temps fini, nous esquissons l'argument de T. Colding et W. Minicozzi ([CM]), plus simple techniquement que celui de Perelman ([Per3]).

Encadré 4

Irréductibilité

On dit qu'une variété orientable M est irréductible si toute sphère plongée $\mathbf{S}^2\subset M$ borde une boule B^3. Cela implique que si M est somme connexe de deux variétés, l'une d'elles est difféomorphe à M et l'autre à $\mathbf{S}^3$. Le théorème de Kneser affirme que toute variété orientable est somme connexe d'un nombre fini de variétés irréductibles et de copies de $\mathbf{S}^2\times\mathbf{S}^1$. En particulier, si M est simplement connexe elle est somme connexe d'un nombre fini de variétés simplement connexes irréductibles.

La largeur de $(M_0,g(t))$ est définie par minimax de l'énergie des sphères $\mathbf{S}^2$ d'un balayage de M_0. C'est une quantité géométrique strictement positive lorsque le lacet définissant le balayage est essentiel dans l'espace $\mathcal{H}$ des applications de $\mathbf{S}^2$ dans $(M_0,g(t))$, continues et d'énergie bornée. L'existence d'un lacet essentiel dans $\mathcal{H}$ est une conséquence de la simple connexité de M_0. On fixe alors une fois pour toute la classe d'homotopie β d'un lacet essentiel dans $\mathcal{H}$. On définit la largeur $W([\beta],g(t))$ de la variété riemannienne $(M_0,g(t))$ par :

$$W([\beta],g(t))=\inf_{\gamma\in[\beta]}\sup_{s\in[0,1]}E(\gamma(s)),$$

où

$$E(f)=\int_{\mathbf{S}^2}|df|^2_{g(t)}\mathrm{dvol}_{\mathbf{S}^2},$$

est l'énergie de l'application $f:\mathbf{S}^2\to(M_0,g(t))$.

La preuve de l'extinction en temps fini repose alors sur les deux faits suivants :

1°) Sur les parties lisses du flot la largeur $W([\beta],g(t))$ décroît assez vite le long du flot d'après l'inégalité suivante de Colding et Minicozzi ([CM]) :

$$\frac{dW([\beta],g(t)}{dt} \leqslant -4\pi + \frac{3}{4(t+C)} W([\beta],g(t)).$$

(C est une constante calculable.) Cela assure l'extinction en temps fini si le flot reste lisse, puisque la largeur atteint 0 en temps fini et, par ailleurs, doit être strictement positive.

2°) Si t_{k+1} est un temps singulier pour le flot $g_k(t)$ sur M_0, on a

$$\lim_{t \to t_{k+1}^-} W([\beta],g_k(t)) \geqslant W([\beta],g_{k+1}(t_{k+1}))\,.$$

Cela résulte de l'existence d'un difféomorphisme $(1+\xi(t))$-lipschitzien entre $(M_0,g_k(t))$ et $(M_0,g_{k+1}(t_{k+1}))$ avec $\xi(t) \to 0$ quand $t \to t_{k+1}$.

Pour en savoir plus

[B] Bessières (L.), Conjecture de Poincaré : la preuve de R. Hamilton et G. Perelman, *Gazette des Mathématiciens*, 106 (2005), 7-35.

[Bes] Besson (G.), *Preuve de la conjecture de Poincaré en déformant la métrique par la courbure de Ricci, d'après G. Perelman* à paraître à Astérisque. Version modifiée du séminaire Bourbaki n° 947, (2004-2005), 57-ième année.

[CM] Colding (T.) et Minicozzi (W.), Estimates for the extinction time for the Ricci flow on certain three-manifolds and a question of Perelman, *J. of the A.M.S*, 18 (2005), n° 3, 561-569.

[DeT] DeTurck (D.), Deforming metrics in the direction of their Ricci tensor, *J. Differential Geom.*, 18 (1983), 157-162.

[Ham-3] Hamilton (R.), Three-manifolds with positive Ricci curvature, *J. Differential Geom.* 17 (1982), n° 2, 255-306.

[Ham-4] Hamilton (R.), Four-manifolds with positive curvature operator, *J. Differential Geom.* 24 (1986), 153-179.

[Ham-sing] Hamilton (R.), The formations of the singularities of the Ricci flow, *In Surveys in Differential Geometry*, volume II, 7-136, International press, Cambridge MA,1995

[Ham-comp] Hamilton (R.), A compactness property for solutions of the Ricci flow, *Amer. Jour. Math.* 117 (1995), 545-572.

[Ham-iso] Hamilton (R.), Four-manifolds with positive isotropic curvature, *Comm. Anal. Geom.* 5 (1997), 1-92.

[Per1] Perelman (G.), The entropy formula for the Ricci flow and its geometric applications. *ArXiv : math.DG/0211159*.

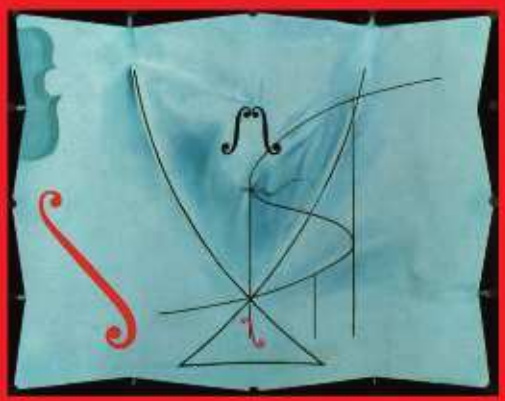
[Per2] Perelman (G.), Ricci flow with surgery on three-manifolds. *ArXiv : math.DG/0303109*.

[Per3] Perelman (G.), Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. *ArXiv : math.DG/0307245*.

[Poin] Henri Poincaré, Œuvres Tome VI, *Gauthier-Villard*, Paris 1953.

[Sco] Scott (P.), The geometries of 3-manifolds. *Bull. London Math. Soc.*, 15 (1983), n° 5, 401-487.

[Thu] Thurston (W.P.), Three dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bull. Amer. Math. Soc.* **6** (1982), 357-381.



CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
3, RUE MICHEL-ANGE 75794 PARIS CEDEX 16 • TÉL 01 44 96 40 00 • TÉLÉCOPIE 01 44 96 50 00

Jean-Pierre Serre et le métier de mathématicien

De Mozart, il a l'élégance, la liberté, l'inventivité, la répugnance aux compromis, l'horreur du toc, la jubilation de faire ce qu'il fait et de le faire si bien, et une œuvre unique, prodigieuse, enthousiasmante. Il s'agit de mathématiques, ici, pas de musique, mais où est la différence ? C'est pourquoi lorsque l'Académie norvégienne a annoncé le nom de Jean-Pierre Serre comme premier récipiendaire du Prix Abel (prix destiné à suppléer à l'absence de Prix Nobel en mathématiques), aucun mathématicien au monde n'en a été surpris. Beaucoup se sont fait la réflexion qu'il n'y avait guère d'autre choix aussi évident, et que c'est bien ainsi.

Il ne fait aucun doute pour les mathématiciens que si l'humanité survit quelques siècles à ses pagailles et à ses folies meurtrières, elle se souviendra du nom de Serre comme celui d'un homme qui aura particulièrement contribué à l'honneur de l'esprit humain. De Harvard à Singapour, lorsqu'on annonce un exposé de Jean-Pierre Serre, les amphis débordent, et on en sort enthousiaste et séduit. Personne ou presque, pourtant, ailleurs que chez les scientifiques, ne connaît son nom… Ce n'est pas qu'il manque de reconnaissance : plus jeune récipiendaire (28 ans) de la médaille Fields à ce jour, plus jeune Professeur au Collège de France (élu à 30 ans), membre d'innombrables académies, titulaire de tous les prix prestigieux. S'il n'est pas une star, c'est sans doute que les mathématiques font peur, et sont souvent difficiles à vulgariser ; c'est aussi que Serre, comme beaucoup de mathématiciens, répugne à la notoriété médiatique : il en craint les déformations, et protège son travail d'entraves inutiles. Et puis, à vrai dire, il n'en a rien à faire…

Ce prix Abel n'a guère été l'occasion de combler un peu cette lacune dans l'information. Les articles auraient pu expliquer, à juste titre, l'importance et la variété des contributions de Jean-Pierre Serre aux mathématiques, l'ampleur de son influence, ainsi que l'extraordinaire constance de cet apport : il avait à peine plus de 20 ans lorsque ses travaux de topologie l'ont révélé au monde mathématique ; à plus de 70 ans, il continue, au même niveau scientifique, à étonner et réjouir ses collègues – mettant ainsi en défaut, ô combien, la rumeur selon laquelle un mathématicien est « fini » à quarante ans.

Ses travaux ont commencé au milieu du siècle dernier, avec sa participation au séminaire de topologie algébrique de Henri Cartan à l'Ecole normale supérieure. Il a découvert que la théorie de Leray des espaces fibrés et de leur suite spectrale pouvait en quelque sorte être appliquée dans un cadre beaucoup plus large qu'on ne l'imaginait alors. Cela l'a amené au calcul des groupes d'homotopie des sphères, *via* des méthodes et des résultats qui ont révolutionné la topologie algébrique – et lui ont valu la médaille Fields en 1954. Dans l'approche que Serre a alors développée on trouve un modèle de ce qui fut, de ce qui est, toute sa vie mathématique : il est sans aucun doute l'un des plus éminents spécialistes de la géométrie algébrique, de la théorie des groupes, de la théorie des nombres, de la topologie, mais dans chacun de ces domaines il importe, utilise, façonne des outils et des idées venant des autres domaines. Il n'est pas surprenant qu'il ait particulièrement aimé travaillé autour des formes modulaires et des représentions ℓ-adiques : on y utilise de la théorie des nombres, de la géométrie, des groupes de Lie, de la combinatoire…

Ici je voudrais essayer de donner le point de vue plus personnel d'un mathématicien. Il me faut, pour cela, enfoncer quelques portes largement ouvertes dans les départements de mathématiques ; j'en demande pardon par avance à mes collègues, mais cet article ne leur est pas destiné.

Nulle part mieux que dans le travail de Serre n'apparaissent à la fois le caractère concret des mathématiques et l'étonnante efficacité de l'abstraction, la dialectique permanente de la théorie et de la pratique que fournissent les mathématiques. Leur caractère concret : Serre connaît un par un les objets mathématiques, les nombres, les groupes, les espaces, il les connaît sous beaucoup d'angles, de plusieurs points de vue, ils lui sont familiers, il connaît leurs particularités et parfois même leurs pathologies ; mieux que quiconque il sait que les mathématiques étudient des objets qui sont ce qu'ils sont, qui ne dépendent ni de notre bon vouloir ni de ce que nous voudrions qu'ils soient. Mais mieux que quiconque aussi il sait, lorsque cela devient nécessaire, prendre le recul, l'envol, monter sur la montagne pour regarder loin et voir globalement : c'est à cela qu'il utilise l'« abstraction » et les outils des mathématiques. La théorie, chez Serre (elle est parfois grandiose, toujours pertinente, et toujours élégante) ne vient que lorsqu'elle s'impose naturellement ; elle n'est jamais là *a priori*, jamais comme un but en soi. Si on peut l'éviter, on l'évite ; sinon, alors, on la fait belle, forte, soignée, lisse, et son efficacité en découle.

S'il est un nom qui symbolise la puissance presque suffocante de l'abstraction, de la théorie, dans les mathématiques du siècle dernier, c'est bien celui d'Alexandre Grothendieck. Grothendieck, et ceux qui connaissent son travail,

savent combien il doit aux intuitions, à la pertinence et aux méthodes de la pensée de Serre. La publication récente de la correspondance mathématique entre Serre et Grothendieck est une contribution majeure à l'histoire des idées.

La beauté et l'élégance des mathématiques de Serre, voilà encore une constatation que l'on entend dans toutes les langues, dans tous les pays où se font des mathématiques. Je viens de tenter d'en aborder l'un des aspects : son usage à la fois retenu et virtuose de la théorie. C'est un des ingrédients de l'« élégance ». Il en est d'autres, difficiles à expliquer au profane ; croyez bien qu'il n'y a là nulle arrogance, juste un grand regret de ne pouvoir partager ce plaisir en quelques lignes.

Générosité : un terme qui vient tout de suite à l'esprit à propos de Serre. Pourtant il n'est pas tout à fait adapté. C'est que, si être « généreux » c'est partager ses intuitions, expliquer ses idées, communiquer sa compréhension, alors la générosité est tout simplement l'un des plaisirs essentiels du mathématicien. Mais peut-être, tout de même, Serre est-il encore plus « mathématicien » que d'autres. Lui qui a tant donné, suscité, éclairé, montré la voie, n'a jamais exercé, ne fût-ce qu'indirectement, un « droit de propriété » sur ses idées, ses découvertes ou ses outils. Il a parfois, certes, défendu la mémoire et l'apport de ceux qu'il respecte ; il s'est parfois battu pour rétablir une vérité historique. Mais jamais il n'a donné l'impression (et sans doute n'a-t-il jamais pensé) que ce qu'il avait donné lui « appartenait » de quelque manière que ce fût. Oui, en fin de compte, généreux.

On ne peut pas évoquer cet aspect de sa pratique scientifique sans mentionner Bourbaki. Serre a joué un rôle de tout premier plan dans les travaux et le fonctionnement de ce groupe de mathématiciens, fondé juste avant guerre par quelques jeunes Français. Parmi eux, André Weil – voilà encore un nom peu connu, si ce n'est comme « grand frère de Simone Weil » ; et pourtant ! –, à qui Serre doit beaucoup. Il est temps, encore, de dire haut et fort ce qu'a été Nicolas Bourbaki : une association de scientifiques de tout premier niveau, créée dans le but d'établir des fondements solides aux Mathématiques ; un modèle de fonctionnement désintéressé, mû par la seule nécessité de la compréhension, l'explication, la clarification, pour la beauté et la raison – et la jubilation des mathématiques ! – sans aucune espèce de rétribution sous quelque forme que ce soit (l'appartenance à Bourbaki était secrète) ; des mois et des mois de travail passé au crible de la critique impitoyable du groupe, de son exigence et de sa passion communes ; une sorte de modèle de service public universel de la science, en quelque sorte.

Les livres tiennent une place privilégiée dans l'œuvre de Serre. Je le tiens (et beaucoup de mathématiciens avec moi) pour un « classique » de la langue française moderne. Voilà encore un exemple d'élégance : son écriture précise, claire, efficace. Peu de scientifiques se sont donné la peine de rédiger autant de livres, avec autant de soin. C'est là que la générosité de Serre apparaît le plus clairement, et c'est sans doute aussi par cela qu'il passera a la postérité. A l'université de Yale, l'examen de Français pour les « graduate students » de mathématiques consiste à traduire un paragraphe de Serre (ou de Bourbaki)…

L'utilité des mathématiques ? Certes plus personne ne songe à la remettre en cause aujourd'hui, puisqu'on sait bien qu'elles sont partout, de la gestion du trafic aérien à la forme des verres de lunettes en passant par l'imagerie médicale, internet et les télécommunications, de la description des particules élémentaires à la maîtrise de l'information génomique. Cependant, l'utilité de ses mathématiques n'a jamais été, pour Serre, un critère de ses choix de travail. Il se plaît à dire que les mathématiciens se contentent de mettre leur production à la disposition de tous, comme sur des étagères où l'on peut venir se servir. Noter, cependant, que ce service-là est gratuit. Et que le mathématicien ne fabrique et ne dépose sur l'étagère que ce qui lui plaît. Ce qui le motive, c'est le plaisir, pas l'appât du gain. Serre dit souvent qu'il s'est étonné toute sa vie qu'on l'ait payé, si longtemps, pour une activité qu'il aimait tant. C'est bien la jouissance et la curiosité qui ont été, sans aucun doute, les moteurs du travail de Serre. Il lui arrive parfois de répondre, à un interlocuteur qui cherche à attirer son attention sur une question : « ça ne m'intéresse pas » ; il n'y a dans cette réponse abrupte aucune arrogance, ni aucun jugement : ça ne l'intéresse pas, c'est tout !

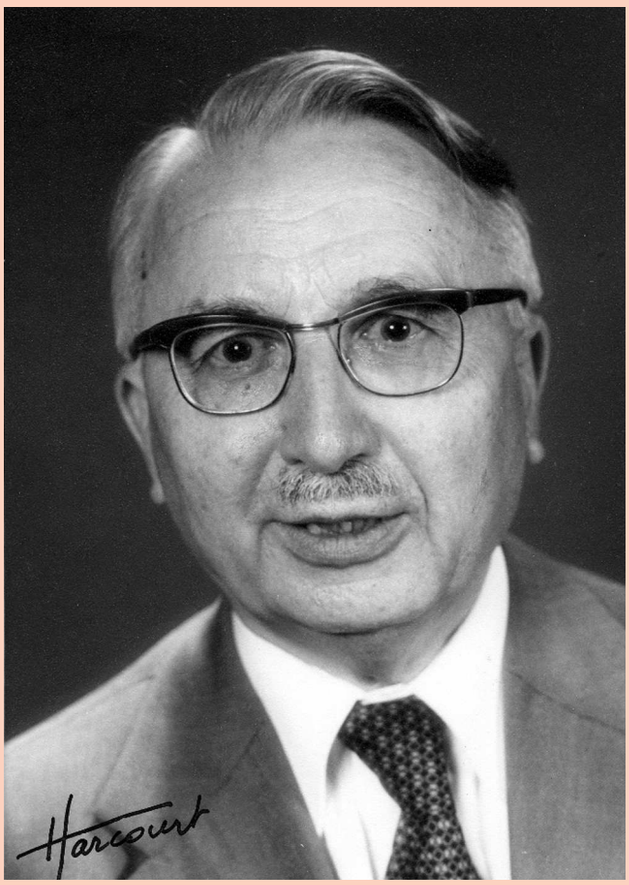
Jean-Pierre Serre incarne cette extraordinaire originalité du mathématicien : il est à la fois interprète du monde et découvreur de ses mystères, révélateur et créateur de sa beauté et de sa complexité, pianiste et compositeur, et, bien sûr ! passionné de l'être. Y a-t-il un plus beau métier, lorsqu'il est pratiqué par un Serre ?

Michel BROUÉ

Institut Henri-Poincaré
11 rue Pierre et Marie Curie, 75005 Paris
broue@ihp.jussieu.fr
http://www.math.jussieu.fr/broue

Juin 2005

Le centenaire d'Henri Cartan



Dans le dossier que *Le Monde* de ce 3 mars 2004 consacre à « *Matière grise : la bataille mondiale* », on lit tout au début des propos du biologiste E. E. Baulieu recueillis par le journal : « *Les mathématiciens français sont parmi les tout premiers du monde.* » Par-delà la remise en cause de certains excès bourbachiques, j'ai la conviction (largement partagée je l'espère) que cette situation est en grande partie héritée d'un groupe de jeunes gens qui, dans les années trente, ont conçu le projet d'un traité « prenant les Mathématiques à leur début. » Dans ce groupe je vois au premier rang Henri Cartan, entouré de ses amis André Weil et Jean Dieudonné.

J'ai eu le privilège de croiser souvent la route d'Henri Cartan. La première fois, c'était (d'après lui) à Strasbourg en 1930 ; j'ai certes oublié cette rencontre – j'avais deux ans – mais j'ai gardé un net souvenir des années qui ont suivi. Cartan était un collègue de mon père et un ami de ma famille, ami admiré mais un peu redouté pour son esprit caustique ; on se faisait aussi du souci pour sa santé fragile ! Puis je l'ai revu en 1940 à Clermont, où l'université de Strasbourg avait été repliée et plus tard à La Bourboule, après la catastrophe qu'il avait pressentie. Et de nouveau à Strasbourg en 1945 après le grand trou d'où ma famille sortait indemne, et pas la sienne. Ensuite à Paris, lui professeur et moi élève à l'école normale : le cours aux élèves de 2^{e} année où il nous a appris ce qu'était une variété différentiable, et où un inconnu (c'était Alexandre Grothendieck) s'est permis de dialoguer avec lui d'égal à égal depuis le fond de la salle ; le premier Séminaire Cartan de Topologie (1948) dont j'ai fait sous sa houlette l'exposé n° 3, ensuite entièrement rédigé par lui ; et souvent dans sa famille, Boulevard Jourdan, où j'étais un peu comme un enfant de plus. Puis comme patron de recherche qui ne donnait pas de sujet, mais qui m'a signalé un jour l'article de Feldbau sur les homéomorphismes des sphères, qui a été le point de départ de ma thèse.

Dès avant cette époque, j'étais assez proche de lui pour recueillir parfois ses conodences, comme celle-ci, dont je me souviens, sur René Thom, qu'il *découvrit* avant tout le monde en dépit de leurs esprits si différents : « *Thom est un garçon rempli d'idées, mais que c'est dur de les lui faire mettre par écrit !* » Puis au fil du temps : ses batailles, le plus souvent victorieuses, pour « *remonter le niveau de la Sorbonne en Mathématiques* » (la formule est de lui) ; ses vains efforts pour qu'une chaire du Collège de France soit créée pour André Weil. Son engagement précoce à une époque où celà demandait une grande hauteur de vue, en faveur de la réconciliation franco-allemande, cette *utopie* dont la réalisation est aujourd'hui un de nos motifs d'espoir face à des conflits en apparence insolubles. Ses engagements pour la défense des Droits de l'Homme partout dans le monde, et pour la construction d'une Europe fédérale, à laquelle il attache tant de prix.

Jean-Pierre Serre a écrit : « *Je crois que le style de Cartan est ce qu'on peut trouver de mieux en Mathématiques* ».

Je crois que le style de Cartan est ce qu'on peut trouver de mieux dans la vie.

Merci, Monsieur Cartan, de nous montrer par votre exemple qu'il est possible de vieillir en devenant de plus en plus humain.

Jean CERF

Texte extrait de la gazette des mathématiciens n° 100, avril 2004

Pour en savoir plus consulter
http://ww.smf.emath.fr/VieSociete/Rencontres/JournéeCartan/

Théorie statistique de l'apprentissage

Olivier CATONI*

Discipline inventée par Vladimir Vapnik ou évolution de l'inférence statistique traditionnelle vers l'analyse de données complexes, la théorie statistique de l'apprentissage se trouve au carrefour de différentes approches, qui touchent aux statistiques, bien entendu, mais aussi à la théorie de l'information ou à la mécanique statistique.

Préliminaires

Le développement des moyens informatiques, des télécommunications et des capteurs électroniques de toutes natures provoque la production et le stockage de données de plus en plus abondantes et de plus en plus complexes. L'exploitation de ces données par des moyens humains devient en conséquence de moins en moins facile, créant un besoin urgent de mettre au point des méthodes automatiques d'analyse permettant de confier à une machine des fonctions de perception, de discrimination et de reconnaissance qui étaient jusque là l'apanage du cerveau humain (ou animal pour bon nombre d'entre elles). Cette recherche s'est avérée plus malaisée que les premiers espoirs de la cybernétique ne l'auraient laissé espérer dans l'immédiat après guerre. En effet, les processus de perception et de traitement réalisés par le cerveau sont mal connus, et leur caractère très largement inconscient ne permet pas de les appréhender par une approche introspective. Nous illustrerons nos propos par deux applications emblématiques, parce qu'elles correspondent aux tâches de perception que le cerveau effectue à jet continu dans la vie courante : la reconnaissance de la parole et la classification d'images (plus généralement l'interprétation de scènes visuelles). Des progrès dans ces domaines en feraient faire à une multitude d'applications dont il serait vain de tenter de faire le tour, comprenant bien entendu la robotique, la navigation assistée, mais aussi le diagnostic médical, la conduite automatisée de processus industriels, et d'une manière plus indirecte (parce que les problèmes d'apprentissage se heurtent à des difficultés génériques indépendantes de l'application envisagée), à l'analyse du génome, la constitution et l'interrogation de bases de données, l'analyse de la langue naturelle, etc.

Nous ne parlerons pas des capteurs, on trouve dans le commerce des caméscopes qui permettent d'enregistrer sur un ordinateur du son et des images de très bonne qualité. Nous ne parlerons pas des recherches faites pour comprendre le fonctionnement du cerveau, parce que nous ne sommes pas certain qu'un ordinateur pourrait le reproduire efficacement : en effet, un cerveau possède un très grand nombre de neurones très fortement interconnectés qui traitent en parallèle des informations à l'aide de processus électrochimiques très lents. Ceci contraste nettement avec l'organisation d'un ordinateur, qui possède une unité centrale (ou un faible nombre d'entre elles) qui traite à une vitesse très élevée des signaux électromagnétiques mais n'accède qu'à une seule donnée à la fois parmi celles qui sont rangées dans une mémoire par ailleurs immobile. De même qu'un muscle et un moteur réalisent en gros la même fonction (fournir du travail mécanique) par des moyens très différents, de même, rien ne permet de penser que la reproduction par une machine de certaines fonctions perceptives du cerveau doive utiliser des algorithmes qui

* *CNRS, Laboratoire de Probabilités, UMR 7599, Université Paris 6, case 188, 4, place Jussieu, F-75252 Paris Cedex 05,*
catoni@ccr.jussieu.fr
http://www.proba.jussieu.fr/users/catoni/homepage/newpage.html

auraient une analogie quelconque avec le fonctionnement interne de celui-ci (notre point de vue paraîtra délibérément polémique à certains, nous espérons que cette prise de position marquée aura au moins le mérite de susciter réactions et réflexions sur la question).

Mises à part les tentatives d'analogie avec le fonctionnement cérébral tel que peuvent le décrire les neurosciences, les chercheurs ont pendant très longtemps essayé de reconnaître des sons ou des images en utilisant la méthode qui fait le succès de la physique depuis Descartes (ou peut-être depuis l'antiquité, nous ne nous prononcerons pas sur ces délicates questions d'histoire des sciences) : établir un modèle du phénomène observé, ici le son ou l'image numérisés, en estimer les paramètres à partir de mesures expérimentales, confronter les prédictions fournies par le modèle (concernant la nature des sons ou des images analysés) avec l'expérience. Cette approche n'a jamais fonctionné correctement pour traiter les problèmes qui nous intéressent ici, sauf dans des situations très simples : personne n'est capable de donner un modèle satisfaisant du bruit que fait une phrase prononcée dans une ambiance sonore quelconque, par un interlocuteur quelconque. Personne n'est non plus capable de donner un modèle de ce qu'enregistre une caméra placée sur le capot d'une voiture, … nous ne sommes pas en présence de phénomènes que l'on puisse décrire à l'aide d'un nombre raisonnable de paramètres et d'équations, comme on le fait des oscillations d'un pendule ou des mouvements d'une planète.

C'est sur cet échec que s'est construite la légitimité des approches statistiques. Bien que la reconnaissance des formes présente toutes sortes de difficultés annexes nous parlerons ici plus spécifiquement de la classification supervisée. Nous supposerons donc disposer d'une base de données $X_1,\ldots,X_N \in \mathcal{X}$ déjà classées et nommerons $Y_1,\ldots,Y_N \in \mathcal{Y}$ les classes correspondantes. L'ensemble $\mathcal{X}$ désignera l'espace mesurable dans lequel les données sont représentées. Dans le cas de la reconnaissance de visages, qui a servi de banc d'essai à de nombreuses méthodes, $X_1,\ldots,X_N$ seront des imagettes centrées soit sur des visages remis à une échelle normalisée, soit sur des contre exemples de ce que l'on rencontre ailleurs dans les images à traiter. Les classes $Y_1,\ldots,Y_N$ prendront alors deux valeurs, correspondant à la présence ou à l'absence de visage. La question posée par l'apprentissage statistique est la suivante : supposons que $X_1,\ldots,X_N$ aient été tirées au hasard parmi une grande « population » d'imagettes, comment choisir une règle de classification qui commette le moins d'erreurs possibles sur la population totale en n'utilisant pour construire cette règle que les exemples observés $X_1,\ldots,X_N$ (et les classes $Y_1,\ldots,Y_N$, supposées fournies par un expert, ou plus généralement tout autre moyen extérieur) ?

Il est important de comprendre que cette approche possède des avantages spécifiques avant d'en commenter les aspects techniques :

– on ne modélise pas les données à analyser, mais seulement une « expérience statistique » qui s'apparente à un sondage. Le seul aléa supposé est celui introduit par le statisticien dans le choix des exemples, les propriétés d'indépendance et d'équidistribution de l'échantillon $X_1,\ldots,\ X_N$ peuvent donc être garanties de façon réaliste ;

– à défaut de modéliser les données à analyser, on doit par contre modéliser les règles de classification qui vont leur être appliquées : ces règles étant construites par le statisticien, et leur complexité étant limitée par la puissance de calcul dont il dispose, cette opération de modélisation est réalisable en pratique. Elle consiste à « structurer » (le terme est de V. Vapnik) l'ensemble des règles dont le statisticien va tester les performances en une réunion de familles « paramétriques », c'est-à-dire de familles d'algorithmes identiques à la valeur d'un certain nombre de paramètres numériques près ;

– l'approche statistique réserve la possibilité de sélectionner des règles de classification différentes pour traiter des jeux de données différents. En ajustant ainsi au plus près le choix de la méthode aux données que l'on souhaite effectivement analyser, on évite d'essayer de résoudre un problème plus compliqué (c'est-à-dire ici plus générique) que nécessaire.

Il faut cependant avoir à l'esprit le fait que la constitution de la base de données sur laquelle va porter la méthode d'apprentissage statistique évoquée ci-dessus pose aussi des questions délicates. En particulier l'extraction et la normalisation des données, que ce soit dans la phase d'apprentissage (c'est-à-dire de sélection de la méthode) ou dans la phase de reconnaissance (c'est-à-dire au moment où on va appliquer la méthode sur des données brutes), est une étape cruciale pour la réussite de l'opération. Elle présente des obstacles tout aussi fondamentaux que l'exploitation de la base de données elle-même, dont les moindres ne sont pas les problèmes dits de « segmentation ». Dans le cas des visages, la segmentation consiste à positionner et à dimensionner convenablement l'imagette autour des zones pouvant contenir un visage. Ce n'est pas trop compliqué parce qu'un visage est un objet assez rigide qui s'inscrit de façon satisfaisante dans un cadre rectangulaire et qu'il y a « peu » de rectangles dans une image. La reconnaissance

Figure 1 – *Les éléments de bords sont des caractéristiques intermédiaire souvent employées en analyse d'images. Ce sont des éléments plus structurés et plus géométriques que les pixels. Ils possèdent une intensité (que l'on peut seuiller) et une orientation. On peut ensuite les regrouper par paquets suivant leurs positions et orientations relatives, et compter le nombre d'apparition de ces configurations dans les images (dans l'illustration ci-dessus, on a regroupé les orientations en trois classes, rouge, verte ou bleue). On obtient ainsi une famille très riche de mesures, à partir de laquelle on peut construire une famille encore plus riche de règles de classification (par exemple en séparant des groupes de mesures par des hyperplans). Les techniques décrites dans cette présentation ont pour but de sélectionner parmi toutes ces règles possibles, une règle dont le taux d'erreur à un niveau de confiance donné soit garanti par une inégalité mathématiquement prouvée.*

d'objets beaucoup plus déformables (un serpent, un chat …) ou partiellement occultés par d'autres poserait de vrais problèmes supplémentaires, qui restent à ce jour largement ouverts. De même, dans le domaine de la reconnaissance de la parole, il existe un saut très important entre la reconnaissance de mots isolés et la reconnaissance d'un discours continu dans lequel les mots s'enchaînent les uns aux autres sans silences permettant d'en identifier facilement les frontières.

L'approche PAC-Bayésienne

Ces remarques faites, venons-en à l'apprentissage d'une règle de classification à partir d'exemples classés $(X_1,Y_1),\ldots,(X_N,Y_N) \in (\mathcal{X}\times\mathcal{Y})$, formant une suite de couples de variables aléatoires indépendants identiquement distribués (i.i.d. en abrégé). Notons $\mathbb{P}$ la loi jointe inconnue de cette suite. Soit $\{f_\theta : \mathcal{X}\to\mathcal{Y};\ \theta\in\Theta\}$ l'ensemble de toutes les règles de classification qui seront envisagées pour classer les données. Comme expliqué dans les préliminaires, Θ se décomposera le plus souvent en une réunion de sous-ensembles de « dimensions » différentes. Un critère naturel, mais malheureusement inaccessible, pour juger de la qualité de la règle f_θ est fourni par son taux d'erreur moyen $R(\theta)=\mathbb{P}\big[f_\theta(X_1)\neq Y_1\big]$ (où l'indice 1 peut être remplacé par n'importe quel autre, l'échantillon étant supposé i.i.d.). Néanmoins ce taux d'erreur moyen est l'espérance d'une variable aléatoire observable, le taux d'erreur empirique (c'est-à-dire constaté sur l'échantillon observé) $r(\theta)=\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\big[f_\theta(X_i)\neq Y_i\big]$. Si nous ne considérions qu'une seule règle f_θ, le lien entre $r(\theta)$ et son espérance serait simple, $r(\theta)$ étant une moyenne de variables de Bernoulli i.i.d. Malheureusement, nous voulons considérer la question bien plus délicate des relations du processus $\theta\mapsto r(\theta)$, où θ varie dans un très grand ensemble Θ, avec le taux d'erreur moyen minimum $\inf_{\theta\in\Theta} R(\theta)$ et la valeur (ou les valeurs) du paramètre θ où il est atteint, soit $\arg\min_{\theta\in\Theta} R(\theta)$. Un phénomène bien mis en valeur par V. Vapnik dès le début de la théorie est celui du « sur-apprentissage » qui peut se décrire qualitativement ainsi : si Θ est « trop grand », les deux minima $\inf_{\theta} R(\theta)$ et $\inf_{\theta\in\Theta} r(\theta)$ peuvent n'avoir aucun lien entre eux, ni les valeurs du paramètre θ pour lesquelles ils sont atteints. Dans ce cas, on obtiendra un meilleur résultat en considérant la relation entre $\inf_{\theta\in\Theta_1} r(\theta)$ et $\inf_{\theta\in\Theta} R(\theta)$, où Θ_1 est un sous-ensemble de Θ de taille convenable. En restreignant Θ à Θ_1, on fait apparaître deux phénomènes antagonistes : un phénomène favorable de « réduction de la variance » qui va faire que $\arg\min_{\theta\in\Theta_1} r(\theta)$ va se rapprocher de plus en plus de $\arg\min_{\theta\in\Theta_1} R(\theta)$, et un phénomène défavorable de biais, qui va faire

que $\inf_{\theta\in\Theta_1} R(\theta)$ va s'éloigner de plus en plus de $\inf_{\theta\in\Theta} R(\theta)$. Pour trouver le meilleur compromis entre ces deux phénomènes, on est conduit à considérer toute une famille de sous-ensembles $(\Theta_j)_{j\in J}$ de Θ. La situation se complique donc : au lieu de chercher le meilleur paramètre θ, nous en sommes maintenant à chercher le meilleur sous-ensemble de paramètres où chercher le meilleur paramètre ! et là encore, une étape de modélisation s'impose : de même qu'il est désavantageux de chercher le meilleur θ dans un ensemble Θ trop grand, de même il est désavantageux de chercher le meilleur Θ_j dans un ensemble de parties de Θ trop grand (en particulier on voit tout de suite que cela n'avance à rien de pousser cette démarche jusqu'à l'extrême en considérant tous les singletons de Θ). Le choix d'une famille $(\Theta_j)_{j\in J}$ de sous-modèles de Θ a été baptisé par V. Vapnik « minimisation structurelle du risque ». La théorie des processus empiriques permet de quantifier les phénomènes que nous venons de décrire en faisant des hypothèses sur la structure du processus $\theta \mapsto r(\theta)$ pour une métrique qui majore les covariances, dans le domaine de la classification on considère souvent $D(\theta,\theta') = \mathbb{P}\big[f_\theta(X_1) \neq f_{\theta'}(X_1)\big]$ qui majore la variance de $\sqrt{N}\big[r(\theta) - r(\theta')\big]$. C'est la voie la plus « classique » d'approche de l'apprentissage statistique. Elle est néanmoins semée d'embûches, la moindre n'étant pas que la distance $D(\theta,\theta')$ n'est en pratique pas plus connue que le reste, et que des hypothèses portant sur le contrôle de l'entropie métrique des sous-espaces (Θ_j, D) sont difficiles à vérifier.

Nous présenterons ici une approche alternative, qui contrôle les quantités évoquées ci-dessus par des moyens détournés. Elle est née dans la communauté du « machine learning », sous l'impulsion séminale de D. McAllester qui l'a baptisée et en a prouvé les premiers résultats. Notons au passage que ce nom de baptême, « Probably Approximately Correct Bayesian theorems », est à comprendre dans une perspective purement historique : la façon de poser le problème que nous venons de décrire n'a rien de Bayésien, de plus l'approche inventée par D. McAllester fournit certes des inégalités de déviations (vérifiées avec probabilité $1-\epsilon$, d'où le préfixe « PAC »), mais peut aussi fournir directement des inégalités en espérance, comme nous allons l'évoquer ; en un mot son contenu n'a rien à voir avec son nom !

Le premier ingrédient de l'approche PAC-Bayésienne consiste à lisser l'étape de minimisation structurelle du risque : au lieu de considérer une famille de sous-modèles, $(\Theta_j)_{j\in J}$, nous allons considérer une mesure de probabilités π sur Θ. Cette mesure n'a pas d'interprétation probabiliste ! Elle peut être vue comme un moyen de spécifier partiellement une représentation des éléments de Θ en choisissant la longueur du code associé (du moins dans le cas où Θ est fini ou dénombrable, mais on peut toujours se ramener à ce cas en pratique en tronquant la représentation des variables réelles). Elle peut aussi être vue comme une sorte de pénalisation *a priori* des différentes parties de Θ.

Le deuxième ingrédient consiste à remplacer le contrôle des fluctuations du processus $\theta \mapsto r(\theta)$ par le contrôle d'une quantité bien connue des physiciens, l'énergie libre, plus sobrement pour les mathématiciens la transformée de Laplace, à savoir $\frac{1}{\lambda}\log\Big[\iint \exp\big[-\lambda r(\theta,\omega)\big]\pi(d\theta)\mathbb{P}(d\omega)\Big]$. En jouant sur le paramètre λ, on va pouvoir se rapprocher plus ou moins de $\inf_\theta r(\theta)$, et donc réaliser quelque chose qui ressemble au compromis sur la taille du modèle recherché par la minimisation structurelle du risque dont nous avons parlé plus haut. En utilisant un peu d'analyse convexe, on pourra alors contrôler le comportement de toutes les « lois *a posteriori* » $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$ possibles, c'est-à-dire de toutes les lois de probabilités sur les paramètres qui dépendent de l'échantillon observé (et sont de ce fait des mesures aléatoires, on supposera plus précisément sans le dire dans ce qui suit que ce sont des probabilités conditionnelles régulières). L'espace probabilisable Ω désigne ici celui sur lequel les variables aléatoires représentant l'échantillon sont construites, on peut en particulier choisir la représentation dite canonique de l'aléa dans laquelle $\Omega = (\mathcal{X}\times\mathcal{Y})^N$ et $(X_i,Y_i)_{i=1}^N(\omega) = \omega$). On pourra en effet utiliser l'identité remarquable :

$$\log\Big\{\int \exp\big[-\lambda r(\theta)\big]\pi(d\theta)\Big\} = \sup_{\rho\in\mathcal{M}_+^1(\Theta)} \lambda\rho\big[r(\theta)\big] - \mathcal{K}(\rho,\pi),$$

où $\mathcal{M}_+^1(\Theta)$ désigne l'ensemble des mesures de probabilités sur Θ et où $\mathcal{K}(\rho,\pi) = \int \log\left(\frac{\rho}{\pi}\right) d\rho$ désigne l'entropie relative de la loi ρ par rapport à la loi *a priori* π (quand ρ n'est pas absolument continue par rapport à π, on pose

$\mathcal{K}(\rho,\pi) = \infty$ par convention). On obtient ainsi facilement les bornes en déviations et en moyenne

$$\mathbb{P}\left\{ \sup_{\rho\in\mathcal{M}_+^1(\Theta)} \int R(\theta)\rho(d\theta) - \frac{\int r(\theta)\rho(d\theta) + \dfrac{\mathcal{K}(\rho,\pi) - \log(\epsilon)}{\lambda}}{1 - \dfrac{\lambda}{2N}} \leqslant 0 \right\} \geqslant 1-\epsilon,\ \lambda < 2N,$$

$$\text{et} \quad \iint R(\theta)\rho(\omega,d\theta)\mathbb{P}(d\omega) \leqslant \int \left[\frac{\int r(\theta,\omega)\rho(\omega,d\theta) + \dfrac{\mathcal{K}[\rho(\omega),\pi]}{\lambda}}{1 - \dfrac{\lambda}{2N}} \right] \mathbb{P}(d\omega),\ \ \rho : \Omega \to \mathcal{M}_+^1(\Theta), \lambda < 2N.$$

(dont nous ne donnons pas les formes les plus précises par souci de simplicité). Les bornes en espérance sont moins intéressantes du point de vue théorique, mais donnent d'un point de vue pratique des constantes plus serrées et sont souvent plus faciles à lire, même si elles ne fournissent que des majorations « sans biais » de l'erreur de généralisation moyenne, qui pourraient s'avérer sans intérêt si une borne en déviation ne permettait de prouver que leurs fluctuations ne sont pas trop grandes.

En travaillant un peu plus, on peut optimiser le paramètre λ dans la première inégalité pour obtenir avec $\mathbb{P}$ probabilité au moins $1-\epsilon$, pour toute loi *a posteriori* $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,

$$\int R(\theta)\rho(\omega,d\theta) \leqslant \left(1 + \frac{2\alpha d}{N}\right)^{-1} \left\{ \int r(\theta,\omega)\rho(\omega,d\theta) + \frac{\alpha d}{N} + \sqrt{\frac{2\alpha d \int r\rho(d\theta)\left[1 - \int r\rho(d\theta)\right]}{N} + \frac{\alpha^2 d^2}{N^2}} \right\},$$

où α est un paramètre réel positif supérieur à 1, que l'on peut prendre par exemple égal à $1 + \left[\log(N)\right]^{-1}$, et où $d = \mathcal{K}\left[\rho(\omega),\pi\right] + \log\left(\dfrac{\log(2\alpha N)}{\epsilon\log(\alpha)}\right)$ est un « terme de complexité ».

Ces inégalités fournissent une première majoration du taux d'erreur moyen d'une règle de classification tirée au hasard suivant la loi *a posteriori* $\rho(\omega,d\theta)$, par une borne qui a le mérite d'être observable, et le défaut d'être infinie pour les masses de Dirac (tout au moins quand π est une mesure diffuse). C'est le prix à payer, semble-t-il, pour obtenir des bornes sans faire d'hypothèses contraignantes sur la structure de (Θ, D). Sous des hypothèses de structure, on pourrait alors montrer que pour $\widehat{\theta}(\omega)$ et $\rho(\omega,d\theta)$ bien choisis $\int D\left[\theta,\widehat{\theta}(\omega)\right]\rho(\omega,d\theta)$ est petit et donc que $R\left[\widehat{\theta}(\omega)\right] \leqslant \int R(\theta)\rho(d\theta) + \int D\left(\theta,\widehat{\theta}\right)\rho(d\theta)$ l'est aussi.

Ces premiers théorèmes PAC Bayésiens peuvent être améliorés d'au moins deux façons : d'une part en jouant sur un choix spécifique de π relié à celui de ρ, menant à des bornes plus « locales », d'autre part en utilisant la structure des covariances du processus $\theta \mapsto r(\theta)$ au lieu d'utiliser la variance de $r(\theta)$. Cependant ces améliorations vont se faire au détriment de la valeur des constantes, si bien qu'elles n'en seront vraiment que pour des valeurs suffisamment grandes de la taille N de l'échantillon. Pour cette raison, les bornes les plus simples gardent tout leur intérêt, en dépit du fait qu'elles ne soient pas asymptotiquement optimales quand N tend vers l'infini.

Localisation

On voit facilement que le choix optimal de la loi *a priori* π dans la borne en espérance est $\pi = \int \rho(\omega)\mathbb{P}(d\omega)$. Malheureusement cette probabilité *a priori* sur les paramètres n'est pas observable (puisque $\mathbb{P}$ est inconnue). Notons qu'elle donne un renseignement intéressant sur le plan théorique : $\int \mathcal{K}\left[\rho(\omega),\pi\right]\mathbb{P}(d\omega)$ est alors égale à *l'information*

mutuelle entre ω (qui représente ici l'échantillon observé) et θ, lorsque ω est tiré suivant $\mathbb{P}$ et θ est tiré suivant $\rho(\omega,d\theta)$ une fois ω choisi. Ainsi, l'écart entre l'erreur de généralisation d'une règle de classification randomisée et l'erreur constatée sur l'échantillon observé est contrôlé par l'information mutuelle entre l'échantillon et le paramètre. En pratique on est tenu de choisir π indépendamment de $\mathbb{P}$ et ce que l'on perd est quantifié par l'identité $\int \mathcal{K}(\rho(\omega),\pi)\mathbb{P}(d\omega) = \int \mathcal{K}\big[\rho(\omega), \int \rho(\omega')\mathbb{P}(d\omega')\big]\mathbb{P}(d\omega) + \mathcal{K}\big[\int \rho(\omega')\mathbb{P}(d\omega'), \pi\big]$. On peut néanmoins aller plus loin de la façon suivante : quand π et λ sont fixés, la loi *a posteriori* optimale (c'est-à-dire qui minimise la borne) a pour densité $\dfrac{d\rho}{d\pi} = \dfrac{\exp\big[-\lambda r(\theta)\big]}{\int \exp\big[-\lambda r(\theta')\big]\pi(d\theta')}$. On la notera $\pi_{\exp(-\lambda r)}$. On peut alors revenir sur le choix de la loi *a priori* et la prendre de la forme $\pi_{\exp(-\beta R)}$. En travaillant un peu sur le lien entre $\pi_{\exp(-\beta R)}$ et sa version empirique $\pi_{\exp(-\beta r)}$, on parvient alors à prouver la borne en espérance

$$\int \left\{ \int r(\theta,\omega)\rho(\omega,d\theta) - \frac{\mathcal{K}\big[\rho,\pi_{\exp(-\beta r)}\big]}{\beta} \right\} \mathbb{P}(d\omega) \leqslant \int\int R(\theta)\rho(\omega,d\theta)\mathbb{P}(d\omega)$$

$$\leqslant \int \left\{ \frac{\int r(\theta,\omega)\rho(\omega,d\theta) + \dfrac{\mathcal{K}\big[\rho,\pi_{\exp(-\beta r)}\big]}{\beta}}{1 - \dfrac{2\beta}{N}} \right\} \mathbb{P}(d\omega)\,,$$

Une inégalité de déviation du même type peut aussi être prouvée. Le cas $\rho = \pi_{\exp(-\beta r)}$ est particulièrement intéressant : les termes d'entropie disparaissent, montrant ainsi que cette « loi de Gibbs » (comme diraient les physiciens) *a posteriori* ne souffre pas de sur-apprentissage : à une constante universelle près, elle a la même performance en espérance et sur l'échantillon observé. De plus la borne inférieure montre que l'encadrement est optimal à un facteur $\left(1 - \frac{2\beta}{N}\right)^{-1}$ près.

Bornes relatives

Une autre amélioration consiste à contrôler $r(\theta) - r(\widetilde{\theta})$, où $\widetilde{\theta}$ est une valeur inconnue du paramètre, par exemple $\arg\min_{\theta\in\Theta_1} R(\theta)$, où Θ_1 est une partie de Θ. On ne contrôle alors pas $R(\theta)$, mais uniquement $R(\theta) - R(\widetilde{\theta})$: dans certaines circonstances, on saura ainsi que l'on se trouve très près du taux d'erreur optimum dans le modèle de classification choisi, sans savoir avec une aussi grande précision quel est ce taux ! Cela se produira par exemple dans le cas binaire bruité où $|\mathcal{Y}| = 2$ et où $P(Y_i = f_{\widetilde{\theta}}(X_i) \mid X_i) = 1 - \alpha$, quand $0 < \alpha < 1/2$.

Voici un exemple d'inégalité en moyenne (une inégalité de déviation de même type est aussi disponible). Considérons une partie Θ_1 de Θ (qui peut éventuellement être égale à Θ tout entier), $\widetilde{\theta} \in \arg\min_{\theta\in\Theta_1} R(\theta)$,

$$\widehat{\theta} \in \arg\min_{\theta\in\Theta} r(\theta),$$

$$d(\theta,\theta') = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\big[f_\theta(X_i) \neq f_{\theta'}(X_i)\big],$$

$$\psi(a) = \sup_{\theta\in\Theta_1} d(\theta,\widehat{\theta}) - a\big[r(\theta) - r(\widehat{\theta})\big],$$

et $g(a) = 2a^{-2}[\exp(a) - 1 - a]$, $a \in \mathbb{R}_+$. Pour tous paramètres réels β et λ tels que $0 \leqslant \beta < \lambda$, toute loi *a posteriori* $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,

$$\int\int R(\theta)\rho(\omega,d\theta)\mathbb{P}(d\omega) \leqslant R(\widetilde{\theta}) + \int \left\{ \frac{\mathcal{K}\left[\rho(\omega),\pi_{\exp(-\lambda r)}\right]}{\lambda-\beta} + \inf_{a,0\leqslant a\leqslant \frac{2N(\lambda-\beta)}{g\left(\frac{2\lambda}{N}\right)\lambda^2}} \frac{g\left(\frac{2\lambda}{N}\right)\lambda^2 \quad (a)}{N(\lambda-\beta)} \right.$$

$$\left. + \left(1 + \frac{g\left(\frac{2\lambda}{N}\right)\lambda^2 a}{2N(\lambda-\beta)}\right)\left[\int r(\omega)\pi_{\exp\left[-(\beta-\frac{g\left(\frac{2\lambda}{N}\right)\lambda^2 a}{2N})r\right]}(d\theta) - r(\widehat{\theta})\right]\right\}\mathbb{P}(d\omega).$$

Cette inégalité fournit une « borne empirique sans biais » permettant de comparer le taux d'erreur moyen de ρ avec celui de la meilleure règle (inconnue) dans Θ_1. On dispose aussi d'une borne théorique correspondante, dans laquelle d est remplacée par D et ψ par $\varphi(a) = \sup_{\theta\in\Theta_1} D(\theta,\widetilde{\theta}) - a\left[R(\theta) - R(\widetilde{\theta})\right]$. Plus précisément

$$\int\int R(\theta)\rho(\omega,d\theta)\mathbb{P}(d\omega) \leqslant R(\widetilde{\theta}) + \inf_{a,0\leqslant a\leqslant \frac{2N(\lambda-\beta)}{g\left(\frac{2\lambda}{N}\right)\lambda^2}} \left(1 - \frac{g\left(\frac{2\lambda}{N}\right)\lambda^2 a}{2N(\lambda-\beta)}\right)^{-1}$$

$$\times \left\{ \frac{\int_\beta^\lambda \left[\int R(\theta)\pi_{\exp(-\gamma R)}(d\theta) - R(\widetilde{\theta})\right] d\gamma}{\lambda-\beta} + \frac{g\left(\frac{2\lambda}{N}\right)\lambda^2\varphi(a)}{2N(\lambda-\beta)} + \frac{\int \mathcal{K}\left[\rho(\omega),\pi_{\exp(-\lambda r)}\right]\mathbb{P}(d\omega)}{\lambda-\beta} \right\}.$$

Cette borne montre que la loi de Gibbs *a posteriori* $\pi_{\exp(-\lambda r)}$ a un taux d'erreur moyen qui peut atteindre dans certains cas des vitesses de convergence vers $\inf_\Theta R$ supérieures à $\sqrt{\frac{R(\widetilde{\theta})}{N}}$ (par exemple dans le cas binaire bruité évoqué plus haut, où $\varphi\left[(1-2\alpha)^{-1}\right] = 0$, la convergence est en $1/N$ dès que $\int R(\theta)\pi_{\exp(-\gamma R)}(d\theta) \leqslant \inf_{\theta\in\Theta} R(\theta) + \frac{c}{N}$, où c est une constante réelle positive).

Echantillon fantôme et bornes de Vapnik

En introduisant un échantillon fantôme $(X_{N+1}, Y_{N+1},\ldots,X_{(k+1)N},Y_{(k+1)N})$, et en utilisant l'échangeabilité de la loi jointe de l'échantillon total $(X_1,Y_1,\ldots,X_{(k+1)N},Y_{(k+1)N})$, on peut montrer des inégalités similaires aux précédentes, dans lesquelles la loi *a priori* π, au lieu d'être fixe, a le droit de dépendre du « design » $(X_1,\ldots,X_{(k+1)N})$, pourvu que cette dépendance soit invariante par permutation des indices. On voit alors, dans cette approche, que seules comptent les restrictions $f_\theta : \{X_i, 1\leqslant i\leqslant (k+1)N\} \to \mathcal{Y}$ des règles de classification à $(k+1)N$ données. Ces restrictions sont en nombre fini (puisque $\mathcal{Y}$ est supposé être un ensemble fini de classes), au plus égal à $|\mathcal{Y}|^{(k+1)N}$. En fait, elles sont souvent bien moins nombreuses, par exemple dans le cas de la classification binaire, quand la famille de règles a une dimension de Vapnik Cervonenkis (dont nous n'avons pas la place de donner ici la définition) inférieure à h, le nombre de règles est inférieur à $\left(\frac{e(k+1)N}{h}\right)^h$ (c'est le cas par exemple de l'ensemble des règles obtenues en séparant $\mathbb{R}^d$ par des hyperplans affines, qui a pour dimension de Vapnik Cervonenkis $h = d+1$). En choisissant π uniforme sur ces règles réduites à l'échantillon total, on ramène le terme d'entropie $\mathcal{K}(\rho,\pi)$ à une valeur maximale de $h\log\left(\frac{(k+1)eN}{h}\right)$ dans les inégalités exposées ci-dessus, y compris lorsque l'on

prend comme loi *a posteriori* $\rho(\omega, d\theta)$ une masse de Dirac. De plus les versions localisées des bornes permettent de réduire le terme d'entropie, voire de l'annuler dans le cas particulier où on considère $\pi_{\exp(-\beta r)}$ comme loi *a posteriori*. Dans ce cadre, les supports vector machines de Vapnik offrent un modèle de classification très intéressant : il consiste, dans le cas binaire, à séparer les données par un hyperplan dans un espace de Hilbert « virtuel » que l'on manipule uniquement à l'aide du « noyau » $K(X_i, X_j)$ qui donne le produit scalaire entre X_i et X_j dans l'espace transformé. Il suffit en fait que la matrice $m_{i,j} = K(X_i, X_j)$ soit symétrique positive pour qu'une telle représentation dans un Hilbert existe, ce qui permet un grand choix de noyaux. En particulier, quand les formes sont représentées initialement dans $\mathbb{R}^d$, on choisit souvent un noyau exponentiel $K(X_i, X_j) = \exp(-\gamma||X_i - X_j||^2)$ qui possède la propriété intéressante d'envoyer les X_i sur des points de la sphère linéairement indépendants les uns des autres dans l'espace transformé. On peut alors séparer dans l'espace transformé les X_i, $1 \leqslant i \leqslant (k+1)N$ de *toutes les manières possibles* : on a fabriqué une représentation linéaire de toutes les règles de classification possibles de l'échantillon total. On peut les ranger en fonction de leur marge, la distance entre l'hyperplan séparateur et le nuage des points transformés des X_i, en pondérant plus fortement sous π les règles de plus forte marge (dans l'approche PAC-Bayesienne). Parmi tous les hyperplans qui séparent les X_i de la même façon on choisira dans cette approche « un hyperplan canonique », c'est-à-dire de marge maximum. Il se trouve que le calcul de cet hyperplan ne fait intervenir que les points placés à distance minimum de l'hyperplan séparateur, appelés *vecteurs de support*. Une autre façon de structurer les modèles consiste à s'appuyer sur le nombre de vecteurs de support. Les support vector machines apparaissent alors comme un cas particulier des « schémas de compression » de Littlestone et Warmuth. En effet, les règles dont la définition ne dépend que de la valeur de h données sur $(k+1)N$ sont au plus au nombre de $\binom{(k+1)N}{h} \leqslant \left(\frac{(k+1)eN}{h}\right)^h$, ce qui permet un contrôle des termes de complexité dans les inégalités dans lequel h joue un rôle similaire à celui de la dimension de Vapnik Cervonenkis. Une règle de classification qui ne dépend que de h données s'appelle un schéma de compression. De tels modèles de règles peuvent être construits de façons extrêmement variées et intuitives. Il suffit pour cela de se poser la question : comment ferais-je pour classer h données ? C'est ensuite cet ensemble restreint de h exemples (appelé ensemble de compression) qui vient paramétrer la famille de règles ainsi construite. En fait, on voit que l'on peut de cette façon construire un schéma de compression à partir de n'importe quelle règle d'apprentissage, en formant la famille de règles obtenue en entraînant la méthode de classification initiale successivement sur tous les sous-ensembles d'apprentissage restreints aux parties à h éléments de l'échantillon de départ. Cette méthode peut en particulier fournir un cadre théorique pour aborder le problème de la sélection et de l'agrégation de caractéristiques (features en anglais). D'autres techniques moins faciles à qualifier sur le plan théorique ont aussi remporté des succès pratiques, comme le boosting, dans lequel on sélectionne pas à pas une suite de combinaisons linéaires seuillées de règles de classification de base, en utilisant un critère pondéré dans lequel le poids des exemples mal classés à une étape augmente à l'étape suivante. On obtient un comportement qui reproduit qualitativement celui des schémas de compression, dans le cas « faiblement bruité » où il y a relativement peu d'exemples mal classés : en effet la règle construite au final dépendra dans ce cas essentiellement d'un petit nombre d'exemples.

Conclusion

L'approche statistique s'est imposée ces dernières années comme l'une des voies les plus prometteuses de l'apprentissage automatique. On dispose en particulier actuellement à la fois de méthodes pratiques (support vector machines, boosting, …) qui donnent des résultats encourageants et d'une (et même plusieurs) théorie mathématique pour les étudier. Il reste néanmoins un certain écart entre la théorie et la pratique, les bornes théoriques ayant tendance à se montrer trop pessimistes par rapport aux performances réellement observées, ce qui limite leur pertinence quand on les utilise pour choisir des modèles de classification. D'autre part la théorie porte essentiellement sur la question de l'apprentissage supervisé, qui n'est, comme nous l'avons mentionné dans l'introduction qu'une partie – centrale mais insuffisante en elle-même – d'une méthode concrète de reconnaissance des formes, qui suppose des prétraitements et des post-traitements des données posant eux-mêmes des problèmes de complexité algorithmique et de choix de représentation difficiles et pouvant dans certains cas se prêter à une analyse mathématique fructueuse que nous n'avons pas la place d'aborder ici.

Pour en savoir plus

[B1] Birgé (L.), Model selection via testing : an alternative to (penalized) maximum likelihood estimators, *preprint PMA-862*, (http ://www.proba.jussieu.fr/prepublications.php) (2003).

[C1] Catoni (O.), Statistical learning theory and stochastic optimization, *Ecole d'été de Probabilités de Saint-Flour XXXI - 2001, J. Picard Ed., Lecture notes in mathematics*, **1851**, pp. 1-272, Springer (2004).

[C2] Catoni (O.), A PAC-Bayesian approach to adaptive classification, *preprint PMA-840* (2003) (http ://www.proba.jussieu.fr/prepublications.php).

[C3] Catoni (O.), Improved Vapnik Cervonenkis bounds, *preprint PMA-942* (2004) (http ://www.proba.jussieu.fr/prepublications.php).

[L1] Littlestone (N.), Warmuth (M.), Relating data compression and learnability, *Technical report*, University of California, Santa Cruz (1986).

[M1] McAllester (D. A.), Some PAC-Bayesian Theorems, *Proceedings of the Eleventh Annual Conference on Computational Learning Theory (Madison, WI, 1998)*, 230-234 (electronic), ACM, New York (1998).

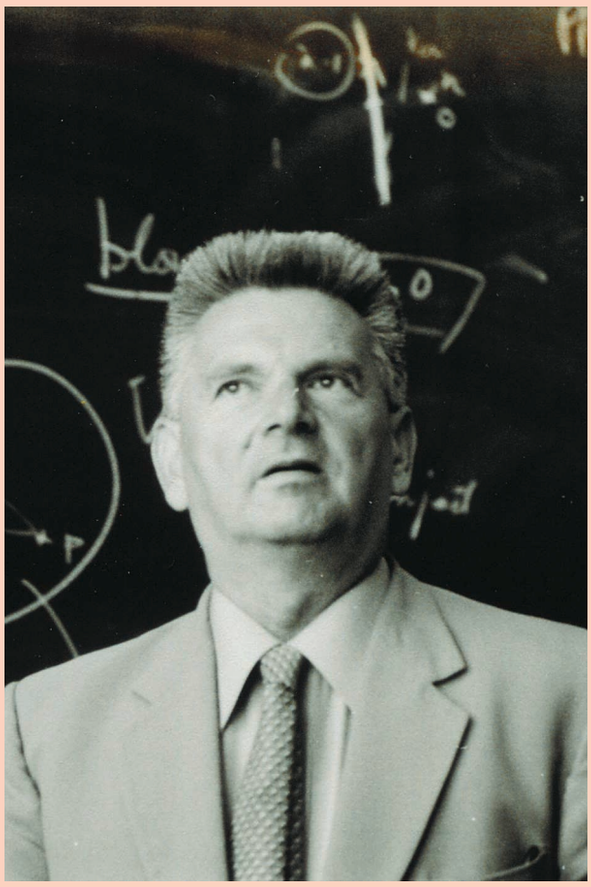
[M2] McAllester (D. A.), PAC-Bayesian Model Averaging, *Proceedings of the Twelfth Annual Conference on Computational Learning Theory (Santa Cruz, CA, 1999)*, 164-170 (electronic), ACM, New York (1999).

[M3] Massart (P.), Concentration inequalities and model selection, Saint-Flour lecture notes, (2003) *Springer*, to appear.

[T1] Tsybakov (A.), Optimal aggregation of classifiers in statistical learning, Annals of Statistics, **32**(1), 2004.

[V1] Vapnik (V. N.), *Statistical learning theory*, Wiley, New York (1998).

René Thom



« Le nouveau Newton est français », proclamait la couverture d'un grand hebdomadaire après la parution en 1972 du livre *Stabilité structurelle et morphogénèse* de René Thom. Ce jugement[1] n'était pas totalement absurde car l'œuvre de Thom, à rebours d'un siècle qui découpait le savoir en rondelles de plus en plus minces, renvoie au temps où les mathématiques n'étaient qu'un aspect de la philosophie naturelle.

Jusqu'à 1970, son activité s'était traduite par un nombre restreint d'articles mathématiques, d'une densité et d'une profondeur exceptionnelles, qui avaient révolutionné la géométrie différentielle et la topologie, lui valant une médaille Fields à trente-cinq ans, en 1958. Avec la rédaction de *Stabilité structurelle et morphogénèse* commence une période plus spéculative, au cours de laquelle il ne touche plus guère aux mathématiques mais s'intéresse à la biologie, à la linguistique et, de plus en plus, à la philosophie, terminant par une exégèse très originale d'Aristote. Voilà pour l'histoire, mais qu'en est-il *au fond* ?

Dans les deux parties si dissemblables de cette œuvre, une même réflexion se poursuit, sur la notion de forme d'abord : qu'est-ce qu'une forme, comment va-t-elle persister par déformation ou au contraire changer, mourir, naître[2] ? C'est sur ces questions qu'est centrée l'œuvre mathématique de Thom, indiscutée parce qu'indiscutable, portant sur des objets idéaux et bien définis : les variétés différentiables (les espaces de la géométrie et de la physique modernes) et les applications entre de telles variétés. Il y fait appel à des idées souvent très simples, qu'il met en œuvre avec l'audace tranquille des visionnaires.

Sur l'autre versant, *Stabilité structurelle et morphogénèse* aborde les mêmes questions, mais du côté des « vraies » formes, et l'audace fait nécessairement place à une certaine témérité car le problème est gigantesque et à peu près inexploré. Comme mise en garde, Thom cite Valéry : « La vie n'a pas le temps d'attendre la rigueur ». Son ouvrage voudrait donc amorcer un dialogue mais se heurte au rejet des biologistes, peut-être indisposés par le tapage médiatique qui l'entoure.

En linguistique, alors que presque tous se concentrent sur l'aspect formel du langage, il part du sens et donc des structures de l'esprit humain, démarche qui, je crois, sous-tend toute son œuvre. Il n'est donc guère étonnant que ce soit dans les neurosciences que certains modèles étudiés aujourd'hui commencent à ressembler à ceux qu'il proposait il y a trente-cinq ans.

Ses œuvres *complètes* font l'objet d'un CD-ROM distribué par l'Institut des Hautes Études Scientifiques.

Marc CHAPERON
Institut de Mathématiques de Jussieu, UMR 7586
Université Paris 7 Denis Diderot, Case Postale 7012
2, place Jussieu, F-75251 Paris Cedex 05
URL: http://www.math.jussieu.fr/ chaperon

[1] Dû à l'enthousiasme du mathématicien britannique E. Christopher Zeeman pour ce qu'il avait baptisé « théorie des catastrophes ».

[2] Les *catastrophes* de Thom et Zeeman, que Vladimir I. Arnold nomme *perestroïkas*.

Les polygones déchaînés et le problème des *n* corps

Alain CHENCINER*

La richesse de la solution la plus triviale du problème des n corps – l'équilibre relatif de n masses égales disposées aux sommets d'un polygone régulier – se révèle si l'on observe globalement et en repère tournant les familles de solutions quasi-périodiques qui en bifurquent dans la direction normale au plan du polygone. Techniquement, l'étude de ce prolongement global se fait en minimisant l'action sous une contrainte de symétrie.

Le problème des *n* corps

Déterminer les mouvements dans l'espace de n masses ponctuelles exerçant l'une sur l'autre une force attractive proportionnelle au produit de leurs deux masses et inversement proportionnelle au carré de leur distance. Tel est le « Problème des n corps », parangon des systèmes « non intégrables » de la mécanique classique dès que $n \geqslant 3$.

Ce n'est que bien des années après les *Principia* de Newton que les équations du mouvement seront écrites sous la forme qu'on leur connaît aujourd'hui : pour $j = 1,\ldots,n$,

$$m_j \ddot{\vec{r}}_j = \sum_{k\neq j} \frac{m_j m_k}{|\vec{r}_k - \vec{r}_j|^3}(\vec{r}_k - \vec{r}_j) = \frac{\partial U}{\partial \vec{r}_j}(\vec{r}_1,\ldots), \qquad (*)$$

où $\vec{r}_j \in \mathbb{R}^3$ et $m_j > 0$ sont la position et la masse du $j^{ème}$ corps,

$$U(\vec{r}_1,\ldots,\vec{r}_n) = \sum_{j<k=1}^{n} \frac{m_j m_k}{|\vec{r}_k - \vec{r}_j|}$$

et où, suivant l'usage des mécaniciens, un point désigne la dérivée temporelle. L'addition de ces équations donne $\sum_{j=1}^n m_j \ddot{\vec{r}}_j = 0$, qui exprime que le centre de gravité $\vec{r}_G = (\sum_{j=1}^n m_j)^{-1} \sum_{j=1}^n m_j \vec{r}_j$ a un mouvement rectiligne uniforme : $\ddot{\vec{r}}_G = \vec{0}$. On choisira un *repère galiléen* dans lequel $\vec{r}_G \equiv \vec{0}$.

Le polygone régulier

Lorsque les masses de tous les corps sont les mêmes (disons égales à 1), il existe pour tout nombre réel positif r une unique fréquence $\omega = \omega(r)$ telle que le mouvement ci-dessous soit une solution périodique de période $T = 2\pi/\omega$ de $(*)$ (on a identifié $\mathbb{R}^2$ au plan complexe et posé $\zeta = e^{2\pi i/n}$) :

$$x(t) = (\vec{r}_1(t),\vec{r}_2(t),\ldots,\vec{r}_n(t)),\ \vec{r}_j(t) = \zeta^j r e^{i\omega t}.$$

* *Université Paris 7 & IMCCE (UMR 8020 du CNRS) 77, avenue Denfert-Rochereau, 75014 Paris.*
chenciner@imcce.fr

En effet, l'attraction qui s'exerce sur le corps j à un instant quelconque est dirigée suivant le vecteur $\vec{r}_j$ puisque ce vecteur porte un axe de symétrie du polygone. De plus, l'intensité de cette force est constante et indépendante de j. Autrement dit, $\ddot{\vec{r}}_j(t) = -\omega(r)^2\vec{r}_j(t)$, ce qui démontre l'assertion. Ces solutions sont les exemples les plus simples d'*équilibres relatifs*, c'est-à-dire de mouvements au cours desquels la configuration tourne rigidement comme un corps solide à vitesse angulaire uniforme. L'existence d'une *forme* est ce qui distingue avant tout le problème des trois (ou $n \geqslant 3$) corps du problème des deux corps et ce n'est pas un hasard si les seules solutions explicites du problème des 3 corps sont les solutions *homographiques* – découvertes par Euler et Lagrange – dans lesquelles ... *les trois Corps pourraient se mouvoir en sorte que leurs distances fussent toujours constantes, ou gardassent au moins entre elles des rapports constants (Lagrange, Avertissement de l'Essai sur le Problème des trois Corps, 1772).* Intimement liées aux symétries du problème (translation, rotation) et à l'homogénéïté de la force newtonienne, ces solutions ne peuvent exister que pour des configurations très spéciales, appelées aujourd'hui c*onfigurations centrales*, celles pour lesquelles la configuration des forces est proportionnelle à celle des corps. Leur détermination lorqu'il y a plus de trois corps est un problème majeur, mais seul le cas « trivial » d'un polygone régulier formé de masses égales va nous intéresser. Notons que, quel que soit le nombre de corps, un mouvement d'équilibre relatif dans $\mathbb{R}^3$ se passe nécessairement dans un plan fixe. Combinées aux forces centrifuges, les forces d'attraction tendent en effet à aplatir la configuration. Ceci n'est plus vrai dans $\mathbb{R}^4$ où un équilibre relatif possède deux axes de rotation orthogonaux.

Minimiser l'action

Les solutions de $(*)$ sont exactement les points critiques de l'*action lagrangienne* qui, à un chemin $[0,T] \ni t \mapsto x(t) = \left(\vec{r}_1(t),\ldots,\vec{r}_n(t)\right)$ fait correspondre l'intégrale

$$\mathcal{A}(x) = \int_0^T \left[\frac{1}{2}\sum_{j=0}^{n} m_j ||\dot{\vec{r}}_j||^2 + U\big(x(t)\big)\right] dt.$$

Cela signifie que $x(t)$ est solution de $(*)$ si et seulement si la variation $\mathcal{A}(x+\delta x) - \mathcal{A}(x)$ de l'action est « du second ordre » par rapport à une variation $\delta x(t)$ du chemin $x(t)$ fixant ses extrémités : c'est le *Principe de moindre action.* Pris au pied de la lettre, ce principe fait rechercher les solutions, non seulement comme points critiques, mais plus précisément comme *minima* de l'action. C'est, au remplacement près de la longueur par l'action, la manière dont les géomètres cherchent une *géodésique* fermée d'un hyperboloïde à une nappe comme une courbe de longueur la plus petite possible parmi celles qui « font le tour » du trou (figure 1).

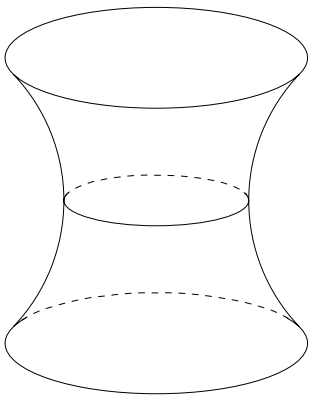


Figure 1 – *Une géodésique minimisante.*

L'action ci-dessus étant toujours positive, cela a un sens de chercher les solutions périodiques de période fixée T de $(*)$ qui minimisent l'action parmi les $x(t)$ périodiques de période T (qu'on appellera des « lacets » de configurations de période T). On peut même s'attendre à ce que les dites solutions soient « les plus simples ». Un théorème célèbre de Tonelli datant d'environ 1925, joint à un résultat de Weierstrass, affirme bien l'existence d'un minimum régulier à condition que soient satisfaites des hypothèses, dites de *coercivité*, assurant qu'un minimum ne peut pas se trouver « à l'infini » (figure 2).

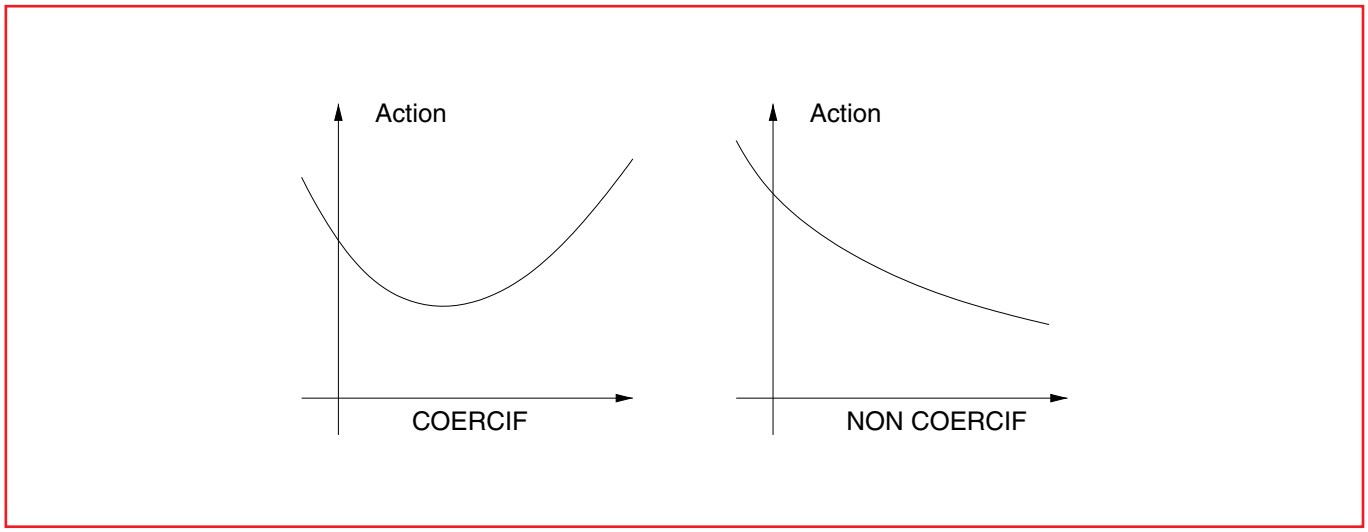


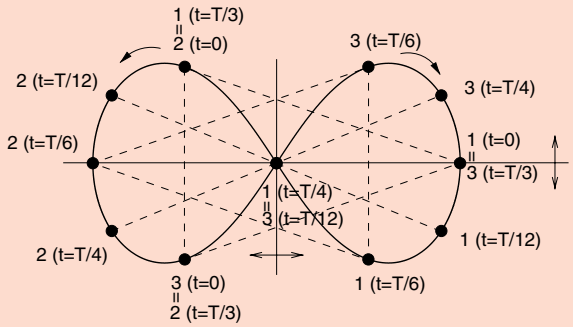
Figure 2 – *Coercivité.*

Malheureusement, c'est exactement ce qui se passe dans notre cas : le minimum absolu de l'action, égal à zéro, est atteint par des corps au repos infiniment éloignés les uns des autres. Eliminer ce problème exige qu'on se restreigne à une classe de lacets soumis à certaines contraintes ; on pense d'abord à des contraintes « topologiques » comme dans l'exemple des géodésiques de l'hyperboloïde mais, comme Poincaré le remarquait déjà dans une note de 1896 [P], la faiblesse de la force newtonienne fait que la minimization sous de telles contraintes conduit bien souvent à des solutions « avec collisions ». Heureusement, il n'en est pas de même avec des contraintes de symétrie. L'exemple le plus simple est la contrainte de *symétrie italienne* : $x(t + T/2) = -x(t)$. Introduite dans les années 80 pour restaurer la coercivité – si un corps s'éloigne, le lacet sera grand puisque ce corps doit, au bout d'une demi-période, occuper une position symétrique par rapport à l'origine : la partie cinétique de l'action sera donc grande elle-aussi – elle exclut également la possibilité de collisions : cela résulte de l'absence de collisions dans les minima de l'action à extrémités fixées, un théorème remarquable de Christian Marchal ([C1]).

Où rien de nouveau n'apparaît

Cherchons donc, dans le cas de trois ou quatre masses égales, les solutions de $(*)$ qui minimisent l'action parmi les lacets de configurations qui d'une part habitent un plan fixe, d'autre part vérifient la symétrie italienne. Il résulte d'un travail avec Nicole Desolneux que les seuls minima sont respectivement l'équilibre relatif du triangle équilatéral et celui du carré. Cela vient de ce que ces deux configurations minimisent la fonction U parmi les configurations planes de taille fixée *(techniquement* : parmi les configurations planes de moment d'inertie I par rapport au centre de gravité fixé). Si l'on admet les configurations spatiales de quatre corps, c'est le tétraèdre régulier qui minimise U à I fixé mais nous avons remarqué à la fin du paragraphe 2 que ce dernier n'a pas de mouvement d'équilibre relatif dans $\mathbb{R}^3$. Par contre, dans $\mathbb{R}^4$, c'est bien un tel mouvement qui minimise l'action parmi les lacets de configurations de quatre corps avec symétrie italienne. En conclusion, la minimisation sous contrainte de symétrie italienne ne fournit rien de nouveau pour 3 et 4 corps de même masse dans le plan ou dans $\mathbb{R}^4$.

Encadré 1

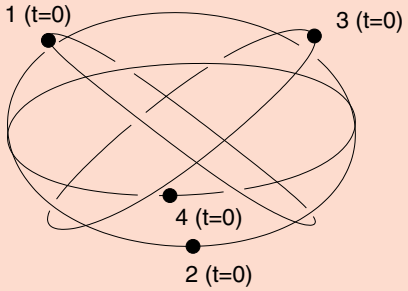


Le « Huit », minimum pour la symétrie $D_6 = \{g_1, g_2 |\, g_1^6 = g_2^2 = 1, g_1 g_2 = g_2 g_1^{-1}\}$:

$$g_1\Big(x_1(t), x_2(t), x_3(t)\Big) = \Big(-\bar{x}_3(t-\tfrac{T}{6}), -\bar{x}_1(t-\tfrac{T}{6}), -\bar{x}_2(t-\tfrac{T}{6})\Big),$$

$$g_2\Big(x_1(t), x_2(t), x_3(t)\Big) = \Big(-x_1(\tfrac{T}{2}-t), -x_3(\tfrac{T}{2}-t), -x_2(\tfrac{T}{2}-t)\Big).$$

La suite exacte $1 \to \mathbb{Z}/3\mathbb{Z} \to D_6 \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \to 1$, *dans laquelle* $\mathbb{Z}/3\mathbb{Z}$ *est engendré par* g_1^2, *implique qu'un lacet de configurations* D_6*-invariant est une* chorégraphie *(les corps subissent une permutation circulaire au bout d'un tiers de période) portée par une courbe symétrique par rapport aux deux axes. La figure est faite dans* $\mathbb{R}^2 \equiv \mathbb{C}$.



Le « Hip-Hop », minimum pour la symétrie italienne $\mathbb{Z}/2\mathbb{Z} = \{g_1\}$: $g_1\Big(x(t)\Big) = -x(t-\frac{T}{2})$, *l'est aussi pour la symétrie* $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} = \{g_1, g_2\}$:

$$g_2\Big(x_1(t), x_2(t), x_3(t), x_4(t)\Big) = \Big(\rho x_4(t), \rho x_1(t), \rho x_2(t), \rho x_3(t)\Big), \text{ où}$$

ρ *est l'isométrie de* $\mathbb{R}^3$ *définie par* $\rho(u,v,w) = (-v,u,-w)$.

Le Hip-Hop est un compromis entre l'équilibre relatif du carré dans $\mathbb{R}^2$ *et celui du tétraèdre régulier dans* $\mathbb{R}^4$. *La figure est faite dans* $\mathbb{R}^3$.

Le Huit et le Hip-Hop

Pour trouver par minimisation de « nouvelles » solutions périodiques, il fallait donc enrichir le groupe de symétrie, ou bien passer du plan à l'espace. L'encadré 1 décrit un exemple de chaque type :

i) passant, pour trois corps, du groupe à 2 éléments $\mathbb{Z}/2\mathbb{Z}$ au groupe dihédral D_6 à 12 éléments, qui est le groupe des symétries de l'« espace des triangles » [C2], on obtient une solution dans laquelle les corps se poursuivent à intervalles de temps égaux le long d'une courbe plane en forme de huit ;

ii) gardant au contraire, pour 4 corps, la symétrie italienne et son groupe à deux éléments mais minimisant l'action parmi tous les lacets de configurations dans l'espace à trois dimensions, on obtient le « Hip-Hop » [C1], [CV] dont la configuration passe continuement du carré au tétraèdre régulier.

Ce qui suit montrera que ces solutions n'étaient, dans un certain sens, pas aussi « nouvelles » qu'elles le paraissaient.

Les polygones se déchainent

(i) L'équation aux variations verticales. La longueur d'un segment de droite ne change pas au premier ordre sous l'influence de variations orthogonales de ses extrémités. Cette conséquence remarquable du théorème de Pythagore implique un scindement de l'*équation aux variations* d'une solution plane en une partie horizontale et une partie verticale : si $x(t)$ est solution de $(*)$, $x(t) + \epsilon y(t)$, où $y_j(t) = (0,0,z_j(t))$, $j = 1,\ldots,n$, est solution au deuxième ordre près en si $z = (z_1,\ldots,z_n)$ vérifie l'*équation aux variations verticales*

$$\ddot{z}_j = \sum_{k \neq j} \frac{m_j m_k}{|\vec{r}_k(t) - \vec{r}_j(t)|^3}(z_k - z_j)\,, \quad j = 1,\ldots,n. \tag{VVE}$$

Pour un équilibre relatif, les coefficients sont indépendants du temps et on montre facilement que, après qu'on ait éliminé la symétrie de translation en exigeant que $\sum m_j z_j = 0$, les solutions sont des combinaisons de solutions de

la forme $z(t) = \mathrm{Re}(Ze^{i\omega_j t})$ où Z est un vecteur propre *complexe* de valeur propre $-\omega_j^2$ de la matrice définissant le second membre de (VVE). Dans le cas de masses égales sur un polygone régulier, les « fréquences verticales » ω_j peuvent être calculées explicitement en fonction des longueurs des diagonales du n-gone régulier. En particulier, ω_1 est la fréquence de l'équilibre relatif considéré. Si $n = 3$, il n'y a qu'une seule fréquence ω_1 ; si $n = 4$, il y en a deux, ω_1 et $\omega_2 = \left(2\sqrt{2}/\sqrt{4+\sqrt{2}}\right)\omega_1$.

(ii) Les familles de Liapunov. Le passage au quotient par les rotations transforme un équilibre relatif en un équilibre. Ce qu'on a dit de l'équation aux variations verticales se traduit par l'existence, pour le champ de vecteurs quotient linéarisé en cet équilibre, d'une décomposition de l'espace des phases vertical » (les $(z,\dot{z})$) en somme directe de sous-espaces de dimensions paires entièrement feuilletés en solutions périodiques. On peut généralement montrer par des techniques classiques de formes normales l'existence locale de surfaces formées de solutions périodiques des équations, appelées *familles de Liapunov*. La période varie en général dans une telle famille, mais, utilisant le fait que si $x(t)$ est une solution de $(*)$, il en est de même de $x_\lambda(t) = \lambda^{-\frac{2}{3}}x(\lambda t)$, on en déduit l'existence locale de familles de solutions à période constante T. Notons que la démonstration de l'existence de ces familles est compliquée par l'existence de résonances avec d'autres fréquences, verticales ou horizontales.

(iii) Les repères tournants. Une façon de « passer au quotient par les rotations d'axe vertical » est d'autoriser des rotations horizontales du repère. Une solution périodique des équations après quotient (les équations *réduites*) apparaitra comme quasi-périodique dans le repère inertiel mais redeviendra périodique dans un repère tournant bien choisi.

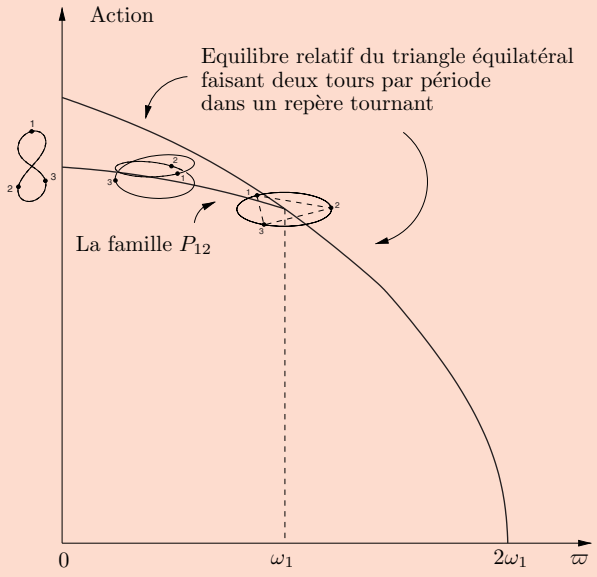
Symétries et bifurcations

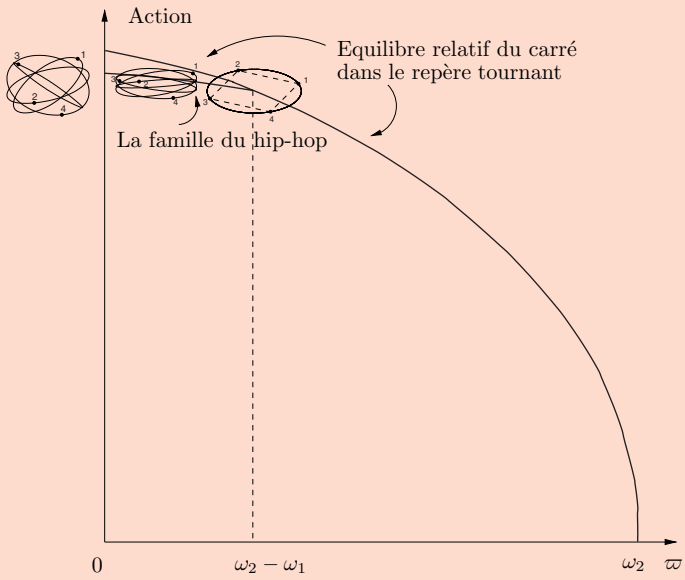
Considérons donc une solution d'équilibre relatif du problème des n corps qui devienne périodique dans un repère tournant dont la vitesse de rotation ϖ, supposée uniforme, va jouer le rôle de paramètre. Si a_C est l'action de l'équilibre relatif $x_C(t) = Ce^{2\pi it}$ de période minimale égale à 1, l'action de l'équilibre relatif correspondant de période T est $a_C T^{\frac{1}{3}}$ et celle de l'équilibre relatif qui parcourt q fois le cercle pendant la période T est $q \times a_C \left(\frac{T}{q}\right)^{\frac{1}{3}} = a_C q^{\frac{2}{3}} T^{\frac{1}{3}}$. Notons $A_C(q,T,\varpi)$ l'action sur un intervalle de temps T de la solution d'équilibre relatif $x_{C,q}^{T,\varpi}(t)$ de configuration normalisée C qui, dans un repère tournant à la fréquence ϖ dans le sens opposé à son mouvement, parcourt q fois le cercle pendant le temps T. Une telle solution est de la forme $x_{C,q}^{T,\varpi}(t) = \lambda^{-\frac{2}{3}}x_C(\lambda t) = \lambda^{-\frac{2}{3}}Ce^{2\pi\lambda it}$. Puisque dans le repère mobile elle devient $\lambda^{-\frac{2}{3}}Ce^{(2\pi\lambda+\varpi)it}$, λ doit être tel que $(2\pi\lambda + \varpi)T = 2\pi q$, c'est-à-dire $\lambda T = q - \dfrac{\varpi}{\omega}$ si l'on note $\omega = \dfrac{2\pi}{T}$. Donc,

$$A_C(q,T,\varpi) = (q - \frac{\varpi}{\omega})^{\frac{2}{3}} T^{\frac{1}{3}} a.$$

Partant de la valeur $\varpi = q\omega$ pour laquelle $A_C(q,T,\varpi) = 0$ (dans le repère inertiel, les particules sont au repos à l'infini), faisons décroître ϖ jusqu'à 0. A chaque valeur ϖ_0 de ϖ telle que l'équation aux variations de $x_{C,q}^{T,\varpi_0}(t)$ possède une solution périodique de période T, le noyau du Hessien de l'action (toujours calculée sur un intervalle de temps T) s'accroît (c'est un raisonnement classique de points conjugués). Afin entre autres de rendre l'action coercive et d'éliminer les solutions « triviales » correspondant à la rotation du plan de l'équilibre relatif ou aux solutions homographiques, choisissons une telle solution $z_0(t)$ *verticale* et minimisons l'action parmi les chemins qui, *dans le repère tournant à la vitesse* ϖ_0, *deviennent un lacet possédant les mêmes symétries que le lacet* $\left(x_{C,q}^{T,\varpi_0}(t), z_0(t)\right)$. Notons que les solutions $\left(x_{C,q}^{T,\varpi_0}(t), 0\right)$ possèdent toujours une telle symétrie, qui est une brisure de

Encadré 2





leur symétrie continue. Dans les deux cas que nous considérons ci-dessous, la contrainte de symétrie est assez forte pour que ϖ_0 soit l'unique point de bifurcation de la famille (encadré 2). Le minimum de l'action est alors réalisé par $x_{C,q}^{T,\varpi}(t)$ lorsque $\varpi_0 \leqslant \varpi \leqslant q\omega$, et par une solution décrivant l'une des familles de Liapunov évoquées plus haut lorsque $0 \leqslant \varpi \leqslant \varpi_0$.

Du triangle au Huit, du carré au Hip-Hop

i) Du triangle au Huit : Dans le cas de l'équilibre relatif équilatéral $x_C(t)$ de trois corps, chaque solution de (VVE) a la même fréquence ω_1 que $x(t)$. L'espace des phases de (VVE), de dimension 4 une fois éliminées les translations, est engendré par la famille – triviale – des équilibres relatifs dans des plans inclinés et par une famille dérivant d'une solution $z_0(t)$. Dans un repère tournant d'un tour complet dans le sens rétrograde pendant une période (*i.e* ; $\varpi_0 = \omega_1$), $\left(x_{C,2}^{T,\varpi_0}, z_0(t)\right)$ devient une chorégraphie portée par une courbe en forme d'huitre entrebaillée et admet le groupe diédral D_6 à 12 éléments comme groupe de symétrie. La minimisation sous cette contrainte de symétrie fournit, pour ϖ variant de $\varpi_0 = \omega_1$ à 0, une famille de chorégraphies dans le repère tournant portées par une courbe qui s'ouvre comme une huitre et aboutit au « Huit » dans un plan vertical. C'est la famille P_{12} ou « déchainement » du triangle. Le germe de la famille de Liapunov associée aux solutions non triviales de (VVE) était décrit dans [M1] comme la famille de solutions (quasi) périodiques du problème des trois corps avec la symétrie maximale (qui est celle de D_6) ; c'est immédiatement après qu'il ait pris connaissance de l'existence du « Huit » que C. Marchal a remarqué [M2] que ces solutions devenaient des chorégraphies dans le repère tournant. C'est le point de départ de [CF], où nous donnons à cette remarque toute sa portée.

ii) du carré au Hip-Hop : Considérées dans un repère qui tourne à la vitesse $\varpi = (\omega_2 - \omega_1)$, les solutions de (VVE) qui correspondent à la deuxième fréquence $\omega_2 > \omega_1$, deviennent $\frac{2\pi}{\omega_2}$-périodiques et possèdent les mêmes symétries que le Hip-Hop. Elles donnent naissance à une famille qui, lorsque la rotation décroît, grandit et se termine en la solution de Hip-Hop (voir [CF] où l'on trouvera également une discussion du cas de 5 corps ; l'absence de collisions dans la minimisation est prouvée dans [TV]).

Dans les deux cas, la seule chose qui n'est pas prouvée (mais qui est claire numériquement) est l'unicité du minimum pour chaque valeur de la rotation, unicité qui impliquerait la continuité de la famille et pas seulement de l'action.

Pour en savoir plus

[C1] Chenciner (A.), Solutions du problème des *n* corps joignant deux configurations, *Gazette des mathématiciens* **99**, 5-12, janvier (2004).

[C2] Chenciner (A.), De l'espace des triangles au problème des trois corps, *Gazette des mathématiciens* **104**, 22-38, (avril 2005).

[CF] Chenciner (A.) & Féjoz (J.), L'équation aux variations verticales d'un équilibre relatif comme source de nouvelles solutions périodiques du problème des N corps, *CRAS*, **340**, n° 8, 593–598 (15 Avril 2005).

[CM] Chenciner (A.) & Montgomery (R.), A remarkable periodic solution of the three-body problem in the case of equal masses, *Annals of Mathematics* **152**, 881–901 (2000).

[CV] Chenciner (A.) & Venturelli (A.), Minima de l'intégrale d'action du Problème newtonien de 4 corps de masses égales dans $\mathbb{R}^3$: orbites « hip-hop », *Celestial Mechanics* **77**, 139-152 (2000).

[M1] Marchal (C.), The three-body problem, *Elsevier* (1990).

[M2] Marchal (C.), The family P_{12} of the three-body problem. The simplest family of periodic orbits with twelve symmetries per period, *Cel. Mech. Dynam. Astron*. **78**, 279-298 (2000).

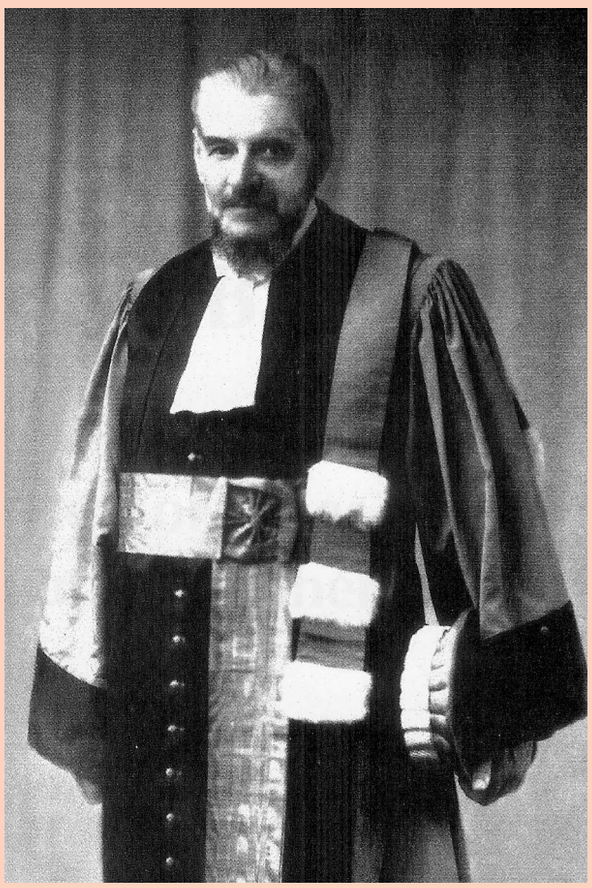
[P] Poincaré (H.), Sur les solutions périodiques et le principe de moindre action, *C.R.A.S*. **123**, 915-918, (1896).

[TV] Terracini (S.) & Venturelli, Symmetric trajectories for the 2N-body problem with equal masses, *to appear in ARMA*.

L'auteur remercie chaleureusement Jacques Féjoz de son aide pour les figures.

Louis Bachelier
11 mars 1870 – 28 avril 1946

Incroyable destinée, pour un mathématicien contemporain, que celle de Bachelier, qui fut d'abord déconsidéré, qui obtint son premier poste fixe à l'université (de Besançon) à 57 ans, et ne devint célèbre que 20 ans après sa mort. L. Bachelier naquit au Havre, dans une famille habituée aux affaires bancaires et commerciales. Il fit des études sans éclat, puis soutint son doctorat sous la direction de Poincaré le 29 mars 1900, avec seulement la mention « honorable », sous le titre « Théorie de la spéculation ». Ce travail, qui a pour objet l'application du calcul des probabilités aux opérations de la Bourse, est suivi de nombreux autres notes, mémoires et ouvrages originaux sur les probabilités, rédigés malgré la situation très précaire de leur auteur.



La thèse de Bachelier contient, et cela de trois façons différentes, la première théorie mathématique du mouvement brownien (cinq ans avant Einstein). Ainsi, en utilisant une terminologie moderne, le mouvement brownien est vu tour à tour comme le processus à accroissements indépendants et homogènes dont les trajectoires sont continues, comme le processus à temps continu limite de marches au hasard symétriques, et enfin comme le processus de Markov dont l'équation « forward » de Kolmogorov est l'équation de la chaleur. Bachelier procède à une étude « fine » des trajectoires du mouvement brownien trente ans avant Paul Lévy, à l'aide du principe de réflexion, de la propriété de Markov forte.

On peut interpréter cette thèse comme le confluent de deux traditions apparemment très éloignées. La première, qui sert de ligne directrice mathématique à son auteur, est celle de la physique mathématique française de J. Fourier, de G. Lamé et évidemment de H. Poincaré. Bachelier tire d'ailleurs explicitement ses analogies (tel le rayonnement de la probabilité) de ces idées. La seconde, ce sont les modèles de raisonnement tacites des spéculateurs en Bourse qu'on trouve, sous des expressions certes moins formalisées, dans la seconde moitié du XIX^{e} siècle. Ainsi un ouvrage de Jules Regnault de 1863 renferme-t-il déjà, en termes plus « littéraires », le cadre conceptuel de l'application du calcul des probabilités aux opérations de Bourse, notamment le fait que « l'écart pris sur un grand nombre d'opérations est en raison directe de la racine carrée du temps ».

Bachelier a été méprisé, fort peu lu et encore moins compris, et il en a beaucoup souffert : Gevrey et P. Lévy ont cru (à tort) qu'il s'était grossièrement trompé ; seul Kolmogorov a reconnu la profondeur de ses travaux, dans les années trente. Or, non seulement, sa thèse était remarquable, mais ses recherches ultérieures l'étaient quelquefois encore davantage et passèrent à peu près inaperçues, malgré un certain soutien de Poincaré.

Par exemple, son mémoire de 1906 donne-t-il les définitions de grandes classes de processus aléatoires apparus par la suite : processus à accroissements indépendants, processus de Markov, processus d'Ornstein-

Uhlenbeck. Ces définitions apparaissent comme conséquences d'une théorie plus générale : celle des équations différentielles stochastiques que Bachelier développe ici, sans toute la rigueur à laquelle nous sommes habitués maintenant, à l'aide d'un vocabulaire issu des jeux de hasard. Deux fonctions jouent un rôle central dans son mémoire, la première, appelée espérance relative à une partie est le terme de « drift » de 1'équation différentielle stochastique, alors que la seconde, appelée fonction d'instabilité relative à une partie, est évidemment le coefficient de diffusion de cette même équation. Bachelier raisonne plutôt en trajectoire : ses équations définissent bien un mouvement, en cela il fait penser à Langevin (1908), puis à Ito et Lévy, plus tard.

L. Bachelier a en outre introduit une théorie intéressante de la « probabilité inverse » et de la « probabilité des causes », c'est-à-dire de l'estimation statistique, qu'il a d'ailleurs reprise dans un traité de calcul des probabilités, aussi clair que remarquable, publié en 1912.

N.B. Pour toutes précisions, voir Courtault J.-M. & Kabanov Y. (dir), Louis Bachelier. Aux origines de la finance mathématique, Besançon, Presse de l'université de Franc-Comtoises 2002.

L. CARRARO, P. CRÉPEL

Pour en savoir plus

BACHELIER (L.), « Théorie de la spéculation », *Annales de l'Ecole Normale Supérieure*, **17**, p. 21-86 (1900).

BACHELIER (L.), « Théorie des probabilités continues », *Journal de mathématiques pures et appliquées*, **2,** p. 259-327 (1906).

BACHELIER (L.), *Calcul des probabilités*, Gauthier-Villars, Paris (1912).

GOBET (E.), Les mathématiques appliquées au coeur de la finance, *Images des mathématiques* (2004).

REGNAULT (J.), *Calcul des chances et philosophie de la Bourse*, Mallet-Bachelier, Paris (1863).

Ondes en milieu aléatoire

Josselin GARNIER*

Un milieu naturel tel que la croûte terrestre a souvent des propriétés physiques possédant des variations spatiales compliquées ou partiellement connues. Il peut alors être modélisé comme une réalisation d'un milieu aléatoire. Lorsqu'une onde se propage dans un tel milieu, on ne peut souvent donner qu'une description statistique de l'onde. Mais parfois, on peut trouver un résultat de nature déterministe : la quantité observée au cours d'une expérience ne dépend que de la statistique du milieu, et pas de la réalisation particulière. Une telle quantité est dite auto-moyennisée, et sa stabilité statistique la rend très attractive pour des applications en imagerie notamment. Un tel phénomène est possible lorsque plusieurs échelles distinctes et bien séparées sont présentes : longueur d'onde, taille des inhomogénéités, distance de propagation.

La propagation d'une onde dans un milieu inhomogène est un problème complexe, dont l'étude passe par une modélisation stochastique du milieu et la détermination des échelles caractéristiques du problème.

Milieu aléatoire. On modélise un milieu inhomogène comme une réalisation d'un milieu aléatoire. Cela veut dire que les évolutions des paramètres physiques du milieu en fonction de l'espace sont décrits par des processus aléatoires. La propagation d'une onde dans un tel milieu est modélisée par une équation aux dérivées partielles à coefficients aléatoires. Cette approche stochastique peut être justifiée *a priori* par les arguments suivants :

1) En certaines circonstances, comme par exemple en géophysique, on ne dispose que de données partielles sur le milieu (la croûte terrestre) dans lequel les ondes se propagent. Dans ce cas, l'approche stochastique vise à modéliser le manque d'information. La modélisation stochastique prend en compte les données disponibles (moyennes, spectres, ...) et complète ces données en utilisant une description statistique.

2) En d'autres circonstances, on pourrait disposer d'une description complète du milieu mais celle-ci serait si compliquée et ferait intervenir tellement d'échelles différentes qu'il serait impossible de résoudre le problème complet, de manière analytique ou numérique. La modélisation d'un tel milieu comme une réalisation d'un milieu aléatoire peut simplifier énormément l'analyse par l'application de théorèmes limites pour des équations à coefficients aléatoires.

Enfin, l'approche stochastique peut se justifier *a posteriori* par la pertinence des résultats qu'elle permet d'obtenir. En particulier, on verra qu'on peut exhiber des quantités auto-moyennisées, dont le comportement est statistiquement stable dans le sens où il dépend seulement de la statistique du milieu, et pas de la réalisation particulière du milieu.

Echelles. Un point essentiel de l'étude consiste à appréhender les différentes échelles caractéristiques du problème, c'est-à-dire les distances typiques sur lesquelles varient les coefficients qui interviennent. Quand on pousse à la limite les rapports entre ces échelles, on peut obtenir un régime asymptotique remarquable. Ainsi le travail se décompose en trois tâches intimement liées. Tout d'abord vient la phase de modélisation, puis des théorèmes limites entrent en jeu. Enfin on tente d'identifier la limite de la manière la plus simple possible, souvent à travers des lois de processus de diffusion.

Laboratoire de Probabilités et Modèles Aléatoires : UMR 7599
Laboratoire Jacques-Louis Lions : UMR 7598 Université Paris VII.
garnier@math.jussieu.fr

Propagation d'ondes dans un milieu inhomogène

On peut distinguer trois échelles de longueur dans un problème de propagation d'ondes en milieu aléatoire : la *longueur d'onde* λ (*i.e.* la largeur du support initial de l'onde), la *distance de propagation* L, et la *taille des inhomogénéités* l_c. L'échelle L peut aussi être l'échelle des variations macroscopiques du milieu (les couches géologiques en géophysique). L'identification de l_c n'est pas toujours facile, mais dans la modélisation stochastique on peut définir l_c précisément comme une *longueur de corrélation*. Je vais me concentrer ici sur le régime le plus couramment rencontré en géophysique. Si on prend les chiffres donnés dans [1], la longueur d'onde $\lambda \sim 100$ m est petite comparée à la taille des couches géologiques $L \sim$ 1-50 km, mais grande comparée à la longueur de corrélation du milieu $l_c \sim 2$ m. On se trouve donc dans le régime où $l_c \ll \lambda \ll L$. C'est un régime particulièrement intéressant d'un point de vue mathématique car c'est une limite haute fréquence par rapport à la distance de propagation, mais c'est une limite basse fréquence par rapport aux fluctuations du milieu.

Pour fixer les choses, nous allons étudier ici l'équation qui régit la propagation des ondes acoustiques en milieu uni-dimensionnel :

$$\partial_t^2 p(t,z) - \partial_z \left[K(z)\rho^{-1}(z)\partial_z p(t,z) \right] = 0 \,.$$

Une onde acoustique est caractérisée par un champ de pression p. Le milieu est caractérisé par deux paramètres : la densité ρ et le module d'incompressibilité K. Dans le cas d'un milieu homogène, les paramètres du milieu ρ et K sont constants. On est alors ramené à l'équation $\partial_t^2 p - c^2 \partial_z^2 p = 0$, qui est l'équation d'ondes standard avec la vitesse de propagation (ou *vitesse du son*) $c = \sqrt{K/\rho}$. La solution générale, connue sous le nom de solution de d'Alembert, est de la forme $p(t,z) = a(z-ct) + b(z+ct)$. Cela veut dire qu'une condition initiale arbitraire donne naissance à deux ondes, une qui se propage vers la droite (a), et une qui se propage vers la gauche (b) avec la vitesse c. En choisissant bien les conditions initiales, on peut générer une onde pure qui se propage vers la droite, sans déformation et à vitesse constante.

Dans un milieu inhomogène, les deux paramètres du milieu ρ et K sont fonctions de la coordonnée spatiale z. Ceci change énormément la propagation d'une onde. La figure 1 est le résultat d'une simulation numérique de pro-

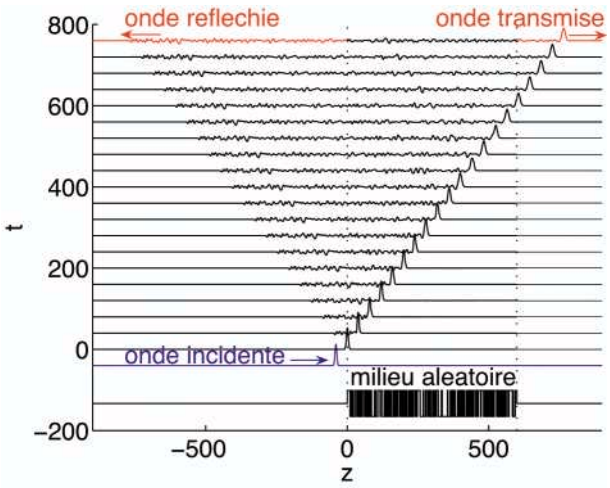


Figure 1 – *Propagation d'une impulsion de forme initiale gaussienne (en bleu en bas) dans une couche de milieu aléatoire qui occupe l'intervalle* $[0,L]$, $L = 600$. *Le profil de densité est dessiné en bas, il résulte d'une alternance de couches d'épaisseurs variables et de densité* $1 \pm \Delta\rho$ *avec* $\Delta\rho$ *: Les profils spatiaux du champ de pression sont dessinés pour une suite d'instants multiples de* 40 *(la vitesse moyenne est* 1*).*

pagation d'une onde dans une couche de milieu aléatoire. A chaque instant, le tracé du signal montre qu'on peut distinguer deux parties :

1) un *front cohérent*, d'amplitude importante et de support étroit, qui garde plus ou moins la forme de l'onde originale,

2) des *ondes incohérentes* (appelées coda en géophysique), de faible amplitude mais dont le support s'accroît au cours du temps, qui sont le résultat de l'interaction de l'onde avec les inhomogénéités du milieu.

Au tout début de la propagation, le front est nettement dominant et il se propage sans changement notable, en émettant un petit train d'ondes qu'on peut pour un temps négliger. En fait, le front se propage comme s'il était dans un milieu homogène. Le calcul des paramètres homogénéisés de ce milieu fictif, et en particulier de la vitesse du son homogénéisée, est discutée section 2.

Au fur et à mesure de sa propagation, le front s'atténue et s'étale. On verra dans la section 3 que cette déformation est parfaitement prévisible et calculable dans le sens où elle ne dépend pas de la réalisation du milieu, mais seulement de ses propriétés moyennes statistiques. C'est le premier phénomène statistiquement stable que nous rencontrerons.

Par conservation de l'énergie totale de l'onde, l'atténuation du front est concominante avec une augmentation de l'énergie des petites ondes incohérentes. On donnera la description statistique des ondes incohérentes en section 4.

Enfin, on expliquera un autre phénomène statistiquement stable dans la section 5 : la refocalisation de l'onde retournée temporellement. En effet, même si les petites ondes incohérentes semblent avoir perdu toute trace de cohérence et ne pas pouvoir apporter d'information utile, ni sur la source originale, ni sur des grandeurs physiques du milieu, on peut régénérer à partir d'elles une onde cohérente de laquelle on peut tirer beaucoup d'information.

Vitesse effective de propagation

Lorsque les inhomogénéités sont de petite taille, *i.e.* $l_c \ll \lambda$, et la distance de propagation pas trop grande, *i. e.* $L \sim \lambda$, le front d'onde est nettement dominant et se propage comme s'il se trouvait dans un milieu homogène, caractérisé par des coefficients homogénéisés. En particulier le front se déplace à une vitesse constante, dont la valeur peut être obtenue par un théorème d'*homogénéisation*. Comme dans l'appendice où on étudie le mouvement d'une particule, on trouve que l'onde se propage à une vitesse qui s'obtient par une procédure de moyennisation des paramètres du milieu aléatoire. Toute la difficulté réside dans le calcul de ces coefficients homogénéisés, et l'analyse montre que les bons coefficients sont $\bar{\rho} = \mathbb{E}[\rho]$ et $\bar{K} = \left(\mathbb{E}[K^{-1}]\right)^{-1}$. Ainsi, la *vitesse effective* du front est $\bar{c} = \sqrt{\bar{K}/\bar{\rho}}$.

Exemple : *des bulles d'air dans de l'eau. L'air et l'eau sont caractérisés par les paramètres suivants :*

$$\rho_a = 1.2\ 10^3\ \text{g/m}^3,\ K_a = 1.4\ 10^8\ \text{g/s}^2\text{/m},\ c_a = 340\ \text{m/s}.$$

$$\rho_e = 1.0\ 10^6\ \text{g/m}^3,\ K_e = 2.0\ 10^{18}\ \text{g/s}^2\text{/m},\ c_e = 1425\ \text{m/s}.$$

Considérons un son audible, de longueur d'onde typique d'ordre 10 cm-100 m. *Les bulles d'air étant beaucoup plus petites, le résultat d'homogénéisation peut être appliqué. Si la proportion volumique d'air dans l'eau est* ϕ, *alors les coefficients homogénéisés sont*

$$\bar{\rho} = \mathbb{E}\left[\rho\right] = \phi\rho_a + (1-\phi)\rho_e = \begin{cases} 9.9\ 10^5 \text{g/m}^3 & \text{si } \phi = 1\% \\ 9\ 10^5 \text{g/m}^3 & \text{si } \phi = 10\% \end{cases}$$

$$\bar{K} = \left(\mathbb{E}\left[K^{-1}\right]\right)^{-1} = \left(\frac{\phi}{K_a} + \frac{1-\phi}{K_e}\right)^{-1} = \begin{cases} 1.4\ 10^{10} \text{g/s}^2\text{m} & \text{si } \phi = 1\% \\ 1.4\ 10^{9} \text{g/s}^2\text{/m} & \text{si } \phi = 10\% \end{cases}$$

En conséquence $\bar{c} = 120$ m/s *si* $\phi = 1$ % *et* $\bar{c} = 37$ m/s *si* $\phi = 10$ %.

Cet exemple montre que la vitesse homogénéisée peut être beaucoup plus petite que le minimum des vitesses des composants du milieu inhomogène. L'inverse (dans le cas des ondes acoustiques) n'est pas possible, dans le sens où la vitesse homogénéisée ne peut pas être plus grande que le maximum (ou le sup essentiel) des vitesses des composants. En effet $\mathbb{E}\,[c^{-1}] = \mathbb{E}\,\left[K^{-1/2}\rho^{1/2}\right] \leqslant \mathbb{E}\,[K^{-1}]^{1/2}\mathbb{E}\,[\rho]^{1/2} = \bar{c}^{-1}$ et donc $\bar{c} \leqslant \mathbb{E}\,[c^{-1}]^{-1} \leqslant \text{ess sup}(c)$.

Note. La théorie de l'homogénéisation dépasse largement le cadre discuté ici. Elle s'applique pour calculer les propriétés effectives des matériaux composites en mécanique, en élasticité, en électromagnétisme, etc. On en trouve des versions valables pour des milieux aléatoires, périodiques ou quasi-périodiques [4].

Propagation du front cohérent

Le résultat d'homogénéisation prédit que le front cohérent se propage à la vitesse constante $\bar{c}$ et sans déformation. Ce résultat néglige les petites ondes incohérentes qui sont générées au fur et à mesure de la propagation, ce qui est valable tant que la distance de propagation reste de l'ordre de grandeur de la longueur d'onde typique $L \sim \lambda$. Pour des distances de propagation plus grandes, $L \gg \lambda$, l'émission de ces ondes incohérentes ne peut plus être négligée dans l'analyse de la dynamique du front. On change alors de régime, et on utilise des résultats d'approximation-diffusion. Après une distance de propagation L telle que $l_c \ll \lambda \ll L$ et $Ll_c \sim \lambda^2$, le profil temporel du front est de la forme

$$K_L * f(t - T_L) \tag{1}$$

– T_L est un retard temporel *aléatoire* qui suit une loi gaussienne de moyenne $L/\bar{c}$ et de variance $\mathbb{E}\,[T_L^2] = \alpha_1 L$ où α_1 est proportionnel à la longueur de corrélation l_c du milieu et ne dépend que de la fonction de covariance du milieu (statistique à deux points). On peut voir sur la figure 2a qu'effectivement, deux réalisations différentes du milieu donnent lieu à deux fronts qui sont décalés temporellement.

– K_L est un noyau de convolution gaussien *déterministe* :

$$K_L(t) = \frac{1}{\sqrt{2\pi D_L^2}} \exp\left(-\frac{t^2}{2D_L^2}\right) \tag{2}$$

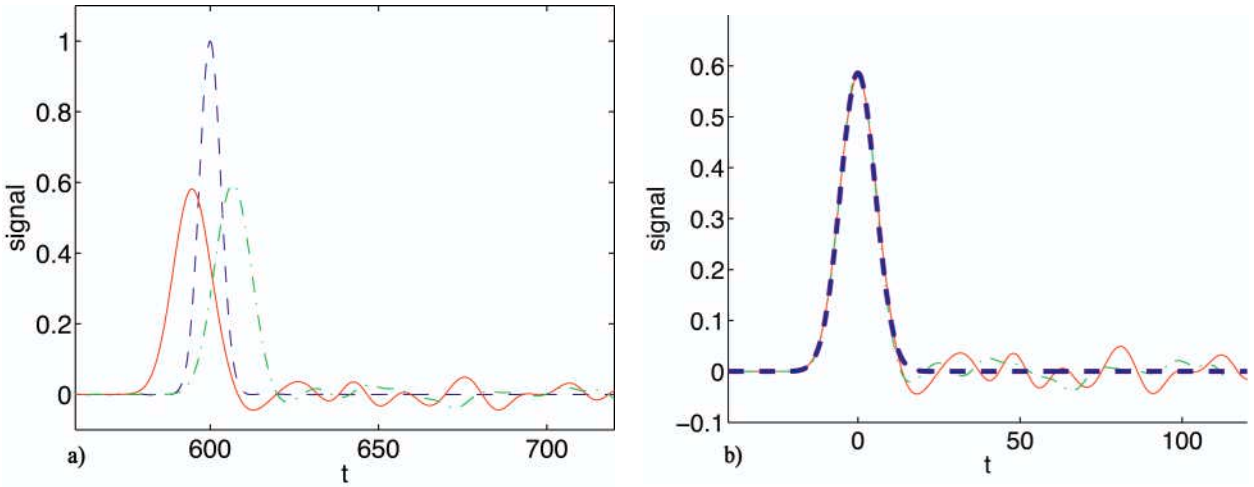


Figure 2 – *Répétition de la simulation de propagation de la figure 1 avec deux réalisations différentes du milieu. Sur la figure a, on dessine les signaux temporels obtenus à la sortie de la couche en $z = L(= 600)$. Sur la figure b, après translation temporelle des signaux, on compare les fronts cohérents obtenus avec la formule théorique (1).*

où $D_L = \alpha_2 L$ et α_2 est proportionnel à la longueur de corrélation l_c du milieu. Ainsi la forme du front cohérent ne dépend pas de la réalisation du milieu, mais seulement de sa statistique. On peut vérifier sur la figure 2b qu'effectivement, la forme du front est parfaitement prédite par la formule (1).

Note. La remarquable stabilité du front cohérent a été mise en évidence par des géophysiciens, O'Doherty et Anstey, dans les années 1970. Il a fallu attendre le milieu des années 1990 pour en avoir une démonstration mathématique [3].

Statistique des ondes incohérentes

L'énergie perdue par le front cohérent se retrouve dans les ondes incohérentes. Cette partie de l'onde souvent négligée est importante en pratique, car on n'a pas toujours accès au front d'onde : ou bien la distance de propagation est si grande que le front s'est complètement évanoui, ou bien, comme c'est souvent le cas en géophysique, on a seulement accès à l'onde réfléchie. Cette onde est uniquement constituée de fluctuations incohérentes, comme on peut le voir sur la figure 1. Apparemment, les informations macroscopiques sur le milieu ou sur la source originale semblent perdues. Les résultats théoriques montrent que ce n'est pas tout-à-fait le cas : on peut trouver dans les propriétés locales du signal réfléchi de l'information. Plus précisément, localement autour d'un instant t_0 d'ordre $L/\bar{c}$, et dans l'échelle de temps de la source originale t d'ordre $\lambda/\bar{c}$, le signal réfléchi est un processus gaussien stationnaire de moyenne nulle et de fonction d'autocorrélation $c_{t_0}(t) = \mathbb{E}\,[p(t_0+t)p(t_0)]$ donnée par

$$c_{t_0}(t) = \frac{1}{2\pi}\int |\hat{f}(\omega)|^2 \Lambda(t_0,\omega) e^{-i\omega t} d\omega \qquad (3)$$

Ici la densité de puissance spectrale Λ est donnée à travers un système d'équations de transport hyperbolique déterministe qui ne dépend que des propriétés macroscopiques et statistiques du milieu [1]. Ce résultat est obtenu par l'application de plusieurs techniques : immersion invariante, analyse de Fourier, analyse stochastique, approximation-diffusion.

Le résultat précédent a des conséquences importantes en imagerie, lorsqu'on cherche à identifier les variations macroscopiques du milieu à partir de l'étude des signaux réfléchis. D'une part, la qualité gaussienne des ondes réfléchies montre que toute l'information sur le milieu est contenue dans la fonction d'autocorrélation, ou de manière équivalente dans la densité Λ. D'autre part, le système d'équations de transport permet, à partir de la connaissance de Λ, de reconstruire les propriétés macroscopiques du milieu. Le problème majeur est en fait l'estimation de la fonction d'autocorrélation, qui s'exprime sous la forme d'une moyenne statistique, c'est-à-dire une moyenne sur le milieu. Or on ne dispose bien souvent que d'une seule réalisation du milieu, et les meilleurs estimateurs statistiques des covariances locales (par transformé de Fourier à fenêtre ou par transformée en ondelettes) se révèlent peu performants. On va voir dans la prochaine section qu'un remarquable estimateur peut être obtenu par retournement temporel des ondes.

Note. Les ondes incohérentes réfléchies sont l'objet de recherches intenses motivées par des problèmes de recherche pétrolière. En théorie, en envoyant un son dans la terre, et en écoutant le signal réfléchi, on doit pouvoir déterminer la structure du sous-sol. Il y a encore beaucoup de progrès à faire dans ce domaine, qui touche l'analyse théorique et numérique, les probabilités et le traitement du signal [1,3].

Retournement temporel des ondes

D'un point de vue expérimental, un miroir à retournement temporel (MRT) est un réseau de transducteurs reliés chacun à une mémoire. Chaque transducteur a deux modes de fonctionnement. Il peut être utilisé comme un microphone, le signal acoustique reçu étant alors stocké dans la mémoire associée. Il peut aussi être utilisé comme un émetteur. Supposons qu'on utilise un tel appareil pour enregistrer le signal réfléchi obtenu lors de la simulation de la figure 1. Ce signal est retourné temporellement dans les mémoires, et le MRT renvoie dans le milieu le signal retourné (figure 3). On observe alors un phénomène remarquable. On voit ressortir une onde cohérente dont la forme est parfaitement calculable :

$$K_{\mathrm{RT}} * f(-t) \qquad (4)$$

où la transformée de Fourier du noyau de refocalisation *déterministe* est

$$\hat{K}_{\mathrm{RT}}(\omega) = \int\int \Lambda(\tau,\omega) G(\tau) e^{-i\omega\tau} d\omega d\tau \qquad (5)$$

Ici G est la fonction de troncation qui délimite la fenêtre temporelle d'enregistrement (sur l'expérience des figures 1-3, on a simplement la fonction indicatrice $G(t) = \mathbf{1}_{[0,640]}(t)$). Le point important est que cette refocalisation est statistiquement stable : la forme de l'impulsion refocalisée ne dépend que des propriétés statistiques du milieu (du noyau Λ), et pas de la réalisation particulière. On peut voir figure 4 qu'effectivement, une répétition de l'expérience avec une nouvelle réalisation du milieu conduit exactement au même résultat. Cette stabilité statistique est très importante pour des problèmes d'imagerie : en comparant l'impulsion refocalisée à la source originale, pour différentes fonctions de troncation G, on dispose d'un estimateur stable du noyau Λ qui permet de reconstruire les propriétés macroscopiques du milieu.

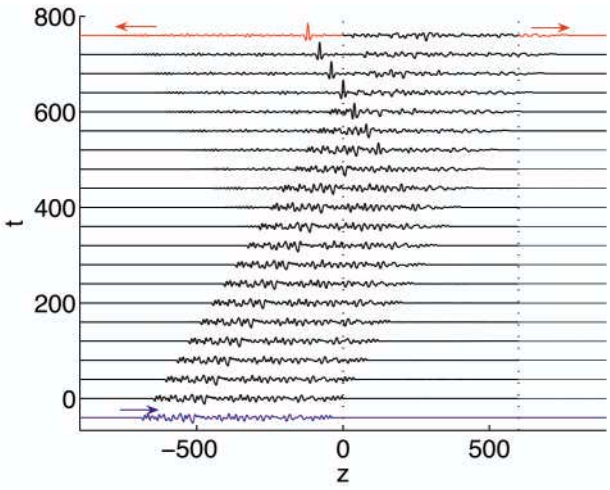
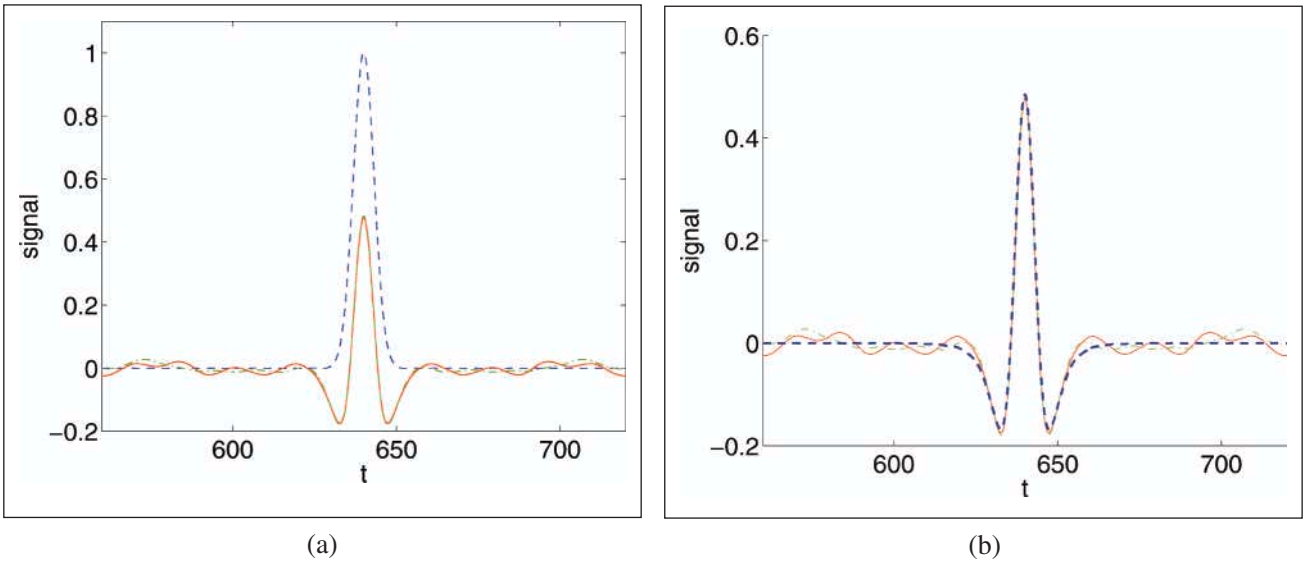


Figure 3 – *Retournement temporel de l'onde réfléchie (amplifiée par un facteur* 1.5*) enregistrée pendant la simulation de l'expérience de la figure 1.*



(a) (b)

Figure 4 – *Répétition de la simulation de retournement temporel avec deux milieux différents. On compare les signaux refocalisés obtenus avec le signal original (a) et avec le signal refocalisé prédit par la formule (4) (b).*

Note. Le retournement temporel a d'abord été étudié expérimentalement par M. Fink et son groupe à l'ESPCI à Paris [2, 5]. Des études [3] sont maintenant menées pour comprendre quantitativement ce phénomène, et en particulier ce fait surprenant : plus le milieu est aléatoire, mieux l'onde refocalise !

Applications du retournement temporel

Imagerie par ultrasons. Certaines applications des ultrasons nécessitent la focalisation précise d'une onde en un point du volume à contrôler. En imagerie échographique ou en contrôle non-destructif, on mesure ainsi la réflectivité locale du milieu. De même en thérapie médicale l'énergie de l'onde ultrasonore est focalisée sur la zone à traiter, soit pour engendrer une élévation locale de la température pour l'hyperthermie, soit pour briser des calculs rénaux pour la lithotritie. Lorsque le milieu de propagation est inhomogène la focalisation de l'onde ultrasonore par les méthodes classiques est dégradée. Les MRT permettent de compenser les distorsions du signal induites par les hétérogénéités réparties sur le trajet de l'onde ultrasonore.

Télécommunication. En matière de télécommunications, la présence d'inhomogénéités dans le milieu ou de réverbérations sur des obstacles a longtemps été ressentie comme un facteur limitant. La présence de diffusion multiple des ondes se traduit pour les communications sans fil par une utilisation sous-optimale de la bande spectrale disponible et une capacité de communication réduite par rapport à la capacité théorique de Shannon. Les MRT devraient permettre d'exploiter au maximum les capacités de transmission d'un milieu.

Appendice : Homogénéisation et approximation-diffusion

On va illustrer sur un exemple particulièrement simple les deux régimes qui conduisent à des théorèmes limites remarquables. On regarde la position d'une particule sur $\mathbb{R}$ soumise à un champ de vitesse aléatoire $\varepsilon F(t)$ où ε est un petit paramètre, F est constant par morceaux

$$F(t) = \sum_{i=1}^{\infty} F_i \mathbf{1}_{[i-1,i[}(t),$$

et les F_i sont des variables aléatoires indépendantes et identiquement distribuées de moyenne $\mathbb{E}[F_i] = \bar{F}$ et de variance $\mathbb{E}[(F_i - \bar{F})^2] = \sigma^2$. La position de la particule partant de 0 à l'instant $t = 0$ est :

$$X(t) = \varepsilon \int_0^t F(s)ds.$$

Clairement $X(t) \overset{\varepsilon \to 0}{\longrightarrow} 0$. Le problème est de trouver la bonne asymptotique, c'est-à-dire l'échelle de temps à laquelle un mouvement macrosocopique de la particule est détectable.

Régime d'homogénéisation. Pour des temps grands, à l'échelle $t \to t/\varepsilon$, $X^\varepsilon(t) := X(t/\varepsilon)$ s'écrit :

$$X^\varepsilon(t) = \varepsilon \int_0^{\frac{t}{\varepsilon}} F(s)ds = \varepsilon \left(\sum_{i=1}^{[\frac{t}{\varepsilon}]} F_i \right) + \varepsilon \int_{[\frac{t}{\varepsilon}]}^{\frac{t}{\varepsilon}} F(s)ds$$

Lorsque $\varepsilon \to 0$, on trouve en appliquant la loi des grands nombres que

$$X^\varepsilon(t) = \underset{\downarrow}{\varepsilon\left[\frac{t}{\varepsilon}\right]} \times \underset{\text{p.s.}\downarrow}{\frac{1}{\left[\frac{t}{\varepsilon}\right]}\left(\sum_{i=1}^{\left[\frac{t}{\varepsilon}\right]} F_i\right)} + \underset{\text{p.s.}\downarrow}{\varepsilon\left(\frac{t}{\varepsilon} - \left[\frac{t}{\varepsilon}\right]\right) F_{\left[\frac{t}{\varepsilon}\right]}}$$
$$\begin{array}{ccc} \downarrow & \text{p.s.}\downarrow & \text{p.s.}\downarrow \\ t & \mathbb{E}[F] = \bar{F} & 0 \end{array}$$

La convergence a lieu presque sûrement (p.s.), c'est-à-dire avec probabilité 1. Ainsi le mouvement de la particule est ballistique dans le sens où sa vitesse est constante $X^\varepsilon(t) \overset{\varepsilon\to 0}{\longrightarrow} \bar{F}t$. Cependant, dans le cas $\bar{F} = 0$, le champ de vitesse aléatoire ne provoque aucun mouvement macroscopique de la particule, ce qui veut dire qu'il faut changer de régime et attendre plus longtemps.

Régime d'approximation-diffusion. Supposons $\bar{F} = 0$. A des temps très grands, à l'échelle $t \to t/\varepsilon^2$, $X^\varepsilon(t) = X(t/\varepsilon^2)$ s'écrit :

$$X^\varepsilon(t) = \varepsilon \int_0^{\frac{t}{\varepsilon^2}} F(s)ds = \varepsilon \left(\sum_{i=1}^{\left[\frac{t}{\varepsilon^2}\right]} F_i\right) + \varepsilon \int_{\left[\frac{t}{\varepsilon^2}\right]}^{\frac{t}{\varepsilon^2}} F(s)ds$$

Lorsque $\varepsilon \to 0$, on trouve en appliquant le théorème de la limite centrale que

$$X^\varepsilon(t) = \underset{\downarrow}{\varepsilon\sqrt{\left[\frac{t}{\varepsilon^2}\right]}} \times \underset{\text{loi}\downarrow}{\frac{1}{\sqrt{\left[\frac{t}{\varepsilon^2}\right]}}\left(\sum_{i=1}^{\left[\frac{1}{\varepsilon^2}\right]} F_i\right)} + \underset{\text{p.s.}\downarrow}{\varepsilon\left(\frac{t}{\varepsilon^2} - \left[\frac{t}{\varepsilon^2}\right]\right) F_{\left[\frac{t}{\varepsilon^2}\right]}}$$
$$\begin{array}{ccc} \downarrow & \text{loi}\downarrow & \text{p.s.}\downarrow \\ \sqrt{t} & \mathcal{N}(0,\sigma^2) & 0 \end{array}$$

La distribution statistique de $X^\varepsilon(t)$ converge quand $\varepsilon \to 0$ vers la distribution gaussienne $\mathcal{N}(0,\sigma^2 t)$ de moyenne nulle et de variance $\sigma^2 t$. Le mouvement de la particule dans ce régime est diffusif, sa distance typique par rapport à sa position d'origine augmente en $\sqrt{t}$.

Pour en savoir plus

[1] Asch (M.), Kohler (W.), Papanicolaou (G.), Postel (M.) et White (B.), Frequency content of randomly scattered signals, SIAM Rev. **33** (1991), 519-625.

[2] Fink (M.), Time reversed acoustics, Scientific American **281** : 5 (1999), 91-97.

[3] Fouque (J.-P.), Garnier (J.), Sølna (K.), and Papanicolaou (G.),Wave propagation and time reversal in randomly layered media, à paraître, Springer, (2006).

[4] Milton (G.), *The Theory of Composites*, Cambridge University Press, Cambridge, (2001).

[5] Tourin (A.), Fink (M.) et Derode (A.), Multiple scattering of sound, Waves Random Media **10** (2000), R31-R60.

Attracteurs des systèmes dynamiques et généricité

Yulij ILYASHENKO*

Les processus d'évolution sont omniprésents, que ce soit à travers le mouvement des atomes, ou bien dans la dynamique des planètes. Newton a pris conscience du fait que ces processus sont décrits par des équations différentielles. Dans les cent cinquante années qui ont suivi, on réalisa que la plupart des équations différentielles ne pouvaient être résolues explicitement. C'est alors Poincaré qui introduisit la théorie qualitative des équations différentielles. Celle-ci s'attache à décrire les propriétés géométriques des solutions, sans connaitre leur forme explicite. Là encore, on s'est rendu compte que le comportement qualitatif des solutions pouvait être extrêmement complexe. La situation se simplifie cependant si l'on ne considère que des équations différentielles génériques. Du point de vue physique, ce sont les plus intéressantes.

En première approche, l'étude des systèmes dynamiques peut être divisée en trois périodes :

- celle de Newton : une équation différentielle est donnée. Résolvez la !
- celle de Poincaré : une équation différentiele est donnée. Décrivez le comportement qualitatif des solutions, sans la résoudre !
- celle d'Andronov : aucune équation différentielle n'est donnée. Décrivez les propriétés qualitatives des solutions !

La dernière affirmation peut paraître paradoxale. Pourtant, elle fait référence à une branche des systèmes dynamiques très développée aujourd'hui, qui étudie non pas un système dynamique particulier, mais cherche plutôt à décrire le comportement d'un système *typique*. Par exemple, les équations différentielles planes ont génériquement d'importantes propriétés communes. Ces propriétés décrivent le comportement asymptotique de toutes les solutions. Nous les présentons plus bas. Ainsi, afin de s'assurer que les solutions d'une équation différentielle plane satisfont ces propriétés, il suffit de savoir que l'équation elle-même est *générique*. En dimension supérieure, on peut également décrire certaines propriétés des équations différentielles génériques. La situation est cependant plus complexe.

Dans ce texte, nous discutons cette approche.

Lois d'évolutions et équations différentielles

Considérons un système physique. Par exemple, un satellite dans le champ gravitationnel de la Terre, ou encore, le système solaire dans son ensemble. A chaque instant, l'état du système est décrit par un nombre fini de paramètres numériques $x = (x_1, \ldots, x_n)$. Dans le cas d'un satellite, le système est décrit par la position du satellite ainsi que sa vitesse, le nombre de paramètres est donc 6. On peut penser à l'ensemble de ces paramètres comme à un point x, évoluant dans un espace $\mathbb{R}^n$ de grande dimension, appelé *espace des phases*. La *loi d'évolution* indique comment le système évolue. Ainsi pour un satellite, la vitesse dans l'espace des phases est décrite par la vitesse à laquelle évoluent les six coordonnées qui décrivent la vitesse et la position du satellite. Autrement dit, elle est déterminée par la vitesse et l'accélération du satellite. Dans l'espace des phases, l'équation différentielle usuelle du second ordre

* *Professeur des Universités de Cornell, d'Etat de Moscou, et Indépendante de Moscou ; Président de l'Université Indépendante de Moscou ; Directeur de recherches à l'Institut Steklov. Institut Steklov, 8 rue Gubkina , 117966 Moscou, Russie.*

$F = ma$ est transformée en une équation différentielle du premier ordre. On note $V(x)$ la vitesse dans l'espace des phases, ainsi l'évolution du système dans l'espace des phases satisfait l'équation :

$$\dot{x} = V(x). \qquad (*)$$

Un théorème important d'existence et d'unicité des solutions pour une telle équation affirme la chose suivante. Etant donné un point x_0 dans l'espace des phases, il existe une unique solution $x(t)$ de l'équation $(*)$ telle que $x(0) = x_0$.

Déterminisme de Laplace et approche géométrique des équations différentielles

Newton fut le premier à comprendre que les processus d'évolution de l'Univers étaient régis par des équations différentielles. Laplace réalisa ensuite que le théorème d'existence et d'unicité des solutions évoqué précédemment, pouvait être appliqué à ces processus, et en tira des conséquences philosophiques. Il s'exprime ainsi dans son *Essai philosophique sur les probabilités* (Œuvres, Gauthier Villars, vol. II, 1, pp 6-7, 1886) :

> « Une intelligence qui, pour un instant donné, connaitrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, seraient présents à ces yeux. »
>
> Personne n'a jamais écrit l'équation différentielle dont Laplace rêvait !

Revenons à des choses plus élémentaires, et considérons des exemples très simples d'équations différentielles, du point de vue géométrique. Quand le temps t évolue, les solutions $x(t)$ du système décrivent des courbes dans l'espace des phases qui sont les *orbites* de l'équation différentielle $(*)$. Géométriquement, trouver la solution de l'équation passant par un point x_0 revient à trouver une courbe issue de x_0 qui soit toujours tangente à V. Considérons deux exemples élémentaires (l'espace des phases est ici le plan $\mathbb{R}^2$). Si $V(x) = x$ est le champ de vecteur radial, toutes les orbites sont des rayons issus de l'origine, à l'exception d'une d'entre elles. L'orbite issue de 0, est réduite à un point. Un tel point est appelé *point d'équilibre*. Supposons maintenant que $V(x)$ est égal à ix hors de l'origine (autrement dit, $V(x)$ est égal au vecteur x tourné d'un angle $\frac{\pi}{2}$), et $V(0) = 0$. Les orbites de ce champ de vecteurs sont des cercles centrés à l'origine, auxquels nous devons ajouter, là encore, l'état d'équilibre situé à l'origine.

Poincaré a été le premier à réaliser que, à défaut de pouvoir résoudre explicitement la plupart des équations differentielles, on peut décrire les propriétés géometriques de leurs orbites, uniquement à partir des propriétés du champ V.

Le théorème de Poincaré-Bendixson

Pour des équations différentielles planes, la description du comportement limite des solutions peut être faite en termes géométriques. Considérons la partition du plan en orbites (éventuellement réduites à un point) associées à une équation différentielle. Faisons l'hypothèse qu'il existe un grand disque dans le plan, dans lequel chaque orbite pénètre à un certain instant, pour ne plus en sortir (un tel système est dit *dissipatif*, dans le cas d'un système physique, cette hypothèse correspond à une perte d'énergie). Alors, chaque orbite a un comportement limite de l'une des formes suivantes.

- L'orbite est un point d'équilibre.
- L'orbite s'accumule autour d'une orbite périodique.
- L'orbite s'accumule autour d'un polygone dont les sommets sont des points d'équilibre et les côtés des orbites qui les relient (de telles orbites sont appelées *connections*).

Ces différentes possibilités sont illustrées dans la figure 1. Cette affirmation constitue le célèbre théorème de Poincaré-Bendixson, qui a été l'un des premiers succès de l'approche topologique des équations différentielles. Le

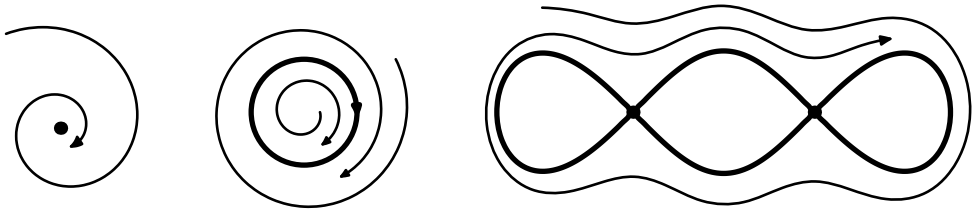


Figure 1 – *Ensembles ω-limite dans le plan.*

comportement que nous venons de décrire n'est pas très compliqué. Cependant nous aimerions le rendre encore plus simple. Pour cela, on peut argumenter de la manière suivante. La troisième situation ci-dessus, qui est la plus complexe, n'est pas générique. On peut s'attendre à ce qu'elle n'apparaisse pas dans des systèmes d'origine physique. Supposons par exemple qu'un système physique ne possède pas de symétrie non-triviale, ni de loi de conservation. Alors, on peut s'attendre à ce que différents opérateurs linéaires associés à l'équation différentielle $(*)$ (tels que le champ de vecteurs linéarisé au voisinage d'un point d'équilibre, ou encore la différentielle de l'application de premier retour de Poincaré) ne possèdent pas de valeurs propres nulles, ou bien de module 1.

Les points singuliers d'une équation différentielle plane générique sont de trois types, illustrés dans la figure 2 : les points selles (a), les noeuds (b), et les foyers (c).

Une propriété importante des champs de vecteurs génériques dans le plan est qu'ils n'ont pas de connexion entre deux points selle. Toute connexion de ce type peut en effet être détruite par une perturbation arbitrairement petite. Le concept naïf de « propriété générique » a été fondé par René Thom, à travers ses théroèmes de transversalité. Thom fut l'un des fondateurs de la théorie des catastrophes, dont l'objet est d'étudier et de classifier les singularités des systèmes génériques ainsi que leurs bifurcations. On peut en fait formuler différents concepts de *généricité*. Par exemple, on peut parler de généricité du point de vue topologique, ou bien du point de vue métrique (c'est-à-dire du point de vue de la théorie de la mesure). Une propriété d'un système dynamique est *topologiquement générique* si elle est vraie pour tous les systèmes qui se trouvent dans une intersection dénombrable d'ouverts denses de l'espace de tous les systèmes dynamiques. Une propriété est *métriquement générique*, si, pour toute famille (f_α) de systèmes dynamiques, où $\alpha \in \mathbb{R}^p$ est un paramètre, la propriété considérée est vraie pour tous les paramètres α en dehors d'un ensemble de mesure nulle. Cette seconde définition peut, au prix de quelques efforts, être généralisée à l'espace (de dimension infinie) de tous les sytèmes dynamiques.

Nous pouvons maintenant formuler le théorème d'Andronov :

Théorème 1. *Un système physique ne possédant ni de symétries, ni de loi de conservation, modèlisé par une équation différentielle plane, possède un nombre fini de régimes limites. Chaque orbite converge vers un régime limite, et chacun de ces régimes limite est ou bien un point critique, ou bien une orbite périodique.*

Au milieu du XX^{e} siècle, quelques experts ont rêvé de généraliser ce résultat en dimensions supérieures. A la fin des années cinquante, Smale a publié une conjecture allant dans ce sens, en décrivant précisement ce que devrait être un système dynamique générique sur une variété compacte. Les systèmes qu'il a alors décrit forment une classe importante de systèmes appelés maintenant *systèmes de Morse-Smale*. Une de leurs caractéristiques est qu'ils ne possèdent qu'un nombre fini d'orbites périodiques. Cependant, contrairement a ce qu'affirmait la conjecture de Smale, ils ne sont pas génériques. Peu après la publication de l'article de Smale, quelques experts de la génération qui le précédait, lui indiquèrent

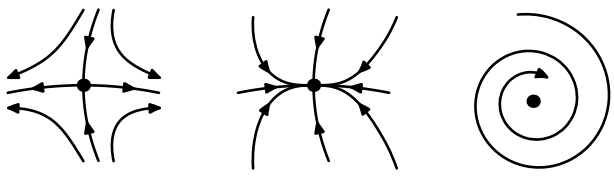


Figure 2 – *Points singuliers génériques dans le plan.*

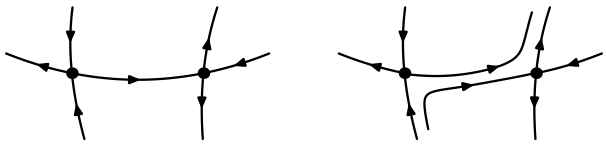


Figure 3 – *Connection entre deux points selle et sa destruction.*

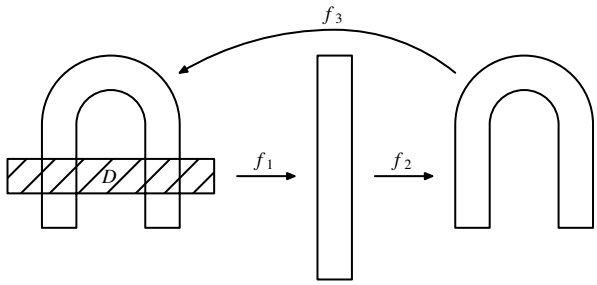


Figure 4 – *Le fer à cheval de Smale.*

que, dans des travaux de Cartwrite, Littlewood et Levinson, étaient construits des systèmes dynamiques ayant une infinité de points périodiques, et cette propriété persistait après une petite perturbation.

Depuis cette époque, l'une des questions majeures de la théorie des systèmes dynamiques a été de savoir quelles sont les propriétés des systèmes dynamiques génériques. Un certain nombres de résultats allant dans cette direction ont été obtenu, mais, de nombreux problèmes restent ouverts en toute généralité. Nous présentons un cas particulier de l'un de ces problèmes à la fin de l'article.

Plutôt que de lire les travaux de ses prédécesseurs, qui étaient longs et fastidieux, Smale a rapidement construit un contre-exemple à sa propre conjecture, pour comprendre comment un système dynamique pouvait posséder, de manière persistante, une quantité infinie dénombrable d'orbites périodiques. L'histoire dit que c'est en se promenant le long de la plage de Copacabana, à Rio de Janeiro, qu'il a inventé l'exemple suivant, appelé maintenant le *fer à cheval de Smale*.

Il était bien compris depuis l'époque de Poincaré que l'étude des équations différentielles, et l'étude des itérations d'un difféomorphisme étaient deux branches de la même théorie. Ainsi, l'exemple de Smale est obtenu en considérant un difféomorphisme, en dimension 2. L'application f est la composition de trois applications que nous décrivons maintenant, et qui sont illustrées dans la figure 4. L'application f_1 contracte le rectangle D dans la direction horizontale et l'étire dans la direction verticale. Puis l'application f_2 plie le rectangle ainsi obtenu pour former un fer à cheval. Enfin, l'application f_3 déplace le fer à cheval afin qu'il intersecte le rectangle initial D comme indiqué sur la figure initiale.

Encadré 1

Considérons une version simplifée de l'application précédente, que nous appelerons encore fer à cheval de Smale. Cette nouvelle application est illustrée dans la figure 5.
Considérons un carré, partitionné en 5 rectangles horizontaux de largeurs égales. Notons D_0 et D_1 respectivement le second et le quatrième rectangle. Nous pouvons également partitionner le carré en cinq rectangles verticaux. On note D_0' et D_1' respectivement, le second et le quatrième rectangle vertical. Enfin, on note $D = D_0 \cup D_1$ et $D' = D_0' \cup D_1'$. Considérons alors l'application $f : D \to D'$, qui contracte D_j cinq fois dans la direction horizontale, le dilate cinq fois dans la direction verticale, puis le translate sur le rectangle D_j' $(j = 0,1)$. L'application f est affine par morceaux, on peut écrire une formule explicite qui la décrit en restriction à chacun des rectangles D_0 et D_1. De plus, l'application f peut être prolongée en un difféomorphisme de la sphère.

Si, pour un point x de D, toutes les itérations (positives et négatives) de f sont définies, nous dirons que x possède une orbite complète *pour f. On note Λ l'ensemble des points de D qui possèdent une orbite complète pour la transformation f. Il n'est pas difficile de voir que Λ est le produit cartésien de deux ensembles de Cantor. Si x est un point de Λ, chacun des points $f^n(x)$ (où n décrit $\mathbb{Z}$) est dans l'un des deux rectangles D_0 ou D_1. Si $f^n(x)$ est dans D_j (où j est égal à 0 ou 1), on pose $\omega_n(x) = j$. On associe ainsi une suite $\omega(x) = (\omega_n(x))_{n\in\mathbb{Z}}$ formée de 0 et de 1 à chaque point de Λ. Il s'avère que toute suite formée de 0 et de 1 est la suite associée à un et un seul point de Λ. Puisque les suites périodiques sont en nombre infini, l'application f possède un nombre infini de points périodiques.*

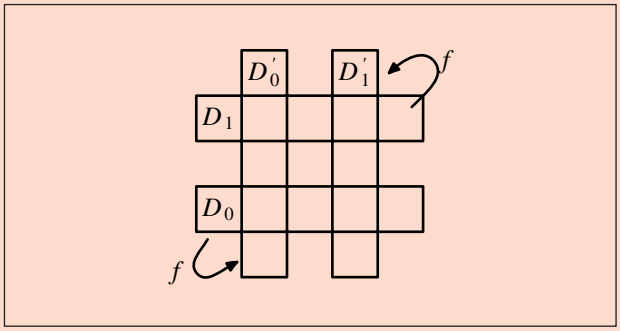


Figure 5 – *Fer à cheval de Smale simplifié.*

Une analyse élémentaire de l'application décrite dans l'encadré précédent, montre le fait suivant : si deux points x et y de Λ se trouvent à une distance l'un de l'autre inférieure à 5^{-n}, leurs n premières images par f (ou f^{-1}) se trouvent dans le même rectangle D_j. Autrement dit, dans les suites $\omega(x)$ et $\omega(y)$ correspondantes, les termes d'indice k compris entre $-n$ et n coïncident. Par contre, après ces n premières itérations, leurs images peuvent varier de manière aléatoire, et leur distance est alors proche de 1. On résume cette propriété en disant que le système est très sensible aux conditions initiales.

Différents types d'attracteurs : coïncident-ils génériquement ?

Remarquons que, dans l'exemple précédent, « presque tout » point du carré ne possède pas d'orbite positive complète pour la transformation considérée. On peut citer un autre exemple, dû à Smale et Williams, qui n'a pas ce défaut. Il est illustré dans la figure 6. C'est une application du tore solide $T = D^2 \times S^1$ dans lui-même, ayant les propriétés suivantes.

- Tous les points du tore solide ont une orbite positive complète.
- Il existe un fermé invariant Λ appelé solénoide, à l'intérieur du tore solide, vers lequel toutes les orbites sont attirées.
- La dynamique à l'intérieur du solénoïde est semblable à celle du fer à cheval, en particulier elle est très sensible aux conditions initiales et possède une infinité d'orbites périodiques.

De plus ces propriétés persistent après de petites perturbations. Le solénoïde de Smale-Williams fournit un exemple d'*attracteur étrange*, dont le comportement est très différent de celui d'un point fixe attractif ou d'une orbite périodique attractive. Tout système dynamique dissipatif possède un ensemble attractif appelé *attracteur*

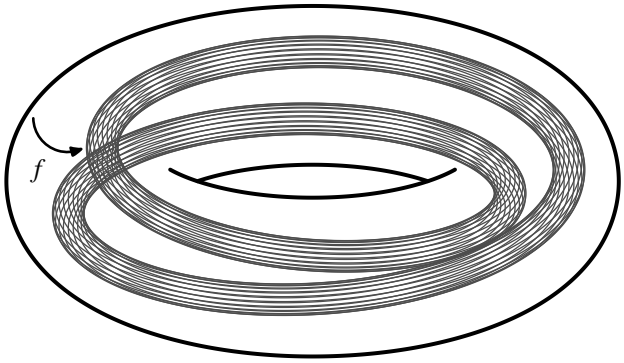


Figure 6 – *Le solénoïde de Smale-Williams.*

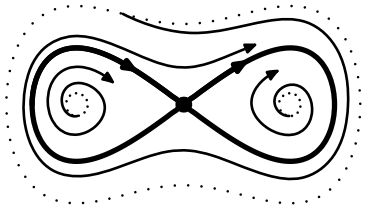


Figure 7 – *Non-coïncidence des attracteurs maximal et statistique.*

maximal vers lequel toutes les orbites convergent. Considérons un difféomorphisme f qui envoie un compact B dans lui-même, mais tel que $f(B) \neq B$. Dans ce cas, l'attracteur maximal A_{max} pour le système dynamique $f : B \to B$ est :

$$A_{max} = \cap_{n=0}^{\infty} f^n(B).$$

Mais en pratique, cet ensemble est trop gros. Lors d'une simulation numérique, un ensemble plus petit que l'attracteur maximal va jouer un rôle : c'est l'ensemble fermé minimal près duquel la majorité des orbites passent la majorité du temps. C'est l'*attracteur statistique*.

Un exemple est representé sur la figure 7. Dans cet exemple l'attracteur maximal est le « huit » et l'attracteur statistique est formé d'un point selle. Mais cet exemple n'est pas générique (il possède deux connections entre points selles, que l'on peut faire disparaître par une petite perturbation). Ainsi, on peut espérer que, génériquement, les attracteurs maximal et statistique coïncident. Cette conjecture reste ouverte aujourd'hui. Une manière plus précise de la formuler est la suivante.

Est-il vrai que, pour un système dynamique générique f, l'attracteur statistique A est égal à l'attracteur maximal de la restriction de f à un voisinage de A ?

Pour en savoir plus

• Sur les sytèmes dynamiques en dimension 2 :

Andronov (A.), Vitt (A.), Khaikin (S.), *Theory of oscillations*, Moscow (1959).

• Sur les attracteurs étranges et la dépendance aux conditions initiales :

Ruelle (D.), Takens (F.), *On the nature of turbulence*, Commun. Math. Phys., vol.20, pages 167-192 (1971).

• Sur les systèmes dynamiques génériques et la théorie des catastrophes :

Katok (A.), Hasselblatt (B.), *Introduction to the modern theory of dynamical systems*, Cambridge University Press (1994).

Arnold (V.), *Catastrophe Theory*, Springer Encyclopaedia in Math., vol.5 (1994).

Palis (J.), *A global perspective for non-conservative dynamics*, Ann. I.H.Poincare – AN, vol.22, pages 485-507 (2005).

• Sur les différentes notions d'attracteurs :

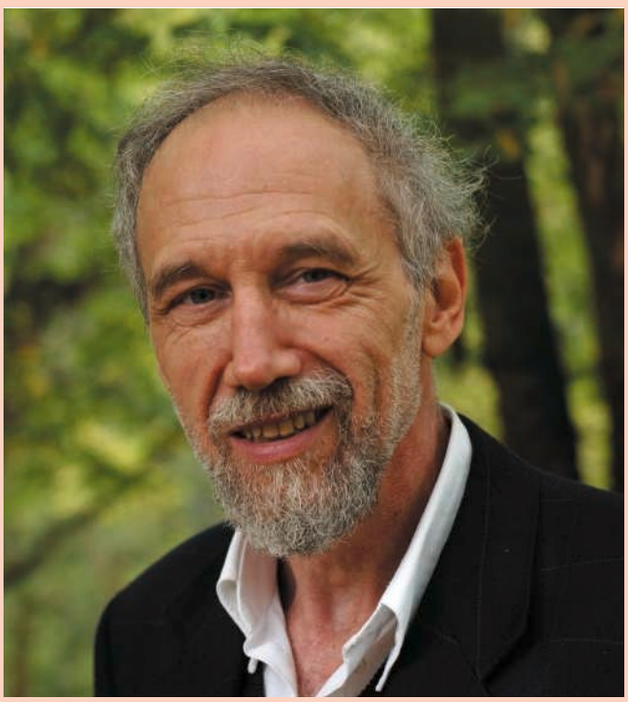
Milnor (J.), *On the concept of attractor*, Commun. Math. Phys., vol.99, no.2, pages 177-196 (1985).

Gorodetski (A.), Ilyashenko (Yu.), *Minimal and strange attractors*, International Journal of Bifurcation and Chaos, vol.6, no.6, pages 1177-1183 (1996).

Remerciements

Je voudrais remercier Pierre Py pour une traduction excellente de la version anglaise de ce texte, Victor Kleptsyn pour la réalisation des figures et Étienne Ghys qui m'a suggéré d'écrire cet article.

Alain Connes : une autre vision de l'espace



Il est difficile de présenter Alain Connes sans commencer par évoquer brièvement son impressionnant curriculum vitae. Né en 1947, élève de l'Ecole Normale Supérieure (1966-1970), après un passage au CNRS, il est successivement professeur à Paris 6, directeur de recherche au CNRS, et depuis 1984 professeur au Collège de France, où il occupe la chaire d'Analyse et Géométrie. Parallèlement, il est depuis 1979 professeur à l'Institut des Hautes Etudes Scientifiques à Bures-sur-Yvette. Il partage sa vie mathématique entre ces deux lieux : l'IHES lui offre le calme pour s'adonner à ses recherches et la possibilité de rencontrer des mathématiciens et des physiciens théoriciens de tous pays et de toutes spécialités ; le Collège de France lui donne l'occasion, chaque année, de présenter dans un cours ses résultats les plus récents. Depuis 2003 il est aussi professeur à l'Université de Vanderbilt aux Etats-Unis. Il a reçu les plus hautes distinctions internationales : la médaille Fields (1982) et le prix Crafoord (2001), deux prix dont le prestige est comparable au prix Nobel, qui, on le sait, n'existe pas pour les mathématiques. En France, où il est membre de l'Académie des Sciences depuis 1983, il vient de recevoir, en décembre 2004, la médaille d'or du CNRS.

La thèse : classification des facteurs de type III

Pour comprendre qui est Alain Connes, il faut discerner quelles furent les différentes étapes de sa vie mathématique. Dans un premier temps, il s'est imposé comme un jeune mathématicien au talent exceptionnel, en résolvant un problème reconnu difficile par les spécialistes, mais considéré par beaucoup à l'époque comme marginal par rapport aux « grandes mathématiques ».

Il suivait alors le séminaire d'algèbres d'opérateurs de Jacques Dixmier à l'Ecole Normale Supérieure, où l'on parlait notamment d'algèbres de von Neumann. Ces algèbres sont des généralisations non commutatives, ou si l'on préfère, quantiques, de la théorie de la mesure. Elles avaient été introduites par von Neumann dès les années 20 ou 30 pour donner un fondement mathématique à la mécanique quantique qui venait d'être découverte.

Etant donné un espace de Hilbert H on considère l'algèbre $\mathcal{L}(\mathcal{H})$ des opérateurs bornés sur H. On considère des algèbres d'opérateurs, c'est-à-dire des sous-algèbres A de $\mathcal{L}(H)$ telles que si un opérateur T appartient à A, il en est de même de son adjoint T^*. On impose de plus à la sous-algèbre A une condition topologique : on dit que A est une algèbre de von Neumann si elle est stable par convergence faible ou forte (c'est la convergence simple sur tout vecteur de H) ; on dit que A est une C^*-algèbre si elle est stable pour la convergence normique (c'est la convergence uniforme sur la boule unité de H).

Parmi les algèbres de Von Neumann, un rôle fondamental est joué par les **facteurs**, c'est-à-dire les algèbres de von Neumann dont le centre est réduit aux scalaires. Murray et von Neumann, dans les années 1930, avaient tenté une classification des facteurs, qu'ils répartirent en trois classes.

Les facteurs de classe I et II sont les plus proches du cas commutatif, avec une notion de trace et de dimension : une trace est une forme linéaire positive τ qui vérifie la propriété $\tau(xy) = \tau(yx)$. Le type I est celui des algèbres de matrices $M_n(\mathbf{C})$, ou en dimension infinie, de l'algèbre $\mathcal{L}(H)$ des opérateurs bornés sur un espace de Hilbert. On a alors la trace usuelle des

opérateurs, qui a la propriété d'intégralité : elle prend des valeurs entières sur les projecteurs (éléments p tels que $p^2 = p = p^*$). Mais il y a aussi des facteurs admettant une trace qui prend des valeurs réelles quelconques sur les projecteurs : c'est le type II.

Murray et von Neumann découvrirent aussi l'existence d'objets hautement non commutatifs, dit facteurs de type III, qui n'admettent aucune trace. Jusqu'aux années 1970, ces facteurs étaient restés mystérieux et résistaient à toute tentative de classification. Lorsque Alain Connes arrive au séminaire Dixmier, on en est encore là, mais les travaux de Powers, Araki et Woods ont produit de nouveaux exemples de facteurs de type III, et même une infinité de tels facteurs deux à deux non isomorphes. Le génie d'Alain Connes a été d'appliquer à ces objets une théorie encore nouvelle et peu exploitée, due au mathématicien japonais Minoru Tomita. Un facteur de type III n'ayant pas de trace, on remplace la notion de trace par celle de poids. Un poids est une forme linéaire positive φ, mais *a priori* $\varphi(yx) \neq \varphi(xy)$. La surprise, c'est que **la non commutativité engendre une dynamique**, une évolution au cours du temps donnée par un groupe d'automorphismes à un paramètre du facteur.

Plus précisément, on a pour tous les $t \in \mathbf{R}$ des automorphismes σ_t du facteur M, avec la loi de groupe $\sigma_t \circ \sigma_s = \sigma_{t+s}$, qui permettent de corriger la non commutativité via la formule dite KMS_β :

$$\varphi(yx) = \varphi(x\sigma_{i\beta}(y))$$

où $\sigma_{i\beta}$ s'obtient par prolongement analytique, pour un certain $\beta > 0$.

Connes a montré que ce groupe à un paramètre est en fait indépendant du poids, modulo les automorphismes intérieurs, et donne lieu à des invariants spectraux qui permettent de classifier les facteurs de type III. Connes a ainsi introduit dans sa thèse (1973) les facteurs dits III_λ où λ est un nombre réel entre 0 et 1. Cette solution, par un jeune thésard encore inconnu, d'un problème ouvert depuis des décennies a profondément impressionné Jacques Dixmier et l'ensemble des spécialistes des algèbres d'opérateurs. Cependant, la majorité des mathématiciens français ignoraient alors la théorie de von Neumann, et le problème de la classification des facteurs n'était pas considéré comme un des principaux défis des mathématiques. Ce sont surtout les physiciens théoriciens qui ont reconnu le génie d'Alain Connes. La mécanique statistique quantique, ainsi que certains modèles de théorie des champs, utilisent en effet les algèbres de von Neumann de façon essentielle.

La statistique de Boltzmann associée à un hamiltonien H est donné (sur une observable A) par le poids

$$\varphi(A) = Trace(Ae^{-\beta H})/Trace(e^{-\beta H})$$

et l'évolution dans le temps par les automorphismes

$$\sigma_t(A) = e^{itH}Ae^{-itH}.$$

Le lecteur vérifiera sans peine la formule KMS_β ci-dessus. Le groupe à un paramètre associé à un poids modélise l'évolution dans le temps d'un système statistique quantique associé à un état de température donné (comme d'habitude, $\beta = 1/kT$ où k est la constante de Boltzmann et T la température absolue).

Aussi, lorsque Connes devient visiteur à l'IHES, c'est en tant que physicien théoricien.

Des facteurs aux feuilletages : vers la géométrie non commutative

Cette arrivée de Connes à l'IHES est un tournant de sa carrière scientifique. En effet, il entre en contact avec des mathématiciens qui, s'ils ignorent tout de la théorie des facteurs de type III, jonglent quotidiennement avec des faisceaux et des feuilletages, avec des groupes d'homologie et des tenseurs de courbure. La géométrie différentielle et la géométrie algébriques sont considérées par la communauté mathématique comme plus centrales, pour ne pas dire plus nobles, que la théorie des algèbres d'opérateurs. Mais Connes ne se laisse pas impressionner. Il raconte, avec sa modestie habituelle, et quelqu'exagération, qu'il ne comprenait rien aux conversations de ses collègues avec qui il déjeûnait à la cafétéria de l'Institut. Mais qu'on ne s'y trompe pas, Alain Connes a une faculté étonnante d'assi-

miler de nouvelles notions. Et pour apprendre une théorie mathématique, au lieu de se plonger dans les livres, il préfère discuter avec d'autres mathématiciens, se faire expliquer puis retrouver par lui-même toute la théorie. Et il fait cela très vite. Aussi a-t-il très rapidement compris la théorie des feuilletages, et vu le lien avec la théorie de von Neumann. Un feuilletage est un objet géométrique qui est localement trivial comme un mille-feuille, mais qui globalement a une structure dynamique non triviale. Connes définit alors l'algèbre de von Neumann associée à un feuilletage, qui est en général un facteur, souvent de type III, le groupe à un paramètre de la théorie de Tomita ayant une interprétation géométrique très simple. Il définit aussi la C^*-algèbre associée à un feuilletage, et dans l'esprit d'Atiyah et Singer montre un théorème d'indice pour les opérateurs elliptiques le long des feuilles d'un feuilletage transversalement mesuré. D'où l'idée de **géométrie non commutative**. L'espace des feuilles d'un feuilletage est un ensemble non dénombrable sur lequel on serait bien en peine de définir, au sens classique, une théorie de la mesure, ou une topologie, voire une structure différentiable.

Expliquons cela par un exemple, celui du feuilletage du tore bidimensionel par une droite de pente irrationnelle. On se donne un carré dans le plan ; après recollage des côtés opposés, on obtient un tore, qui est une variété compacte de dimension 2. On se donne une direction D dans le plan. Partant d'un point x_1 du bord inférieur I, on construit une suite de segments de droites dans le carré de la façon suivante : le premier segment est issu de x_1 ; chaque segment est parallèle à D et a son origine et son extrêmité sur le bord du carré ; chaque fois qu'un point du bord est extrêmité d'un segment, l'origine du segment suivant est le point opposé du bord. On obtient ainsi une famille de segments tracés à l'intérieur du carré, et après recollage des bords, on obtient une trajectoire sur le tore, appellé **feuille** (du point de vue de la géométrie riemannienne, c'est une géodésique de la métrique plate sur le tore). Le **feuilletage** est donné par la partition ainsi obtenue de la variété (ici le tore) en feuilles. Voir les figures ci-contre où l'on a noté $x_2, x_3, \cdots$ les points obtenus sur le segment I.

Il faut distinguer deux cas. Supposons que la pente de D soit rationnelle. Le processus est périodique (sur la figure, la pente est égale à 2). Il n'y a qu'un nombre fini de segments, la trajectoire sur le tore est périodique. L'espace des trajectoires est bien décrit par l'algèbre commutative des fonctions continues sur un intervalle (ici la moitié de l'intervalle I). C'est un espace usuel, dit commutatif.

Prenons au contraire pour la pente de D un nombre θ irrationnel. On a tracé les quatre premiers points situés sur I de la trajectoire de x_1 et les deux premiers points de celle de y_1. Chaque trajectoire est infinie (on ne revient jamais au point de départ) et dense. Si on veut décrire l'espace des trajectoires par des fonctions continues sur I, celles-ci doivent avoir la même valeur sur les points d'une même trajectoire, donc être constantes. L'algèbre de ces fonctions est l'algèbre des nombres complexes. Elle ne donne aucun renseignement sur l'espace $\mathcal{T}_\theta$ des feuilles. Alain Connes propose de décrire $\mathcal{T}_\theta$ par un algèbre non commutative. En ce sens, **l'espace des feuilles est un espace non commutatif**.

Un autre exemple d'espace non commutatif est l'espace des orbites de l'action d'un groupe discret sur une variété compacte. Le cas le plus simple, intimement lié au feuilletage ci-dessus, est celui où la variété est le cercle muni de l'action du groupe $\mathbf{Z}$ engendrée par une rotation d'angle $2\pi\theta$. Là encore les orbites sont denses si θ est irrationnel et l'espace des orbites doit être décrit par une algèbre non commutative.

Pour définir l'algèbre du feuilletage du tore ci-dessus, on considère des noyaux $k(x,y)$ définis sur les couples de points x et y situés sur une même feuille, continus et à support compact sur chaque feuille. On multiplie ces noyaux par le produit usuel de convolution des noyaux :

$$k_1 * k_2(x,y) = \int k_1(x,z)k_2(z,y)dz$$

où l'intégrale est prise sur la feuille, pour la mesure de Lebesgue. De tels noyaux peuvent être interprêtés comme des familles d'opérateurs indexés par les feuilles. En complétant l'algèbre ainsi obtenue, on définit alors d'une part l'algèbre de von Neumann du feuilletage, et d'autre part la C^*-algèbre du même feuilletage. De même, dans le cas de l'action de $\mathbf{Z}$ on considère des matrices à support fini $a_{i,j}$ indexées par des couples (i,j) d'éléments d'une même orbite, et qu'on multiplie par le produit usuel des matrices, et on définit ainsi l'algèbre de von Neumann et la C^*-algèbres associées.

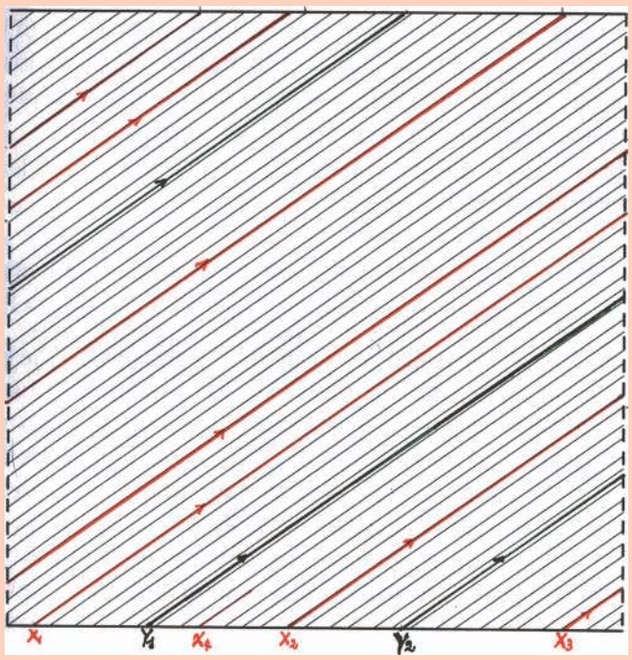


Figure 1 – *Pente irrationnelle.*

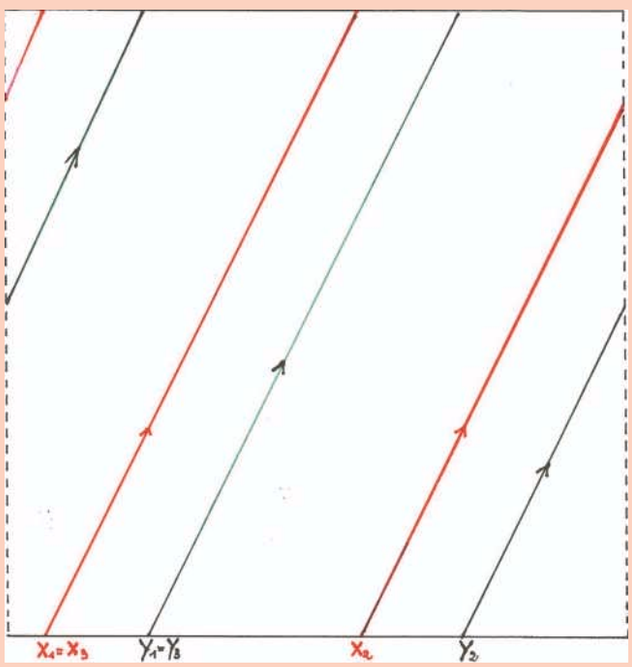


Figure 2 – *Pente rationnelle.*

Notons que dans les deux exemples ci-dessus les facteurs obtenus sont de type II. Cela est dû au fait que l'action de la rotation sur le cercle préserve la mesure de Lebesgue (ou pour le feuilletage, le flot des trajectoires préserve une mesure transverse), d'où une trace sur l'algèbre. Mais on construit facilement des exemples sans mesure invariante, et on obtient alors le type III.

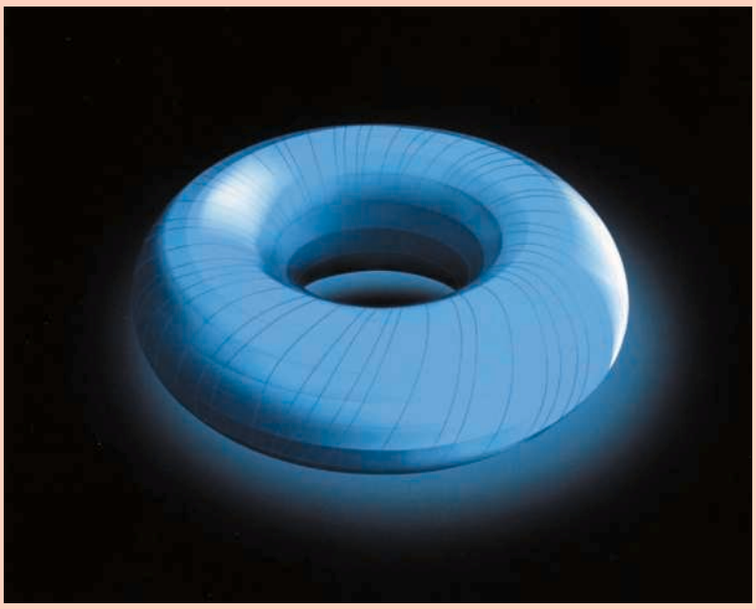


Figure 3 – *Feuilletage du tore.*

Ces géométries « non commutatives » sont donc décrites au moyen d'algèbres non commutatives qui jouent le rôle d'espaces de fonctions. Ainsi un espace mesuré X est décrit par l'algèbre $L^\infty(X)$ des fonctions mesurables bornées, qui est une algèbre de von Neumann commutative. Pour l'espace des feuilles d'un feuilletage, c'est l'algèbre de von Neumann (non commutative) associée qui joue le rôle d'algèbre de fonctions mesurables bornées sur l'espace des feuilles. De même pour la topologie : un espace topologique compact X est décrit par l'algèbre des fonctions continues sur X. Pour un feuilletage, la C^*-algèbre (non commutative) associée est considérée comme l'algèbre des fonctions continues sur l'espace « non commutatif » des feuilles. Ce qui est fait pour un feuilletage peut aussi se faire pour l'action d'un groupe discret sur un espace compact. L'espace des orbites est lui aussi un espace « non commutatif » décrit par une C^*-algèbre (du point de vue topologique) ou une algèbre de von Neumann. D'autres espaces non commutatifs sont des espaces d'orbites de relations d'équivalences, ou bien sont définis par des groupoïdes. Un autre cas intéressant est celui des groupes : le dual d'un groupe, c'est-à-dire l'espace des classes d'équivalences de représentations unitaires irréductibles, est aussi un espace non commutatif : dans le cas des groupes de Lie ou des groupes p-adiques, cet espace désingularise le dual décrit par la théorie des représentations, mais dans le cas des groupes discrets, on obtient un espace non commutatif hautement non trivial et encore mal connu.

K-théorie et homologie cyclique

Un invariant de topologie algébrique qui se généralise sans difficulté au cas non commutatif est la **K-théorie**. C'est un groupe abélien $K(X)$ attaché à un espace topologique X (disons, compact) et qui classifie (à isomorphisme stable près) les fibrés vectoriels sur X.

Deux fibrés E_1 et E_2 sur X sont stablement isomorphes s'il existe un fibré F tel que les fibrés sommes directe $E_1 \oplus F$ et $E_2 \oplus F$ soient isomorphes. Sur l'ensemble des classes d'isomorphisme stable de fibrés, on définit l'addition par la somme directe des fibrés, est $K(X)$ n'est autre que le groupe (dit de Grothendieck) des différences formelles de classes de fibrés, exactement comme dans la construction de l'anneau $\mathbf{Z}$ *des entiers relatifs à partir des entiers naturels.*

On sait depuis Atiyah et Singer que cet invariant joue un rôle crucial dans la théorie de l'indice. Par un théorème dû à Serre, les fibrés sur X correspondent à certains modules sur l'algèbre $C(X)$. On peut alors définir la K-théorie d'une C^*-algèbre et donc d'un espace topologique non commutatif. Dans le début des années 1980, Connes s'est intéressé à cet invariant, et a compris, avec Georges Skandalis, l'intérêt de la K-théorie bivariante développée à ce

moment là par le mathématicien russe Gennadi Kasparov. Ils ont montré, dans l'esprit de Grothendieck et en réinterprétant une des preuves d'Atiyah et Singer, comment le théorème de l'indice pour les opérateurs elliptiques sur des variétés compactes se ramenait à une propriété de fonctorialité en théorie de Kasparov. De là la généralisation aux feuilletages était naturelle. Il devenait donc naturel de calculer la K-théorie d'espaces non commutatifs comme les feuilletages, les groupes ou les actions de groupes. Dès 1980 lors d'un congrès à Kingston (Ontario), Alain Connes a eu l'intuition d'une interprétation géométrique de la K-théorie de la C^*-algèbre d'un feuilletage ou d'un groupe. De ses discussions avec le topologue Paul Baum, qu'il rencontre alors, jaillira l'idée d'une conjecture, désormais célèbre sous le nom de **conjecture de Baum-Connes** : le groupe de K-théorie analytique (c'est-à-dire la K-théorie de la C^*-algèbre du groupe ou du feuilletage) est isomorphe, *via* une flèche d'indice, à un groupe de K-théorie dit géométrique, construit à partir du classifiant des actions propres du groupe (ou du groupoïde d'holonomie du feuilletage). Cette conjecture est liée à la théorie des représentations des groupes de Lie (séries discrètes), à la topologie (conjecture de Novikov sur l'invariance par homotopie des hautes signatures), à la géométrie riemannienne (conjecture de Gromov-Lawson sur les obstructions à l'existence de métriques riemanniennes à courbure scalaire positive), à l'algèbre (conjecture des idempotents).

Précisée par la suite avec N. Higson, la conjecture de Baum-Connes sera le point de départ des travaux de nombreux mathématiciens pendant au moins 20 ans. Parmi les résultats les plus spectaculaires, citons :

– *le travail de N. Higson et G. Kasparov en 1996 qui établit la conjecture pour les groupes ayant la propriété dite de Haagerup ou a-T-menabilité (Gromov) ; c'est une classe de groupes relativement vaste, elle contient les groupes moyennables, mais exclut par exemple les groupes discrets possédant une propriété de rigidité dite propriété T de Kazhdan.*

– *la thèse de Vincent Lafforgue en 1998, qui pour la première fois prouve la conjecture de Baum-Connes pour certains groupes discrets ayant la propriété T. Elle a également permis de démontrer la conjecture pour tous les groupes localement compacts connexes.*

Malgré le nombre de résultats établissant la conjecture pour des groupes particuliers, un cas aussi simple que celui de $SL(3,\mathbf{Z})$, le groupe des matrices 3 sur 3 à coefficients entiers et de déterminant égal à 1 (ce groupe a la propriété T), est ouvert, sans qu'on ait aucune idée d'approche. Inversement, on peut tenter de construire des groupes suffisemment sauvages pour être des contrexemples à la conjecture. L'idée puissante de M. Gromov de construire des groupes aléatoires en sorte qu'ils aient des propriétés très éloignées des groupes pour lesquels la conjecture est actuellement prouvée, est prometteuse. Mais pour l'instant , elle n'a réussi à fournir que des contrexemples à une conjecture, dite conjecture de Baum-Connes à coefficients, qui est plus forte que la conjecture de Baum-Connes classique.

Mais la K-théorie n'est pas le seul invariant intéressant en topologie algébrique. Une question naturelle est de définir l'homologie, ou la cohomologie d'un espace topologique non commutatif. Pour cela, il faut faire un détour par la géométrie différentielle. On sait que dans le cas classique, l'homologie (ou la cohomologie) peut être définie comme (co)homologie singulière (au moyen de triangulations par des simplexes ou des cycles singuliers) ou bien de façon équivalente comme cohomologie de Čech, décrite au moyen de cocycles associés à un recouvrement par une famille d'ouverts. Il s'agit là de définitions topologiques de la (co)homologie à coefficients entiers, naturellement invariantes par homéomorphisme, mais difficilement généralisables dans le cas non commutatif. Si l'espace topologique est de plus muni d'une structure de variété différentiable, alors il y a une autre définition de la cohomologie (à coefficients réels ou complexes), celle de de Rham, obtenue à partir du complexe des formes différentielles. C'est cette dernière définition qui est retenue par Connes pour le cas non commutatif. Mais le prix à payer, c'est qu'il faut choisir une certaine sous-algèbre dense de la C^*-algèbre, jouant le rôle de l'algèbre des fonctions lisses (de classe C^∞, par exemple). Le point de départ est dans des calculs de géométrie non commutative : le calcul de caractères de Chern de modules de Fredholm fait apparaître une généralisation de la notion de trace. Ainsi un 2-cocycle cyclique est une forme trilinéaire sur une algèbre vérifiant les formules :

$$\tau(a_0, a_1, a_2) = \tau(a_1, a_2, a_0)$$

$$\tau(a_0a_1,a_2,a_3) - \tau(a_0,a_1a_2,a_3) + \tau(a_0,a_1,a_2a_3) - \tau(a_3a_0,a_1,a_2) = 0$$

A partir de la notion de n-cocycles, Alain Connes définit en 1981 la **cohomologie cyclique** d'une algèbre. Il développe cette théorie purement algébrique, découvre la longue suite exacte qui permet de la calculer. L'homologie

cyclique a depuis été abondemment utilisée par les algébristes, indépendemment des motivations d'Alain Connes qui, lui, revient toujours à son idée : la géométrie non commutative. L'un des problèmes les plus difficiles est de bien choisir la sous-algèbre dense de la C^*-algèbre : il faut qu'elle soit suffisamment petite pour qu'on puisse calculer sa cohomologie cyclique, mais suffisamment grosse pour avoir la même K-théorie que la C^*-algèbre. Une bonne partie des articles d'Alain Connes dans les années 1980 tournent autour de cette problématique, appliquée au cas des feuilletages ou au cas essentiellement équivalent des actions de groupes discrets sur des variétés. Il définit dans ce cadre la classe fondamentale transverse, analogue de la classe fondamentale d'une variété (qui appartient à l'homologie de la variété). En particulier, il donne une interprétation de la **classe de Godbillon-Vey** d'un feuilletage de codimension 1 en terme de cohomologie cyclique et donc d'accouplement avec la K-théorie.

Un feuilletage de codimension 1 est donné par une équation $\vartheta = 0$, où ϑ est une 1-forme satisfaisant la condition d'intégrabilité $d\vartheta \wedge \vartheta = 0$. On a donc $d\vartheta = \alpha \wedge \vartheta$ où α est une 1-forme. On considère alors la 3-forme $\alpha \wedge d\alpha$, et un calcul simple montre que sa classe de cohomologie ne dépend que du feuilletage. C'est l'invariant de Godbillon-Vey du feuilletage.

Alain Connes tire de son interprétation de cette classe un corollaire frappant : si la classe de Godbillon-Vey est non nulle, alors le flot des poids de l'algèbre de von Neumann préserve une mesure de masse finie (et elle est de type III). Ce superbe théorème de Connes peut aussi se démontrer de manière élémentaire sans aucune homologie cyclique, mais il montre toute la force et la beauté du point de vue géométrie non commutative. On voit là une des caractéristiques de la pensée d'Alain Connes qui est sa profonde unité. En développant la K-théorie et l'homologie cyclique des feuilletages, il n'oublie pas son point de départ, la classification des facteurs de type III.

Un autre succès important de cette méthode de géométrie *différentielle* non commutative est d'avoir donné la première preuve de la conjecture de Novikov pour les groupes hyperboliques au sens de Gromov. Dans ce cas l'algèbre de « fonctions lisses » sur le dual du groupe (vu comme espace non commutatif) est l'algèbre de fonctions à décroissance rapide définie par Paul Jolissaint.

Triplets spectraux et indice transversal

Jusque là, les théorèmes d'indices rencontrés dans le cadre des feuilletages sont des théorèmes d'indice **longitudinaux** : on considère des opérateurs elliptiques le long des feuilles du feuilletage, et l'indice d'un tel opérateur est un élément de la K-théorie de la C^*-algèbre du feuilletage, c'est-à-dire la K-théorie de l'espace non commutatif des feuilles. Pour obtenir un indice qui soit un nombre, on l'évalue sur une classe d'homologie de cet espace, c'est-à-dire de la cohomologie cyclique de l'algèbre du feuilletage. Mais il y a un problème plus difficile auquel Alain Connes va s'attaquer dans les années 1990, en collaboration avec Henri Moscovici. C'est la question du théorème d'indice **transverse** : cette fois on veut vraiment considérer un opérateur elliptique sur l'espace non commutatif des feuilles, c'est donc un opérateur transversalement elliptique, dont Connes et Moscovici montrent qu'il définit un élément de la K-homologie du feuilletage (la K-homologie est la théorie duale de la K-théorie) , donc une application de la K-théorie vers $\mathbf{Z}$, l'anneau des entiers. Le but du théorème d'indice transversal est de donner cette application de la K-théorie vers $\mathbf{Z}$ de façon concrète, c'est-à-dire au moyen d'un cocycle cyclique, pour lequel une formule explicite est donnée. Il y a d'abord une première difficulté : fabriquer de tels opérateurs transversalement elliptiques, sans aucune hypothèse sur le feuilletage : ainsi on ne veut pas se restreindre au cas où le feuilletage aurait une métrique riemannienne transversale invariante par holonomie. Pour cela, Connes et Moscovici procèdent en deux étapes ; d'abord on grossit l'espace pour prendre celui de toutes les métriques transversales, ensuite on admet, au lieu d'opérateurs elliptiques, des opérateurs hypoelliptiques. Moyennant quoi on obtient un **triplet spectral** sur l'espace non-commutatif des feuilles.

Etant donné un espace non commutatif dont la topologie est décrite par une C^*-algèbre A, on se donne un espace de Hilbert H dans lequel est représentée A, et un opérateur (non borné) autoadjoint D à résolvante compacte (l'opérateur $1 + D^2$ est d'inverse compact) et tel que les commutateurs $[D, a]$ soient bornés pour a dans une certaine sous-algèbre dense de A. Dans le cas de l'opérateur de Dirac sur une variété riemannienne, ce commutateur redonne la métrique, c'est-à-dire l'élément de longueur infinitésimal ds^2. La notion de triplet spectral permet de définir l'analogue non commutatif de la notion de variété riemannienne.

Il reste alors à calculer le caractère de Chern de ce triplet spectral. Ce qu'ils font en utilisant la notion de trace de Dixmier et de résidu de Wodzicki. La formule obtenue est à la fois très simple par son élégance, et très compliquée par le nombre de termes impliqués dès qu'on veut l'expliciter. Ainsi, même dans le cas d'un feuilletage de codimension 1, il faut une bonne centaine de pages pour mener le calcul ... Dans le cas général, cela est quasiment impossible, sauf si l'on peut comprendre un principe permettant d'organiser et de simplifier ces calculs. Comme c'est en général le cas en mathématiques, un tel principe simplificateur est fourni par la notion de **symétrie**. Classiquement, la symétrie est décrite par un groupe ; ici, dans une situation non commutative, c'est une **algèbre de Hopf** ou **groupe quantique** : il s'agit d'un objet qui se comporte comme un groupe, sauf que l'ensemble des éléments du groupe n'est pas vraiment un ensemble ou un espace, mais un espace non-commutatif. L'idée est que ce groupe quantique agit sur l'espace non commutatif, d'où l'on déduit une application caractéristique qui permet de pousser la cohomologie cyclique du groupe quantique (qui se calcule, c'est la cohomologie de Gelfand-Fuchs) dans la cohomologie cyclique de notre espace non commutatif. Il se fait alors que le caractère du triplet spectral est dans l'image de cette application caractéristique, et alors tout se calcule explicitement grâce à la cohomologie de Gelfand-Fuchs, au moyen de polynômes universels analogues à ceux qui apparaissent dans les théorèmes d'indice classiques.

Retour à la physique

C'est là qu'a lieu, de façon inattendue, un retour à la physique, via les algèbres de Hopf. Les calculs des physiciens en théorie quantique des champs reposent sur des méthodes de développement perturbatifs où les termes sont des intégrales divergentes, qui nécessitent une **renormalisation**. Ces techniques de renormalisation font apparaître la combinatoire des **diagrammes de Feynmann**. Des formules empiriques dites de Bogoliubov-Parashiuk ramènent le calculs de diagrammes compliqués à des diagrammes plus simples. Or en 1998, le physicien Dirk Kreimer découvre que ces formules qui n'étaient *a priori* que de simples recettes, traduisent l'existence d'un objet mathématique, qui n'est autre qu'une algèbre de Hopf ou groupe quantique. C'est là que Connes rencontre Kreimer, et ils découvrent ensemble que l'algèbre de Hopf de Kreimer et celle de Connes-Moscovici sont essentiellement les mêmes. Autrement dit, ce sont les mêmes règles de symétries quantiques qui régissent d'une part les calculs de théorie quantique des champs (permettant de calculer, par des méthodes perturbatives, des quantités physiquement observables), d'autre part les calculs de géométrie non commutative (donnant explicitement l'indice d'opérateurs transversalement elliptiques sur des feuilletages).

Un pas de plus a ensuite été franchi par Connes et Kreimer pour comprendre l'origine mathématique de cette algèbre de Hopf et son rôle dans le processus de renormalisation : ce processus n'est autre que la décomposition de Birkhoff, et ceci établit un lien direct et très simple avec le **problème de Riemann-Hilbert**. Ainsi, ce qui était au départ recette empirique, justifiée par l'expérience physique, est maintenant relié à un des grands problèmes des mathématiques, et non des moindres puisque le problème de Riemann-Hilbert est le 21$^{\text{ième}}$ de la liste des 23 problèmes proposés par David Hilbert au congrès international de Paris en 1900.

Enfin, dans sa collaboration récente avec Matilde Marcolli, Alain Connes a trouvé la signification mathématique de cette correspondance de Riemann-Hilbert : cette dernière est reliée à la **théorie de Galois motivique**, introduite par Grothendieck. Elle fait apparaître un groupe de symétrie dont Pierre Cartier avait conjecturé l'existence sous le nom de « groupe de Galois cosmique » et qui est donc de nature arithmétique. Ainsi la géométrie non commutative relie la physique à la théorie des nombres. On entrevoit ainsi le lien entre ces deux univers mystérieux, celui des particules élémentaires et celui des nombres premiers, qui ont toujours fasciné Alain Connes.

Pierre JULG
Université d'Orléans et CNRS
UMR 6628-MAPMO
BP 6759
45067 Orléans Cedex 2
Pierre.Julg@univ-orleans.fr

L'auteur remercie Alain Connes d'avoir relu son texte et de lui avoir suggéré quelques améliorations. Il remerci également Claire Anantharaman de lui avoir permis d'utiliser les posters qu'elle avait conçus en vue de la Journée Portes Ouvertes de l'Université d'Orléans.

Le mouvement brownien et son histoire, réponses à quelques questions

Jean-Pierre KAHANE*

Au moment où j'écris cet article, il y a dans la base de données MathSciNet 3454 références sur le mouvement brownien, plus 807 sur le bruit blanc, 476 sur le processus de Wiener, et plusieurs dizaines sur des sujets voisins (chaos homogène, chaos de Wiener, série de Fourier-Wiener). En 1905, il n'y avait rien. Que s'est-il passé dans l'intervalle ? Que peut-on saisir de l'histoire et de l'actualité du mouvement brownien en quelques pages ? L'article tente de répondre à une série de questions posées par un faux naïf, comme introduction à un domaine mathématique neuf et fascinant.

D'abord, qu'appelle-t-on mouvement brownien ?

Deux choses : un phénomène naturel, et un objet mathématique.

De quoi s'agit-il ?

Le phénomène naturel est le mouvement désordonné de particules en suspension dans un liquide. Il a été observé dès le 18ème siècle, sinon avant. L'objet mathématique est un processus gaussien dont la variance des accroissements est égale au temps écoulé. Norbert Wiener, qui l'a défini en 1923, l'appelait « the fundamental random function ».

Quel rapport entre les deux ?

C'est toute une histoire, dans laquelle les physiciens jouent un rôle majeur. Certains, au 19ème siècle, avaient pressenti que le mouvement des particules pouvait tenir à l'agitation moléculaire. Mais le grand départ est venu en 1905 avec Einstein. A l'origine, Einstein voulait tester la théorie cinétique moléculaire de la chaleur dans les liquides. Cela l'a mené à une formule qui permettait, à partir de l'observation du mouvement brownien, de calculer le nombre d'Avogadro. Jean Perrin a réalisé ce programme, et achevé ainsi d'établir la réalité des atomes ; il faut lire le grand classique qui en est résulté, *Les Atomes* (1912). Les observations de Jean Perrin ont inspiré Norbert Wiener. Dès ses premiers mots sur le mouvement brownien, Wiener cite un article de Perrin de 1909, qui évoque à ce sujet les courbes sans tangente des mathématiciens. Et Wiener se propose en effet de batir un modèle dans lequel les trajectoires sont continues, avec une vitesse infinie en tout point.

* *Département de mathématiques, Université Paris-Sud Orsay, Bâtiment 425, F-91405 Orsay Cedex*
Jean-Pierre.Kahane@math.u-psud.fr

D'où vient l'appellation de «mouvement brownien » ?

De Richard Brown, un grand botaniste écossais du début du 19ème siècle, qui s'intéressait à l'action du pollen dans la reproduction des plantes. Il a été amené, comme d'autres, à observer le mouvement irrégulier et incessant de particules de pollen en suspension dans l'eau. *A priori*, il s'agissait là d'un phénomène vital. Cependant les expériences que Brown a su monter avec des particules inorganiques montrent que c'est faux. L'apport du biologiste a été de sortir le phénomène de la biologie.

Mais, dans un autre sens, l'appellation vient de Paul Lévy. Ce sont les écrits de Paul Lévy qui ont fixé l'usage de nommer « mouvement brownien » le processus de Wiener.

L'histoire parait donc simple. Jusqu'à Brown, le phénomène est du ressort de l'histoire naturelle. Avec Einstein et Perrin, il est l'objet d'une théorie physique et il donne lieu à des expériences de physique. Avec Wiener et Lévy, il est défini mathématiquement et son étude mathématique commence. Est-ce bien celà ?

Oui et non. Si l'on ne retient comme origine du mouvement brownien des mathématiciens que le mouvement de particules de pollen en suspension dans l'eau, il est exact que Brown, Einstein, Perrin, Wiener et Lévy représentent les maillons essentiels de la chaîne qui va de la botanique à la mathématique en passant par la physique. Mais il y a bien d'autres maillons dans la chaîne, et surtout d'autres sources et d'autres liens dans l'histoire du mouvement brownien.

D'autres sources et d'autres liens ? Lesquels ?

On ne va pas pouvoir tout détailler, parce que le mouvement brownien occupe aujourd'hui une place centrale en mathématiques et qu'il est lié à la plupart de leurs branches : les équations d'évolution, l'analyse de Fourier, la théorie du potentiel, la théorie des fonctions d'une variable complexe, la géométrie et la théorie des groupes, l'analyse numérique,… A ces liens correspondent d'autres sources historiques, parmi lesquelles trois me semblent devoir être signalées en priorité.

1. L'équation de la chaleur (Fourier 1808) et sa diffusion ; la mise en évidence par Louis Bachelier dans sa thèse (1900) du processus de fluctuation des cours en Bourse et le fait que sa probabilité $p(x,t)dx$ que ce processus se trouve entre x et $x + dx$ au temps t obéit à l'équation de la chaleur (Bachelier parle du « rayonnement » de cette probabilité) ; c'est la source historique du lien entre mouvement brownien et mathématiques financières.
2. Les promenades au hasard, qui remontent au début du calcul des probabilités, avec l'image qu'en donne l'évolution de la fortune d'un joueur au jeu de pile ou face, puis, en 1905 de nouveau, les « random flights » de Pearson, qui sont des marches aléatoires isotropes dans le plan, et, en 1921, la première étude par Georges Polya des marches au hasard sur $\mathbf{Z}^d$ (récurrence ou transience, existence ou non de points multiples).
3. Les séries de puissances et les séries trigonométriques à coefficients aléatoires, dont l'idée remonte à Emile Borel en 1896, mais qui n'ont pu faire l'objet d'études rigoureuses, à partir de 1920, que lorsque se sont formalisées les notions de probabilité et de propriétés presque sûres (Steinhaus ; Paley et Zygmund ; Paley, Wiener et Zygmund) ; c'est d'ailleurs en collaboration avec Paley et Zygmund, en 1932, que Wiener a achevé son programme en montrant que la non-dérivabilité en tout point de sa fonction aléatoire était presque sûre.

Dans ces liens, il y a les retombées. Pouvez-vous en signaler quelques unes ?

Il y en a tant . . . D'abord, en me bornant aux trois sources que je viens de signaler, voici quelques éléments :

1. la thèse de Bachelier a été longtemps ignorée et elle est maintenant très populaire ; le mouvement brownien est l'outil de base des mathématiques financières ;

2. les promenades au hasard sur les groupes, les arbres, les graphes, les surfaces, sont de bons moyens pour explorer leur structure à l'infini ;
3. Wiener a proposé comme programme d'unifier la présentation des séries trigonométriques à coefficients aléatoires, dont fait partie la série de Fourier-Wiener qui représente le mouvement brownien, et ce programme a débouché sur une nouvelle théorie, les probabilités dans les espaces de Banach.

Par ailleurs, une retombée essentielle du processus de Wiener est l'axiomatique de Kolmogorov en 1933, qui ne se borne pas au classique $(\Omega,\mathcal{A},\mathbf{P})$, mais qui montre, exactement à la manière de Wiener, comment construire l'espace de probabilité à partir d'une famille $(\Omega_{\mathcal{C}},\mathcal{A}_{\mathcal{C}},\mathbf{P}_{\mathcal{C}})$, adaptée au processus à probabiliser. Il faudrait ensuite citer les équations différentielles stochastiques, l'intégrale d'Itô etc.

Peut-on revenir à l'histoire esquissée tout à l'heure ? Comment est-on passé des équations d'Einstein au processus de Wiener ?

Einstein a procédé par étapes et présenté sa théorie sous différents angles. L'essentiel est la formule

$$\overline{(\Delta x)^2} = \frac{RT}{N}\,\frac{1}{3\pi\mu a}\tau$$

où R est la constante des gaz parfaits, T la température absolue, N le nombre d'Avogadro, μ la viscosité du liquide, a le rayon de la particule, supposée sphérique, et τ le temps correspondant au déplacement Δx. Quant à $\overline{(\Delta x)^2}$, c'est la moyenne du carré des déplacements, dans une direction donnée, d'un grand nombre de particules pendant un intervalle de temps donné, de durée τ ; et c'est aussi la moyenne du carré des déplacements d'*une* particule au cours d'intervalles de temps consécutifs de durée τ. C'est surtout sous la dernière forme que Jean Perrin l'a utilisée. Mais c'est la première que la version mathématisée traduit le mieux : d'après Wiener, l'ensemble des fonctions continues de $\mathbf{R}^+$ dans $\mathbf{R}^+$, nulles en 0, porte une mesure de probabilité, la mesure de Wiener, telle que l'intégrale suivant cette mesure du carré du déplacement entre deux temps t_1 et t_2 est $t_2 - t_1$:

$$\mathbf{E}\left((X(t_2) - X(t_1))^2\right) = t_2 - t_1 \quad (t_2 > t_1)$$

et que, de plus, pour tout $t_0 > 0$, les déplacements à partir du temps t_0, $X(t) - X(t_0)$ $(t > t_0)$ soient indépendants de ce qui s'est passé avant t_0.

Pour les physiciens, le coefficient était important, n'est-ce pas ?

Très important. C'est ce qui permettait de passer de l'observation à la détermination de N. Il est bon de dire un mot de Smoluchowski et de Langevin. Quand Einstein l'a faite, la théorie du mouvement brownien était dans l'air. Indépendamment d'Einstein, Marian Smoluchowski en a publié sa version en 1906, et le résultat est le même, à l'exception d'un facteur $64/27$ dans l'expression de $\overline{(\Delta x)^2}$. En 1908, Paul Langevin publie une note aux Comptes rendus pour dire que, rectification faite, l'approche de Smoluchowski mène exactement à la formule d'Einstein. Puis, en tout petits caractères, il expose sa propre approche, qui est lumineuse.

Lumineuse au point d'être exposée ici ?

Pourquoi pas ? Mais il faut commencer, de façon un peu surprenante, en considérant que les particules browniennes ont une vitesse, soit :

$$u = \frac{dx}{dt}$$

dans la direction Ox. Si leur masse est m, l'énergie moyenne correspondante est $1/2m\overline{u^2}$. Selon l'hypothèse fondamentale de la mécanique statistique, c'est aussi l'énergie moyenne d'une molécule, soit RT/N :

$$m\overline{u^2} = \frac{RT}{N}.$$

Pour une particule sphérique de masse m et de rayon a, dans un liquide viscosité μ, l'équation du mouvement est *a priori* :

$$m\frac{du}{dt} = -6\pi a\mu u$$

et, selon cette équation, le mouvement s'arrête rapidement. Or on le voit se poursuivre. Il y a donc des chocs moléculaires, qui entretiennent ce mouvement. L'équation que propose Langevin est

$$m\frac{du}{dt} = -6\pi a\mu u + X, \qquad (1)$$

avec une hypothèse minimale sur X : « sur la force complémentaire X nous savons qu'elle est indifféremment positive et négative, et sa grandeur est telle qu'elle maintient l'agitation de la particule ». De cette équation (1) Langevin tire

$$m\frac{d(xu)}{dt} = mu^2 - 6\pi a\mu xu + xX,$$

puis, en prenant des valeurs moyennes, avec $\overline{xX} = 0$,

$$m\frac{d(\overline{xu})}{dt} = m\overline{u^2} - 6\pi a\mu\overline{xu} = \frac{RT}{N} - 6\pi a\mu\overline{xu}.$$

Cela donne

$$2\overline{xu} = \frac{RT}{N}\frac{1}{3\pi a\mu} + C\exp(-\frac{6\pi a\mu}{m}t)$$

et, compte tenu des valeurs numériques, le dernier terme est négligeable pour $t >> 10^{-8}$ sec. Le régime est donc pratiquement permanent. En intégrant,

$$\overline{x^2(t+\tau) - x^2(t)} = \frac{RT}{N}\frac{1}{3\pi a\mu}\tau.$$

Or l'accroissement $x(t+\tau) - x(t)$ a une valeur moyenne nulle et il est indépendant de $x(t)$, donc

$$\overline{x^2(t)} + \overline{(x(t+\tau) - x(t))^2} = \overline{x^2(t+\tau)},$$

et l'équation d'Einstein en résulte.

Est-ce que ⟨1⟩ est une équation stochastique au sens où vous l'entendiez tout a l'heure ?

Tout à fait. C'est le prototype des équations stochastiques. J.-L. Doob en a repris la théorie en 1942 en prenant pour X le bruit blanc, qui est formellement la dérivée du processus de Wiener. Il n'est pas bien difficile alors d'intégrer l'équation (1), et la solution, $u(t,\omega)$, s'appelle le processus d'Ornstein-Uhlenbeck. Ainsi, une fois faite la théorie mathématique du mouvement brownien sous la forme du processus de Wiener, qui est presque sûrement non dérivable en tout point, on peut bâtir une théorie donnant la loi explicite, en fonction du processus de Wiener, de la dérivée du mouvement brownien. Ce n'est pas un tour de passe-passe : c'est seulement affaire de changement d'échelle et de choix de modèle. Mais cela justifie la prudence de Wiener, parlant de « fundamental random func-

tion » et non de « mouvement brownien » : il y a plusieurs idéalisations possibles du mouvement brownien. Après Doob, c'est K. Itô qui a élaboré une théorie de l'intégration des équations différentielles stochastiques, à partir d'un outil nouveau, l'intégrale d'Itô, et c'est maintenant d'usage courant dans les mathématiques financières.

Une fois établie l'équation d'Einstein, que restait-il donc à faire aux mathématiciens ?

L'équation d'Einstein est une loi physique. Il n'y a pas besoin de prouver l'existence du mouvement brownien : il est là, on l'observe, on l'étudie et on l'utilise.

Pour les mathématiciens, il reste à en faire un modèle mathématique. L'idée, *grosso modo*, est que ce modèle doit être pour les fonctions ce que la variable aléatoire normale (gaussienne normalisée) est pour les nombres : « the fundamental random function ». Wiener construit le modèle en se restreignant d'abord aux temps entiers : c'est simplement la suite des sommes partielles d'une série de v.a. normales indépendantes. Puis, aux temps demi-entiers, il procède par interpolation, et ainsi de suite. C'est le travail, monumental pour l'époque, constitué par « Differential space », l'espace des différences. Les différences sont des variables aléatoires gaussiennes centrées.

Est-ce toujours un travail monumental que de définir « the fundamental random function », le processus de Wiener ?

Non. Pour faire vite, on part d'un espace de probabilité Ω tel que $L^2(\Omega)$ contienne un espace de Hilbert $\mathcal{H}$ constitué de variables gaussiennes centrées. Commençons par des espaces $L^2(\Omega)$ et $\mathcal{H}$ réels. Ainsi $X \in \mathcal{H}$ signifie que X est une fonction de $\omega \in \Omega$, et que

$$\mathbf{E}(e^{iuX}) = e^{-u^2/2||X||^2} \quad (u \in \mathbf{R})$$
$$||X||^2 = \mathbf{E}(|X|^2),$$

$\mathbf{E}()$ désignant l'espérance, c'est-à-dire l'intégrale sur Ω. Il s'agit de construire une fonction $X(t,\omega)$, où $\omega \in \Omega$ et t, le temps, appartient à un intervalle réel I, avec les propriétés voulues. La clé de la construction est une application linéaire et isométrique de $L^2(I)$ dans $\mathcal{H}$, qu'on désignera par W. Le processus de Wiener est l'image par W de la fonction indicatrice de l'intervalle $[0,t]$. On a ainsi un modèle du mouvement brownien à valeurs dans $\mathbf{R}$. En prenant $L^2(\Omega)$, $\mathcal{H}$ et $L^2(I)$ complexes, on a un modèle du mouvement brownien à valeurs dans $\mathbf{C}$. En choisissant les composantes indépendantes, on a un modèle dans $\mathbf{R}^n$.

C'est si rapide que je n'y comprends rien. En nous limitant au cas réel, où cela nous mène-t-il ?

Pour fixer les idées, prenons d'abord $I = \mathbf{R}$, et regardons $X(t,.) = W(1_{[0,t]})$, $t \in \mathbf{R}$. C'est une courbe paramétrée dans $\mathcal{H}$ (c'est-à-dire un processus gaussien) dont le carré de la longueur d'une corde est bien facile à calculer ; c'est $||X(t,.) - X(s,.)||^2_{\mathcal{H}} = ||1_{[0,t]} - 1_{[0,s]}||^2_{L^2(\mathbf{R})} = |t-s|$, ce qui est l'équation d'Einstein normalisée. Si l'on effectue une translation sur le paramètre t, la courbe glisse isométriquement sur elle-même : c'est une *hélice*. Si l'on prend trois points X_1, X_2, X_3 sur cette hélice dans l'ordre des paramètres, on a $||X_3 - X_2||^2 + ||X_2 - X_1||^2 = ||X_3 - X_1||^2$, le triangle $X_1X_2X_3$ est rectangle en X_2 : l'*hélice brownienne* est donc une figure très remarquable, qui ne peut se réaliser que dans un espace de Hilbert de dimension infinie. L'orthogonalité dans $\mathcal{H}$ de $X_2 - X_1$ et $X_3 - X_1$ signifie leur indépendance : les accroissements dans le futur sont indépendants du passé. Beaucoup de propriétés se lisent sur cette hélice brownienne. Si l'on partage un intervalle de temps $[s,t]$ au moyen de points de subdivision $s = t_0 < t_1 < \ldots < t_n = t$, on a $\sum_{i=1}^{n} ||X(t_j,.) - X(t_{j-1},.)||^2 = t - s$: la variation quadratique le long de l'arc image de $[s,t]$ est la lon-

gueur de $[s,t]$. L'hélice brownienne est de dimension 2 au sens de Hausdorff et la mesure en dimension 2 d'une portion de l'hélice est la mesure linéaire de l'intervalle de temps correspondant. C'est une courbe très régulière.

Pourtant on insiste sur l'irrégularité du mouvement brownien. Comment est-ce compatible avec la régularité de l'hélice brownienne ?

C'est tout le charme des processus gaussiens. Leur image sous forme d'une courbe dans $\mathcal{H}$ peut être très lisse, et leurs réalisations très capricieuses.

Qu'est-ce donc qu'une réalisation ?

C'est $X(t,\omega)$ comme fonction de t, quand ω est donné. On s'intéresse aux propriétés presque sûres des réalisations, c'est-à-dire qui ont lieu pour presque tout ω. Par exemple, quand il s'agit du processus de Wiener, la fonction $t \to X(t,\omega)$ est presque sûrement continue, hölderienne d'ordre $1/2-\varepsilon$ pour tout $\varepsilon > 0$ et nulle part dérivable.

Est-ce difficile à voir ?

Oui et non. Wiener a donné dans Differential space beaucoup de propriétés des réalisations, mais il lui a fallu attendre la collaboration avec Paley et Zygmund pour établir la non-dérivabilité partout. Si l'on part de l'application $W : L^2(I) \to \mathcal{H}$, il est commode d'introduire une base de $L^2(I)$, disons (u_n), et son image dans $\mathcal{H}$, disons (ξ_n), qui est une suite de variables normales indépendantes. L'image d'une $f = \sum \hat{f}_n u_n$ est $W(f) = \sum \hat{f}_n \xi_n$. Ainsi $X(t,\omega) = \sum a_n(t)\xi_n(\omega)$ $(a_n(t) = \int_0^t u_n(s)ds)$, série convergente dans $\mathcal{H}$ quand t est donné. Il s'agit d'étudier la convergence de cette série quand ω est donné. Or on connait très bien le comportement asymptotique de la suite $\xi_n(\omega)$ pour presque tout ω : si $n \in \mathbf{N}$, $\xi_n(\omega) = O(\sqrt{\log n})$ p.s. Connaissant les $a_n(t)$, l'étude de la série aléatoire $\sum a_n(t)\xi_n(\omega)$ donne des propriétés presque sûres de la fonction $X(.,\omega)$. Si l'on prend pour (u_n) le système trigonométrique, on obtient la série de Fourier-Wiener. Dès 1923 dans Differential space, Wiener avait montré que, sur tout intervalle fini, sa « fundamental random function » était développable, à un terme affine près, en série trigonométrique aléatoire. En 1933, dans le livre où il décrit les travaux faits en commun avec Paley, il part de cette série pour définir la fonction. Dans cette approche, on doit d'abord vérifier que la série converge presque sûrement vers une fonction continue, et la mesure de Wiener apparaît comme l'image de la probabilité donnée sur Ω par la série de Fourier-Wiener. Dans un article postérieur de 1938, « The homogeneous chaos », Wiener prend $I = \mathbf{R}$ et les fonctions de Hermite pour u_n. Mais le plus commode pour l'étude des propriétés locales de la fonction $X(.,\omega)$ est de prendre $I = [0,1]$ et pour u_n le système des fonctions de Haar ($u_0 = 1$ et les autres u_n sont portées par des intervalles dyadiques où elles prennent des valeurs opposées sur les moitiés gauche et droite.) Les $a_n(.)$ sont alors des fonctions triangles, et les calculs sont facilités.

Peut-on lister quelques propriétés presque sûres de la fonction $X(.,\omega)$ que l'on obtient de cette façon ?

Allons y.

1. C'est une fonction continue.
2. Sur tout intervalle borné son module de continuité est $O(\sqrt{h\log 1/h})$.
3. En tout point t $\limsup_{h\to 0} \dfrac{X(t+h,\omega)-X(t,\omega)}{\sqrt{|h|}} > 0$ (Dvoretzky 1963).

4. En presque tout point t $\limsup_{h\to 0} \dfrac{X(t+h,\omega)-X(t,\omega)}{\sqrt{2|h|\log\log 1/h}} = 1$ et à l'infini $\limsup_{t\to\infty} \dfrac{X(t,\omega)}{\sqrt{2t\log\log t}} = 1$. (c'est la « loi du logarithme itéré » de Khintchine, ramenée à distance finie par Paul Lévy).

5. Mais il y a des points « rapides », où

$\limsup_{h\to 0} \dfrac{X(t+h,\omega)-X(t,\omega)}{\sqrt{|h|\log 1/h}} > 0$ (Orey-Taylor 1974) et des points « lents », où $X(t+h)-X(t) = O(\sqrt{|h|})$ $(h \to 0)$ (Kahane 1974).

Est-ce tout ?

Bien sûr que non. Comme j'ai évoqué le « chaos homogène » de Wiener, il est bon de dire ce qu'il appelle le chaos gaussien pur, que l'on appelle couramment le bruit blanc. C'est, au choix, l'opérateur W, ou son expression dans la base (u_n) sous la forme $\sum u_n(t)\xi_n(\omega)$, ou la distribution de Schwartz aléatoire représentée par cette série. Si l'on prend pour E un ensemble mesuré et que l'on part de $L^2(E)$ au lieu de $L^2(I)$, on a le bruit blanc sur E. On définit le bruit blanc complexe aussi bien que réel. Par exemple, sur tout groupe abélien localement compact, muni de la mesure de Haar, on a un bruit blanc complexe, et on peut définir sa transformée de Fourier : c'est le bruit blanc sur le groupe dual. En particulier, la transformée de Fourier du bruit blanc sur $\mathbb{R}^n$ est le bruit blanc sur $\mathbb{R}^n$. Mais surtout, il y a toutes les merveilleuses propriétés découvertes par Paul Lévy.

Celui qui a désigné le processus de Wiener comme «mouvement brownien » ?

Celui-là même. C'est à partir de Paul Lévy qu'on parle du mouvement brownien comme objet mathématique. Pour marquer le coup nous allons l'écrire $B(t)$ au lieu de $X(t,\omega)$. On peut définir $B(t)$ à valeurs réelles, ou complexes, ou dans $\mathbb{R}^n$, et ses propriétés – par exemple l'existence de points multiples, ou la récurrence au voisinage d'un point – dépendent de la dimension. En dimension 1, Paul Lévy a donné la loi de l'ensemble des zéros de $B(t)$ et montré comment construire le graphe à partir d'« excursions » sur les intervalles contigüs. C'est un très beau sujet, mais je préfère vous parler de la dimension 2.

C'est-à-dire du mouvement brownien complexe ?

Exactement. Paul Lévy a découvert sa relation avec les fonctions analytiques d'une variable complexe. Elle est très simple et intuitive. A un changement de temps près, la loi du mouvement brownien plan est bien définie par le fait qu'elle est localement isotrope et que les trajectoires sont continues. Cette propriété est conservée par représentation conforme. Donc, si l'on applique une fonction analytique $F(z)$ à $B(t)$, la fonction aléatoire obtenue, $F(B(t))$, est encore un mouvement brownien, au changement de temps près. Cela permet d'utiliser des fonctions analytiques pour établir des propriétés du mouvement brownien plan, et aussi d'utiliser le mouvement brownien plan pour démontrer des propriétés des fonctions analytiques.

Pouvez-vous donner des exemples ?

En voici un. Prenons $B(t)$ partant de 0 au temps $t = 0$, et $F(z) = e^z - 1$, de sorte que $F(B(t))$ a le même point de départ. Comme $F(B(t))$ ne prend jamais la valeur -1, il est presque sûr qu'il en est de même pour $B(t)$. De même en remplaçant -1 par n'importe quel nombre complexe $\neq 0$. Donc l'aire de la courbe décrite par $B(t)$ $(t \geqslant 0)$ est nulle p.s.. En voici un autre, plus subtil. Prenons pour $F(z)$ une fonction analytique et bornée dans le disque $|z| < 1$ (on écrit $F \in H^\infty(D)$), $\neq 0$ et arrêtons $B(t)$, partant de 0, au moment θ où il rencontre le cercle $|z| = 1$, disons, en ζ. Quand t tend en croissant vers θ, $F(B(t))$, étant borné, tend vers une limite, disons Z. Comme (presque sûrement),

$B(t)$ coupe une infinité de fois au voisinage de $t = \theta$ tout segment de droite donné dans le disque $|z| \leqslant 1$ et aboutissant à ζ, il s'ensuit que $F(z)$ tend vers Z quand z tend vers ζ dans un angle compris entre deux tels segments, c'est-à-dire « non-tangentiellement ». Cela valant presque sûrement, vaut pour presque tout ζ à la frontière du disque. De plus, la probabilité que $F(B(t))$ prenne la valeur 0 en dehors de $t = 0$ étant nulle, il est presque sûr que $Z \neq 0$, donc l'ensemble des ζ pour lesquels $Z = 0$ est de mesure nulle. On vient de démontrer à l'aide du mouvement brownien un théorème de Fatou sur les fonctions $F \in H^\infty(D)), \neq 0$: une telle fonction admet presque partout à la frontière une limite non-tangentielle $\neq 0$. Ce n'est pas la démonstration la plus simple du théorème de Fatou, mais elle s'étend bien au delà des $F \in H^\infty(D)$: elle est valable dès que l'arrêt de $B(t)$ à la frontière de D se traduit presque sûrement par l'arrêt de $F(B(t))$, et cela vaut, par exemple, quand $F \in H^p(D)$ avec $p > 0$, ce que la théorie classique n'établit qu'assez laborieusement[1]. La théorie classique repose sur la considération des zéros de $F(z)$ dans D et la théore de Nevanlinna. Le mouvement brownien a l'avantage d'ignorer les zéros, en les contournant.

Il y a eu des travaux récents sur le mouvement brownien plan. Peut-on en avoir une idée ?

On trouve maintenant facilement des images du mouvement brownien plan, c'est-à-dire de l'ensemble aléatoire $B([0,1])$. Nous savons que cet ensemble est d'aire nulle (sous-entendu : presque sûrement). Son aspect est celui d'un tapis déchiquetté sur les bords et presque entièrement dévoré par les mites. Mais il lui reste quand même de l'étoffe : sa dimension de Hausdorff est 2. Sa frontière est constituée de contours disjoints, ayant tous le même aspect fractal, avec des tailles différentes. La statistique de leurs diamètres (Werner) est connue. L'attention s'est portée sur leur dimension, qui est la dimension de la frontière extérieure. Benoit Mandelbrot a conjecturé en 1982 que cette dimension est 4/3. En s'inspirant des résultats de G.F. Lawler et W. Werner, B. Duplantier a donné de cette conjecture une interprétation physique convainquante *via* la « gravité quantique » (une « démonstration heuristique »). En 2000, en utilisant un processus introduit par O. Schramm, le SLE (décrit dans l'article de Kenyon et Werner des Images des mathématiques 2004), Lawler, Schramm et Werner ont donné la démonstration mathématique attendue.

Y a-t-il encore des choses à trouver sur le mouvement brownien plan ?

Oui, le principal peut-être, dont parle l'article de Kenyon et Werner. La conjecture de Mandelbrot était en réalité que la frontière du mouvement brownien plan donne une image du « mouvement brownien auto-évitant », dont on a une bonne idée, mais qu'on ne sait pas encore définir. C'est un sujet d'intérêt commun aux physiciens et aux mathématiciens.

Ce qui est fascinant est la variété et la puissance des outils mis en oeuvre dans cette étude, et l'appui que se prêtent mutuellement physiciens et mathématiciens. Au moment d'écrire cet article, j'ai eu le bonheur d'entendre au séminaire Poincaré (le séminaire Bourbaki des physiciens) un exposé de Bertrand Duplantier sur le mouvement brownien. C'est à la fois une mise au point historique, avec toutes les références souhaitables, et un aperçu stimulant sur les recherches en cours. Oui, il reste encore bien des choses à trouver, et si vous voulez vous en convaincre, lisez Werner, lisez Duplantier.

1. $F \in H^p(D)$ signifie que les intégrales de $|F|^p$ sur les cercles de centre O et de rayon < 1 sont uniformément bornées.

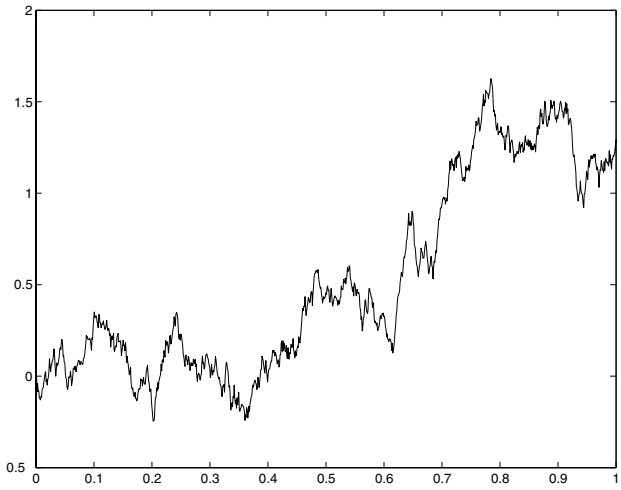


Figure 1 – *Graphe du mouvement brownien linéaire*

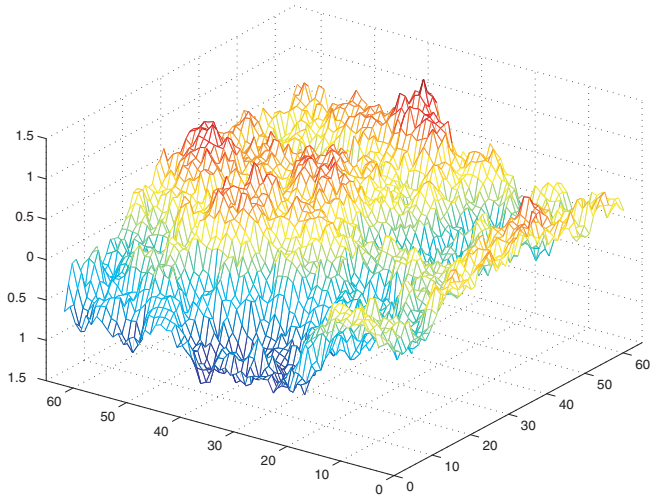


Figure 2 – *Drap brownien*

Pour en savoir plus

WERNER (W.), Les chemins de l'aléatoire, *Pour la Science,* n° 286, pp. 68-74 (août 2001).

DUPLANTIER (B.), *Le mouvement brownien*, Séminaire Poincaré 1, pp. 155-212 (avril 2005).

Post-scriptum

Bertrand Duplantier m'a fait connaître deux textes de grand intérêt. Le premier est « Brownian motion in Clarkia pollen : a reprise of the first observations », par Brian J. Ford, dans *The Microscope* 40(4) (1992), pp. 235-241. Il s'avère que les particules observées par Brown n'étaient pas les grains de pollen eux-mêmes, mais des particules beaucoup plus ténues contenues dans ces grains. L'auteur a refait l'expérience et en a tiré une vidéo.

Le second explicite une coïncidence remarquable, que signale Duplantier dans son article p. 163 : en même temps qu'Einstein et indépendamment, un physicien australien, William Sutherland, donnait une théorie et une formule analogues, dans le but de déterminer le poids moléculaire de l'albumine. L'histoire est racontée par Bruce H.J. Mc Kellar (Australien lui aussi) sous le titre « The Sutherland-Einstein Equation », en tête du AAPPS Bulletin, February 2005.

Duplantier fait remarquer que les approches de Sutherland et d'Einstein sont thermodynamiques, et celle de Smoluchowski probabiliste.

Wendelin Werner

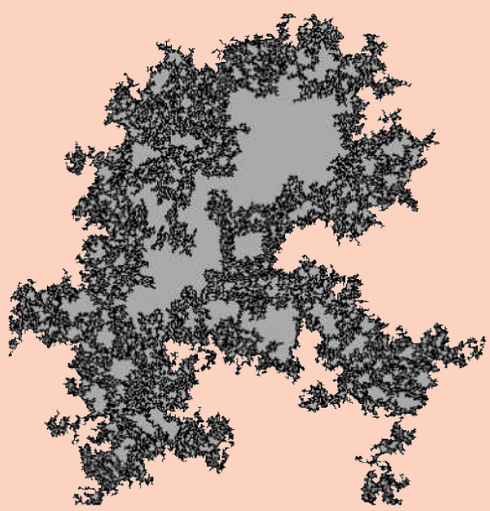
Les travaux du mathématicien français Wendelin Werner ont été récompensés par la prestigieuse Médaille Fields lors du dernier Congrès International des Mathématiciens qui s'est tenu à Madrid du 22 au 30 août 2006. Ancien élève de l'Ecole normale supérieure de Paris, Wendelin Werner a d'abord été chercheur au CNRS et il est depuis 1997 professeur à l'Université Paris-Sud (Orsay). Il est aussi membre de l'Institut universitaire de France et professeur à temps partiel à l'Ecole normale supérieure.

Avec Wendelin Werner, la Médaille Fields distingue pour la première fois un spécialiste de la théorie des probabilités. Ses travaux se placent à l'interface entre cette théorie et la physique statistique. Le fait que les modèles étudiés possèdent des propriétés asymptotiques d'invariance conforme conduit aussi à l'utilisation d'outils sophistiqués d'analyse complexe.

Un exemple simple mais significatif des résultats de Wendelin Werner est fourni par l'étude de la probabilité de non-intersection de deux marches aléatoires planes. Considérons une particule qui se déplace de manière aléatoire sur le réseau $\mathbf{Z}^2$ selon les règles suivantes : à l'instant initial la particule se trouve à l'origine puis, à chaque instant entier strictement positif, elle saute en l'un des quatre plus proches voisins du point occupé précédemment, avec la même probabilité $1/4$ pour chacune des possibilités, indépendamment du passé. La trajectoire de la particule entre les instants 0 et n est l'ensemble des points qu'elle visite entre ces deux instants. Considérons aussi une seconde particule qui se déplace selon les mêmes règles, indépendamment de la première. On s'intéresse alors à la probabilité que l'origine soit le seul point commun aux trajectoires des deux particules entre les instants 0 et n. On savait depuis assez longtemps que cette probabilité se comporte comme (une constante fois) n^{-a} quand n est grand. La valeur exacte de l'exposant $a = 5/8$, conjecturée par les physiciens théoriciens Duplantier et Kwon en 1988, n'a pu être calculée rigoureusement que grâce aux travaux récents de Wendelin Werner et de ses collaborateurs Gregory Lawler et Oded Schramm. De manière inattendue, ce calcul a nécessité l'introduction de nouveaux processus aléatoires, les évolutions stochastiques de Loewner ou SLE en anglais. Les processus SLE ont beaucoup d'autres applications spectaculaires à différents modèles de physique statistique, comme la percolation, les marches aléatoires auto-évitantes ou modèles de polymères, ou encore les arbres couvrants sur un réseau. Le développement de telles applications, par Wendelin Werner et ses collaborateurs, a constitué un pas de géant dans la compréhension mathématique de ces modèles.

Après celle obtenue par Laurent Lafforgue en 2002, la Médaille Fields de Wendelin Werner témoigne une nouvelle fois de la grande vitalité de l'école mathématique française.

Jean-François Le Gall,
École normale supérieure



La théorie des sondages

Michel LEJEUNE*

On présente ici les principes et enjeux simples de la théorie des sondages pouvant susciter des désirs d'approfondissement.

Introduction

La théorie des sondages ne fait pas partie, en France, des connaissances usuelles des statisticiens même si la pratique des sondages, quant à elle, est très répandue. Elle s'est développée à partir des années 1930 dans le monde anglo-saxon ainsi qu'en Inde. Bien que reposant sur les mêmes principes que la statistique mathématique classique elle en diffère sensiblement dans son esprit en raison d'objectifs spécifiques.

Cette théorie se consacre essentiellement au problème de la sélection de l'échantillon (voir ci-après la notion de plan de sondage qui fait le pendant des plans d'expériences en théorie classique) et à la recherche d'estimateurs. Du fait qu'elle porte sur des populations finies, d'existence bien concrète, elle ne peut ignorer les contraintes du monde réel, ce qui n'est peut-être pas sans lien avec le faible intérêt qu'elle suscite chez nous.

Plan de sondage et probabilités d'inclusion

On considère une *population* comprenant N *individus* parfaitement identifiés par un numéro d'ordre. Pour ce qui suit il nous suffira de ne retenir que ces numéros d'ordre et nous définissons ainsi la population $U = \{1,\ldots,k,\ldots,N\}$. Notons que les vocables de population et individus sont purement conventionnels. Les parties de cette population sont appelées *échantillons*. Dans cette présentation nous n'envisagerons que la situation où l'échantillon à sélectionner est de *taille* (cardinal) fixée, notée n, et désignerons simplement par S l'ensemble des échantillons de taille n. Commençons par donner quelques définitions.

Définition 1. *On appelle plan de sondage une loi de probabilité définie sur S.*

Concrètement le plan de sondage définit, pour chaque échantillon, la probabilité qu'il soit sélectionné *via* le mécanisme aléatoire utilisé.

Définition 2. *Soit un plan de sondage p et S_k l'ensemble des échantillons contenant l'individu k. On appelle probabilité d'inclusion de l'individu $k \in U$:*

$$\pi_k = \sum_{s \in S_k} p(s)$$

* *Université Pierre Mendès France, LABSAD (BSHM), 38040 Grenoble Cedex 09.*
michel.lejeune@upmf-grenoble.fr

Il s'agit donc de la probabilité que cet individu appartienne à l'échantillon sélectionné. Le fait d'avoir un échantillon de taille n se traduit par $\sum_{k\in U}\pi_k = n$. En effet, soit I_k la variable indicatrice de la sélection de l'individu k, on a $\pi_k = E(I_k)$ et :

$$\sum_{k\in U}\pi_k = \sum_{k\in U}E(I_k) = E(\sum_{k\in U}I_k) = n.$$

On définit de même la probabilité qu'un couple d'individus $\{k,l\}$ soit dans cet échantillon, en considérant l'ensemble S_{kl} des échantillons contenant ce couple.

Définition 3. *On appelle probabilité d'inclusion d'ordre 2 des individus k et l :*

$$\pi_{kl} = \sum_{s\in S_{kl}} p(s)\,.$$

Pour illustrer ces notions considérons le *plan simple sans remise* (PSSR), correspondant au cas où p est uniforme sur S :

$$\forall s \in S, p(s) = \binom{N}{n}^{-1}.$$

Comme il y a $\binom{N-1}{n-1}$ échantillons comprenant un individu donné, on a :

$$\forall k \in U, \quad \pi_k = \binom{N-1}{n-1}\binom{N}{n}^{-1} = \frac{n}{N}.$$

Ce résultat, à savoir que la probabilité d'être sélectionné, pour un individu donné, est égale au *taux de sondage* $\frac{n}{N}$, est tout à fait intuitif. On montre aisément que l'on a $\pi_{kl} = \dfrac{n(n-1)}{N(N-1)}$.

Remarque 4. *Notre définition d'échantillon exclut le cas du plan simple avec remise (PSAR), lequel correspond à la théorie classique de l'échantillonnage aléatoire. Les notions précédentes et la théorie générale peuvent aussi être développées dans cette situation. En pratique on n'effectue pas de plans avec remise car, on le sent bien intuitivement, le risque d'observer plusieurs fois le même individu constitue une perte d'information. La théorie montre, dans diverses situations, qu'un estimateur sans remise est meilleur qu'un estimateur avec remise et, même, qu'il est préférable, dans le cas d'un échantillon avec remise, de ne prendre en compte chaque individu qu'une seule fois.*

Les estimateurs

On s'intéresse maintenant à une certaine « variable » réelle observable sur chaque individu. Notons y_k la valeur prise par l'individu k. On souhaite estimer une caractéristique de la variable, le plus souvent son total (ou sa moyenne). Pour le total t_y, Horvitz et Thompson ont proposé un estimateur pour un plan quelconque. Soit $Y_1, \ldots, Y_n$ les variables aléatoires correspondant à la sélection de n individus selon le plan de sondage, cet estimateur est :

$$\widehat{t}_y^{HT} = \sum_{i=1}^{n}\frac{1}{\pi_i}Y_i\,.$$

Les $\frac{1}{\pi_i}$ sont appelés *poids de sondage.* Cet estimateur est sans biais. En effet :

$$\widehat{t}_y^{HT} = \sum_{k\in U} \frac{1}{\pi_k} I_k y_k, \text{ d'où } E(\widehat{t}_y^{HT}) = \sum_{k\in U} \frac{1}{\pi_k} E(I_k) y_k = \sum_{k\in U} y_k = t_y.$$

Il est généralement retenu car on montre qu'il est le seul à avoir cette propriété parmi les estimateurs fonctions linéaires des Y_i. De plus on sait exprimer sa variance et un estimateur sans biais de sa variance, *via* les probabilités d'inclusion d'ordres 1 et 2.

On peut évidemment étudier cet estimateur par une voie directe. Pour les plans PSAR et PSSR il est clair que $\frac{N}{n}\sum_{i=1}^{n} Y_i$ est sans biais pour t_y (chaque Y_i ayant une loi uniforme sur les y_k, son espérance est $m_y = \frac{1}{N} t_y$, la moyenne de la population). Notons que l'estimateur de Horvitz-Thompson pour la moyenne de la population est la moyenne de l'échantillon $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Le calcul de la variance montre les difficultés inhérentes à la théorie des sondages. Pour le plan PSAR les Y_i sont indépendantes et le calcul est immédiat. Pour le PSSR interviennent les covariances et le calcul demande une certaine adresse (le lecteur pourra tenter une démonstration en utilisant $\sum_{i=1}^{n} Y_i = \sum_{k\in U} I_k y_k$, $Var(I_k) = \pi_k(1-\pi_k)$ et $Cov(I_k, I_l) = \pi_{kl} - \pi_k \pi_l$). On trouve :

$$Var(\overline{Y}) = \left(\frac{N-n}{N-1}\right)\frac{v^2}{n}$$

où v^2 est la variance de la population :

$$v^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - m_y)^2.$$

Un apport essentiel de la théorie des sondages est d'établir la *précision* d'une estimation (vulgairement : la fourchette). Celle-ci est définie par la demi-longueur d'un intervalle de confiance à 95 %. En vertu du théorème central limite qui donne une bonne approximation pour les tailles usuelles d'échantillons, cet intervalle repose sur la loi de Gauss (moyennant une deuxième approximation due à l'estimation de la variance v^2). Dans le cas PASR on adopte ainsi pour précision sur l'estimation de la moyenne :

$$1,96\sqrt{\left(1-\frac{n}{N}\right)\frac{s^2}{n}}$$

où s^2 est la variance de l'échantillon qui estime sans biais $\frac{N}{N-1}v^2$.

Le principe de stratification

Dans cette brève présentation il n'est pas possible de présenter les principaux plans de sondage. Nous nous concentrons sur le plan stratifié en raison de ses liens avec les pratiques très répandues d'utilisation de quotas et de redressement d'échantillons. Ces procédures reposent sur la même idée : toute partition de la population en *strates* fortement homogénes (variances internes aux strates faibles vis-à-vis de la variance globale) doit permettre d'accroître la précision.

Dans le plan stratifié (PSTRAT) on effectue des sondages, par exemple de type PSSR, indépendamment dans chaque strate et, *de facto*, se pose le problème du choix des tailles des « sous-échantillons ». L'estimateur naturel de t_y qui est aussi celui d'Horvitz-Thompson est obtenu, dans le cas de sous-échantillons de type PSSR, en recomposant les estimateurs des moyennes par strate, soit $\sum_{h=1}^{H} N_h \overline{Y}_h$ pour les H strates. Par l'indépendance mutuelle sa variance est immédiate. Un problème intéressant, mais purement théorique, est celui dit de « l'allocation optimale »,

à savoir de trouver les tailles n_h minimisant cette variance sous la contrainte $\sum_{h=1}^{H} n_h = n$ (pour le lecteur intéressé : les taux de sondage n_h/N_h sont alors proportionnels aux écarts-type des strates).

Généralement on stratifie à taux de sondage constants dans les strates ce qui garantit un gain de précision par rapport à un sondage PSSR. En effet la « fourchette » est multipliée par un facteur $(1-\eta^2)^{1/2}$ où η^2 est un coefficient bien connu des statisticiens, compris entre 0 et 1, qui mesure en quelque sorte le lien existant entre le critère de stratification et la variable d'intérêt (rapport de la variance intra-strates à la variance totale pour cette variable). Cette stratification, dite aussi « à la proportionnelle », est qualifiée par les praticiens comme produisant un échantillon *représentatif* vis-à-vis du critère de stratification. Notons que dans ce cas particulier les probabilités d'inclusion sont identiques pour tous les individus.

Ceci nous amène à la « *méthode des quotas* » dont la pratique est systématique dans les instituts. Elle consiste à sélectionner un échantillon qui soit un modèle réduit de la population sur certains critères de partitionnement dont on peut penser qu'ils ont un lien avec la thématique de l'enquête. Comme les effectifs des strates dans la population doivent être connus pour chaque critère d'une part et qu'il y a de fortes contraintes de mise en œuvre d'autre part, on se limite généralement à un critère géographique (région), à la classe d'âge, au sexe et, parfois, à la catégorie socio-professionnelle. La différence avec un plan PSTRAT tient au fait que, pour des raisons de faisabilité, la proportionnalité des effectifs de l'échantillon à ceux de la population ne concerne que les effectifs des critères « à la marge » et non des critères croisés entre eux.

Il n'est pas inutile, ici, pour les citoyens que nous sommes tous, d'ouvrir une parenthèse pratique. Dans la presse on déclare communément que les résultats d'un sondage proviennent d'un échantillon obtenu par la méthode des quotas et/ou d'un échantillon représentatif de la population selon la région, l'âge, etc. Ainsi il y a, chez les sondeurs, un véritable culte des quotas, la plupart d'entre eux étant convaincus qu'il suffit d'effectuer ce modèle réduit pour garantir de bonnes estimations. Or l'essentiel n'est pas là et, en fait, la mention consacrée dans la presse n'est d'aucun intérêt quant à juger de la qualité réelle de l'échantillon. Ce qui nous importe avant tout est de savoir dans quelle mesure un plan aléatoire, incluant ou non des contraintes de quotas, a-t-il été respecté.

L'approche théorique permet de déterminer le gain apporté par l'utilisation de quotas par rapport à un PSSR. Il est de même nature que celui présenté en stratification à ceci près que le coefficient η^2 repose ici sur la variance expliquée par un modèle à effets additifs des critères de quotas (pour les initiés : modèle d'analyse de variance sans interactions). Le constat empirique est bien décevant pour les critères usuels, le facteur de réduction de la fourchette $(1-\eta^2)^{1/2}$ ne passant que très rarement sous la valeur 0,90. Ce constat indique au passage que ces critères de quotas sont des déterminants extrêmement ténus du comportement ou des opinions des individus (il n'en reste pas moins que les quotas sont utiles, voire même nécessaires, pour limiter les biais importants découlant des obstacles de terrain).

La pratique des *redressements* relève des mêmes principes que les quotas mais elle intervient en aval du recueil de l'échantillon. Elle consiste, dans le but d'améliorer les estimations, à « caler » l'échantillon observé sur la population pour divers critères disponibles. Comme pour les quotas le calage se fait à la marge, critère par critère. Il s'effectue en attribuant des poids aux individus en perturbant le moins possible leurs poids de sondage. Mathématiquement il s'agit de déterminer les poids w_i tels que les contraintes de marges soient respectées pour les différents critères retenus et tels que :

$$\sum_{i=1}^{n} d\left(w_i, \frac{1}{\pi_i}\right)$$

soit minimal (où d est la distance choisie).

La résolution d'un tel problème nécessite une procédure itérative. En fait l'algorithme utilisé en pratique a été introduit tout à fait empiriquement : la recherche de poids correcteurs étant immédiate pour un seul critère, on calera successivement chacun des critères de façon cyclique jusqu'à quasi convergence vers les bonnes marges. Ce n'est que postérieurement que les travaux initiés par J.-C. Deville à l'INSEE ont donné un cadre théorique aux procédures de calage, notamment en introduisant la notion de distance mentionnée ci-dessus (l'algorithme usuel relève d'une distance implicite peu commune mais proche de la distance quadratique).

Les estimateurs par redressement n'ayant pas d'expression explicite leurs propriétés sont difficiles à établir. Ainsi, pour déterminer biais et variance, on doit se contenter d'approximations pour n grand. Brièvement disons que le

biais, pour des distances raisonnables, reste négligeable et que leur précision est très proche de celle découlant des estimateurs avec quotas (rapport égal à $1 + O(1/n^{1/2})$). Ainsi, comme pour ces derniers, le gain par rapport à un PSSR, pour une variable d'intérêt donnée, sera d'autant plus élevé que les critères retenus auront un lien fort avec elle. Ceci permet d'orienter le choix de ces critères sur la base des liens observés dans l'échantillon.

La modélisation des non-réponses

Il existe diverses sources d'erreur dans les sondages : erreur de mesure, biais de sélection, biais de couverture de la population et, souvent la plus importante, la *non-réponse*, à savoir qu'un individu dûment sélectionné par le plan de sondage n'a pu être observé pour une raison ou une autre. Les non-réponses, systématiquement présentes et même dans de fortes proportions, sont une cause potentielle de biais. Pour y remédier différents schémas du mécanisme de non-réponse ont été proposés pour lesquels s'appliquent des modèles appropriés. Par exemple, si le fait de répondre ou non à une enquête est indépendant de la variable d'intérêt conditionnellement à un ensemble de variables auxiliaires (par exemple âge, sexe, profession, statut matrimonial, etc.) le recours à un redressement sur ces variables est efficace. Il existe aussi des non-réponses partielles (seules certaines variables n'ont pu être observées ici ou là) débouchant sur des modèles *d'imputation*.

Le traitement des non-réponses totales ou partielles donne lieu actuellement à de nombreux développements théoriques.

Pour en savoir plus

Ardilly (P.), *Les techniques de sondage*, Technip, (1994).

Deville (J.-C.) et Sarndal (C.-E.), Calibration estimators in survey sampling, *Journal of the American Statistical Association*, vol. 87, p. 376-382, (1992).

Tillé (Y.), *Théorie des sondages*, Dunod, (2001).

Compression d'image

E. LE PENNEC*

La compression de données est une activité ancienne : l'utilisation d'abréviations en est une preuve. Les langues elles-mêmes utilisent des mots de longueurs variées, les plus fréquents étant les plus courts, afin de réduire la taille des phrases. Il est cependant une application plus visible que les autres : la compression des images numériques. C'est elle qui a permis la diffusion des images sur Internet ou encore la démocratisation des appareils photos numériques. Elle constitue également la base de la compression vidéo. L'objectif de cet article est de présenter à travers certains algorithmes classiques de compression des images numériques (GIF, JPEG, PNG) les fondations mathématiques de ceux-ci : théorie de l'information, modélisation statistique, approximation non-linéaire… La dernière partie sera consacrée à l'introduction récente de la géométrie pour améliorer la compression des images naturelles.

Images numériques et compressions

Une image numérique, telle qu'on peut la voir sur un écran d'ordinateur, est une mosaïque de pixels (picture elements) dont la couleur est choisie dans un ensemble fini : il s'agit d'un objet naturellement discret. Il s'identifie à une matrice à h colonnes et v lignes dont les éléments appartiennent à un ensemble fini E. Typiquement, pour une image en niveau de gris, E est constitué des entiers compris entre 0 et 255 et correspond à l'intensité lumineuse de chaque pixel. Ces 256 valeurs distinctes se codent avec 8 bits ($2^8 = 256$), d'où le nom d'image 8 bits. Pour les images couleurs, chaque pixel est caractérisé par 3 intensités lumineuses, celles des canaux rouge, vert et bleu, définissant des images $3 \times 8 = 24$ bits.

Le stockage en mémoire d'une image couleur requiert donc $h \times v \times 24$ bits. Ceci représente rapidement une grande quantité : une image de 4 mégapixels nécessite ainsi $4 \times 2^{20} \times 24$ bits soit 12 mégaoctets, ne permettant le stockage que d'une dizaine de photos sur une carte de 128 mégaoctets. Les modems téléphoniques fournissaient des contraintes encore plus grandes puisque une image couleur de 320×240 non compressée nécessitait pas loin de 4 minutes pour être transmise.

L'objectif de la compression est de réduire la quantité de mémoire nécessaire pour le stockage d'une image ou de manière équivalente de réduire le temps de transmission de celle-ci. Cette compression peut soit conserver l'image intacte, on parle alors de compression sans perte, soit autoriser une dégradation de l'image pour diminuer encore l'empreinte mémoire, on parle ici de compression avec perte. La première méthode est limitée à des facteurs de compressions (rapport entre la taille mémoire originale et la taille comprimée) de l'ordre de 3 tandis que la seconde permet des facteurs beaucoup plus grand au prix de cette dégradation de l'image. Nous allons maintenant voir comment ce procédé, illustré par la figure 1, est possible.

* *Laboratoire de Probabilités et Modèles Aléatoires UMR 7599, Université Paris 7. lepennec@math.jussieu.fr*

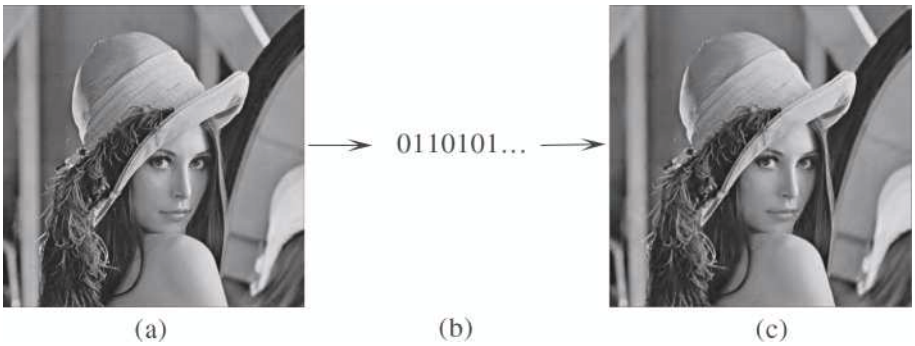


Figure 1 – *Compression : une image numérique (a) est transformée en un train de bits (b) qui permet de reconstruire l'image (c) (ici, dans le cas de la compression avec perte, déformée). Le nombre de bits utilisés est plus petit que celui,* a priori, *nécessaire pour décrire l'image.*

Compression sans perte

Le premier algorithme de compression d'image ayant connu un réel succès est un algorithme de compression sans perte et ceci n'est pas un hasard : il est issu de la compression de données arbitraires dans laquelle il est crucial de ne pas perdre d'information. L'algorithme GIF est ainsi l'héritier direct des algorithmes de compression de données génériques introduits à partir de la fin des années 40 par Shannon, le fondateur de la théorie de l'information.

L'entropie de Shannon

L'une des nombreuses questions que s'est posée Shannon à la fin des années 40 (et qu'il a bien souvent résolues) est la suivante : étant donnée une source X produisant des symboles x_i choisis dans un ensemble fini avec une probabilité respective $p(x_i)$, quel est le nombre minimal de bits nécessaire pour transmettre l'information que c'est le symbole x_{i_0} qui a été produit.

Shannon a pu montrer qu'il existe une quantité $H(X)$, qu'il a appelé entropie de la source, définie par

$$H(X) = -\sum_i p(x_i)\log_2 p(x_i)$$

telle que si un codage est possible en $\bar{n}$ bits en moyenne alors nécessairement

$$\bar{n} \geqslant H(X)\,.$$

L'entropie définit une borne inférieure sur le nombre moyen de bits nécessaire pour coder une source donnée. Ce résultat est complété par la construction explicite à partir des probabilités $p(x_i)$ d'un code de longueur moyenne $\bar{n}$ satisfaisant

$$\bar{n} \leqslant \lceil H(X) \rceil$$

où $\lceil H(X) \rceil$ est le plus petit entier plus grand où égal à $H(X)$. Il est donc possible de construire un code quasi optimal dès que l'on connaît la loi de probabilités et la longueur moyenne de ce code est comprise entre $H(X)$ et $\lceil H(X) \rceil$.

L'entropie mesure l'incertitude dans la production de la source X : l'entropie d'une source X produisant N symboles distincts est maximale, égale à $\log_2 N$, si tous les symboles sont équiprobables et minimale, égale à 0, si l'un des symboles est de probabilité 1.

Supposons maintenant que la source X produise des pixels de 24 bits, soit 2^{24} valeurs distinctes possibles et que l'on souhaite coder une image de 4 mégapixels, la source correspondante pour les images est X^n où $n = 4 \times 2^{20}$ est le nombre de pixels. On vérifie que le nombre d'images possibles est $N = (2^{24})^{4\times 2^{20}}$. L'entropie $H(X^n)$ de cette source est nécessairement inférieure à $\log_2 N = 4 \times 2^{20} \times 24$ et il est donc possible de coder ces images en moins de $4 \times 2^{20} \times 24$ bits en moyenne. Ceci suppose cependant de connaître la distribution de probabilités de la source qui est inconnue. Il va donc falloir procéder sans connaître cette distribution.

Codage de type dictionnaire

En 1977, Lempel et Ziv propose un algorithme de codage universel : leur algorithme permet automatiquement de coder une suite de n symboles issues d'une même source X avec une longueur moyenne qui tend vers la longueur optimale $\bar{H}(X) = \lim_{n\to+\infty} \frac{1}{n} H(X^n)$ sous une simple hypothèse de stationnarité de la source. On assure ainsi l'optimalité asymptotique de la longueur moyenne du code sans la connaissance de la loi de la source.

L'algorithme proposé par Lempel et Ziv (LZ77) comme ses nombreuses variantes (LZ78, LZW,…) est, de plus, simple à programmer et efficace numériquement. Le principe de ces méthodes est de parcourir la chaîne de symboles et de coder les nouvelles occurrences des sous-chaînes déjà observées par une simple référence à l'occurrence précédente. Le terme de codage de type dictionnaire provient de l'implémentation dans laquelle un dictionnaire de sous-chaîne est construit au fur et à mesure du parcours de la chaîne. Sans mentionner toute la littérature existante sur le sujet, il suffit de noter que c'est la famille utilisée dans les algorithmes de compression de données de type ZIP (ou ARJ) pour en mesurer le succès.

L'algorithme de compression d'image GIF est basé sur cette technique ou, plus précisément, sur la variante LZW proposée par Welsh en 1984. En 1987, Compuserve, le leader du moment des fournisseurs d'accès au réseau, introduit ce format afin d'accélérer la transmission des images sur celui-ci. Pour comprimer une image, il suffit de réordonner la matrice des pixels en une liste de valeurs et de comprimer cette liste à l'aide de l'algorithme LZW. Cette technique simple est suffisante pour obtenir un taux de compression de l'ordre de 2 et donc de réduire de moitié le temps de transmission.

Prédiction

Le taux de compression obtenu à l'aide d'un algorithme de compression universel n'est optimal qu'asymptotiquement et, bien qu'il soit impossible de franchir la barrière de $(\log_2 N)/\bar{H}(X)$, il est possible d'accélérer la convergence vers celle-ci.

Les méthodes prédictives proposent de transformer de manière réversible les chaînes de symboles en des chaînes plus simple pour l'algorithme de compression. Il s'agit d'aider celui-ci en introduisant un modèle *ad hoc* permettant de prédire chaque symbole apparaissant dans la chaîne en fonction des symboles précédents et de coder à l'aide de l'algorithme de codage universel l'erreur de prédiction au lieu du symbole lui-même.

Ceci suppose que les symboles possèdent une interprétation, ce qui est le cas pour les images. Il n'est donc pas étonnant que, en 1995, lorsque l'algorithme PNG, conçu comme une alternative non encombrée par des brevets de GIF, a été proposé, cette prédiction ait été ajoutée. Les modèles proposés utilisent des prédictions linéaires de l'intensité lumineuse en fonction de celle des voisins déjà connus. Les erreurs de prédictions sont alors codées par une autre variante de l'algorithme LZ77, l'algorithme deflate.

L'amélioration de performance est notable puisque l'on peut atteindre des taux de l'ordre de 3 avec l'algorithme PNG.

Codage statistique

Les techniques de codage universel ne font pas intervenir la distribution de probabilité de la source mais s'y adaptent asymptotiquement. A l'opposé, on a vu que si l'on connaît cette distribution, on peut construire un code quasi optimal. Deux algorithmes se disputent la prédominance pour ces codages dits statistiques : le plus ancien, l'algorithme de Huffman, est simple et efficace mais est moins performant que le plus récent, l'algorithme de compression arithmétique, dont la complexité est plus grande. Le choix entre ces deux algorithmes se fait suivant les contraintes de performance et de complexité.

La voie de la modélisation statistique est une voie intermédiaire. Elle remplace la distribution inconnue par un modèle connu et utilise cette nouvelle distribution pour coder les symboles à l'aide d'un algorithme de codage statistique. Si le modèle n'est pas trop éloigné de la réalité (au sens de la distance de Kullback-Leibler), il permet de comprimer les données efficacement sans avoir recours à un comportement asymptotique.

Les modèles utilisés varient en complexité : les modèles les plus simples font l'hypothèse que tous les éléments de la chaîne sont indépendants et identiquement distribués tandis que les modèles les plus complexes conditionnent le choix de la distribution de probabilités pour un nouveau symbole à tous ceux déjà codés dans le passé.

Les meilleurs résultats de compression sans perte sont obtenus avec des modèles contextuels de type chaîne de Markov où la loi utilisée dépend du voisinage et est apprise au fur et à mesure du parcours de l'image. Ces modèles permettent des taux de compression dépassant 4 au prix d'un algorithme complexe et lent.

Compression avec perte

Ce taux de compression d'un facteur 4 n'est pas toujours suffisant pour la transmission et le stockage des images numériques de grandes tailles. Pour l'améliorer, il va falloir perdre quelque chose : dégrader l'image. Ceci permet d'atteindre des taux arbitrairement grand au prix d'une dégradation toujours plus importante. L'objectif des algorithmes de compression avec perte est de minimiser cette dégradation pour un taux de compression donné.

Dégradation et Quantification

La clé de la compression avec perte est dans une modification non réversible de la source permettant d'obtenir une nouvelle source dont l'entropie est plus faible.

La modification la plus simple pour les images est, sans doute, le changement de résolution : en diminuant le nombre de pixels, l'entropie de la source est diminuée. Cette astuce est largement utilisée, les images transmises le sont, le plus souvent, à une résolution adaptée au récepteur. Cependant, si l'on souhaite conserver la résolution, ce principe doit être abandonné. Une alternative simple est alors de réduire le nombre de couleurs possibles pour chaque pixel.

Cette phase qui consiste à utiliser une partition de l'espace des couleurs et à remplacer chaque couleur par un représentant de la classe à laquelle elle appartient a pour nom la quantification. L'incarnation la plus élémentaire de cette modification est la quantification scalaire uniforme : toute valeur x de luminosité dans l'intervalle $[(n-.5)\Delta,(n+.5)\Delta[$ est remplacée par $n\Delta$ où Δ est un pas de quantification à choisir et codée par l'entier n. On note que ceci a déjà été utilisé de manière implicite puisque l'intensité lumineuse est une valeur continue et qu'elle a déjà été quantifiée par des valeurs entières entre 0 et 255. L'effet de cette quantification est assez rapidement trop perturbant pour qu'elle soit utilisée : elle engendre des *aplats* désagréables pour l'oeil. D'autres choix de partitions pour la luminosité sont possibles et la question de l'optimisation de cette partition se pose.

Cette recherche d'une bonne partition est encore plus cruciale lorsque la quantification des couleurs est vectorielle : la bonne partition se trouve parmi celles de l'ensemble des couleurs et non plus celles de la luminosité canaux par canaux. Cette opération dite de palettisation fait par exemple partie intégrante de l'algorithme GIF qui débute par une partition de l'ensemble des couleurs en 256 classes. On constate une amélioration par rapport à la quantification scalaire mais la encore la limite perceptuelle se rencontre assez vite.

Enfin, la quantification peut-être réalisée en travaillant sur des blocs de pixels plutôt que de travailler pixel par pixel. La recherche d'une partition optimale de l'ensemble des blocs possibles pose alors un problème de complexité mais améliore encore les résultats. Cette technique n'est pas très utilisée car en ajoutant un ingrédient la quantification scalaire suffit.

Transformation

La quantification scalaire devient en effet suffisante si elle est précédée d'un changement de base orthonormée. Les valeurs à quantifier ne sont alors plus les couleurs des pixels mais les coordonnées de l'image dans cette nouvelle base. Ceci crée de manière simple une partition appropriée non pas de l'ensemble des couleurs mais directement de l'ensemble des images pourvu que la base soit bien choisie. L'image est alors reconstruite à l'aide des coordonnées quantifiées.

Ainsi, pour $\{b_n\}_{n\in N}$ une base orthonormée de l'ensemble des images d'énergie finie (l^2) et I l'une de ces images, la relation entre I et ses coefficients c_n dans la base $\{b_n\}$ est donnée par la formule classique suivante

$$I = \sum_{n\in N} c_n b_n \text{ avec } c_n = \langle I, b_n \rangle .$$

Le codage par transformée est obtenue en quantifiant les coefficients par un quantificateur Q_n connu et en codant, sans perte cette fois, les coefficients quantifiés par des méthodes similaires à celles de la section précédente. L'image dégradée $\tilde{I}$ effectivement compressée est alors donnée par

$$\tilde{I} = \sum_{n\in N} Q_n(c_n) b_n .$$

L'application la plus connue de ce principe est sans doute l'algorithme JPEG proposé en 1994 par un comité d'experts (Joint Picture Expert Group). Ce standard possède un statut de quasi monopole pour la compression avec perte des images couleurs et on verra que même le nouveau standard JPEG 2000 a du mal à le détrôner. La base choisie

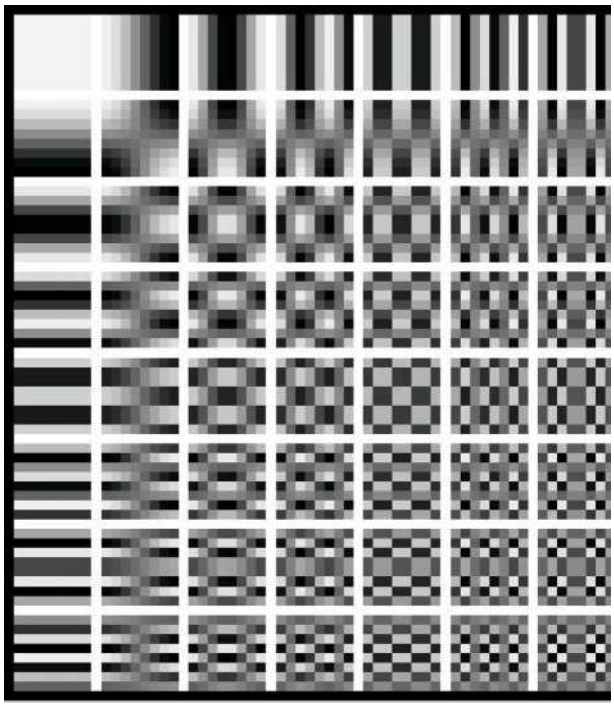


Figure 2 – *Base de la DCT-IV : chacun des 64 carrés correspond à un élément de la base utilisée dans l'algorithme JPEG. Le caractère oscillant des éléments (la fréquence) augmente lorsqu'on se déplace sur la droite ou vers le bas.*

par le comité est dérivée des bases de Fourier : l'image est découpée en blocs de taille 8×8 et chacun de ces sous-blocs est décomposé dans une base $\{b_{j_1,j_2}\}$ de cosinus locaux (DCT-IV) représentée dans la figure 2 :

$$b_{j_1,j_2}[k_1,k_2] = \frac{1}{4}\cos\left(\frac{\pi}{8}\left(j_1+\frac{1}{2}\right)\left(k_1+\frac{1}{2}\right)\right)\cos\left(\frac{\pi}{8}\left(j_2+\frac{1}{2}\right)\left(k_2+\frac{1}{2}\right)\right).$$

Ces coefficients sont alors quantifiés et codés à l'aide d'un codage statistique (Huffman le plus souvent) dans lequel les coefficients sont modélisés comme indépendants. La transformée a un effet décorrélant qui justifie cette approximation. Un modèle psycho-visuel est de plus utilisé pour quantifier les coefficients selon leur importance visuelle. Cet algorithme est très efficace pour une large gamme de taux de compression mais présente l'inconvénient de faire apparaître ces blocs 8×8 à fort taux de compression.

Le choix de la base est crucial pour les algorithmes par transformée comme on va le montrer maintenant avec une petite étude théorique.

Codage et approximation non linéaire dans des bases

Nous allons ici construire un codeur par transformée très simple mais qui permet de saisir les enjeux du choix de la base et sa relation avec la théorie mathématique de l'approximation : l'image I est *transformée* en ses coefficients dans la base $\{b_n\}$, ceux-ci sont quantifiées de manière uniforme à l'aide d'un pas Δ et l'on code alors les valeurs quantifiées par deux listes : une liste binaire donnant pour chaque coefficient s'il est nul ou non et une liste des valeurs quantifiées non nuls. Cette stratégie s'explique par le caractère particulier des coefficients quantifiés à 0.

En effet, en reprenant les notations précédentes, l'erreur introduite par la compression est donnée par la différence entre I et $\tilde{I}$ se mesure aisément en norme quadratique :

$$\|I-\tilde{I}\|^2 = \sum_{\substack{n\in N\\ |c_n|\leqslant\Delta/2}} c_n^2 + \sum_{\substack{n\in N\\ |c_n|>\Delta/2}} (c_n - Q(c_n))^2$$

et en introduisant M_Δ le nombre de coefficients c_n tel que $|c_n| \geqslant \Delta/2$

$$\|I-\tilde{I}\|^2 \leqslant \|I-I_{M_\Delta}\|^2 + M_\Delta\Delta^2/4$$

où I_{M_Δ} est obtenu à partir de I en conservant les M_Δ plus grands coefficients en valeurs absolues.

L'étude du terme $\|I-I_{M_\Delta}\|^2$ est le coeur de la théorie de l'approximation dans les bases orthonormées. Pour approcher une fonction (une image) avec M coefficients à choisir librement pour minimiser l'erreur quadratique de reconstruction, la bonne stratégie est de conserver les M plus grands coefficients en valeurs absolues. L'un des objets de la théorie de l'approximation est d'étudier les possibilités d'approximations de classes de fonctions dans une base donnée. Ceci s'exprime bien souvent par une relation entre l'erreur d'approximation $\|I-I_M\|^2$, M et Δ, le seuil associé, de la forme : $\|I-I_M\|^2 \leqslant CM^{-\gamma}$ et $M_\Delta \leqslant C'\Delta^{-2\gamma/(\gamma+1)}$, où α est lié à une forme de régularité propre à la classe. L'erreur de compression satisfait alors

$$\|I-\tilde{I}\|^2 \leqslant C'' M_\Delta^{-\gamma}$$

La taille du code nécessaire pour spécifier les coefficients quantifiés est également reliée à cette quantité M_Δ. La proportion de coefficients quantifiés non nuls est de M_Δ/N. L'entropie de la carte binaire de non nullité des coefficients est donc, si les coefficients sont considéréscomme indépendants, de $-M_\Delta \log_2(M_\Delta/N) - (N-M_\Delta)\log_2(1-M_\Delta/N)$. Enfin, chacun des $M_\Delta$coefficients non quantifiés à 0 requiert auplus $C - \log_2\Delta$ bits pour être spécifiés. Il en résulte, après un développement limité en M_Δ/N, que le nombre total R de bits nécessaire pour coder l'image satisfait

$$R \simeq M_\Delta(1 + \log_2(M_\Delta/N) + C - \log \Delta)$$
$$\simeq M_\Delta(C' + \log_2(M_D/N)).$$

La combinaison des estimations de $\|I - \tilde{I}\|^2$ et R donne alors

$$\|I - \tilde{I}\|^2 \leqslant C\, R^{-\gamma}\, \log_2^{\gamma+2}(R)$$

de sorte que l'efficacité de cet algorithme de codage est lié à une performance d'approximation non linéaire dans la théorie de l'approximation. Les performances de l'algorithme JPEG sont ainsi reliées à la capacité de la base de Fourier à approcher les fonctions régulières. La base de Fourier n'est cependant pas optimale pour les images et les bases d'ondelettes, introduites plus récemment, possèdent de meilleures propriétés. C'est donc tout naturellement qu'elles ont été utilisées dans le nouveau standard JPEG 2000.

Ondelettes et JPEG 2000

Les performances des ondelettes, introduites par S. Mallat en 1989, ont pour origine une adaptativité à la taille des structures des images inaccessible pour la DCT et ses blocs 8×8. Elles s'obtiennent par le procédé récursif de la figure 3. A chaque étape, l'image basse résolution est décomposée en une image de plus basse résolution vivant dans un espace de dimension 4 fois plus petite et une image de détails. Chacune de ces composantes est représentée à l'aide de fonctions de bases dont la taille double à chaque étape. Ainsi, la base finale comprend des fonctions de supports variés : de fonctions à large support pour les grandes tendances à des fonctions à support très petits pour des détails très précis en passant par les situations intermédiaires. Le tour de force de cette construction est de rester dans le cadre des bases orthonormées : la technique de codage par transformée s'applique immédiatement. Pour être un peu plus précis, les bases d'ondelettes sont construites à partir de deux fonctions monodimensionnelles : une fonction d'échelle ϕ, qui est d'intégrale 1, et une ondelette ψ de moyenne nulle. Trois ondelettes bidimensionnelles sont obtenues par simple produit tensoriel comprenant au moins une ondelette : $\phi(x_1)\psi(x_2)$, $\psi(x_1)\phi(x_2)$, $\psi(x_1)\psi(x_2)$. Pour un bon choix de ϕ et ψ, la famille obtenue en dilatant ces fonctions par des puissances de 2 et en les translatant de manière adaptée

$$\left\{ \frac{1}{2^j}\phi\left(\frac{x_1 - k_1 2^j}{2^j}\right)\psi\left(\frac{x_2 - k_2 2^j}{2^j}\right),\ \frac{1}{2^j}\psi\left(\frac{x_1 - k_1 2^j}{2^j}\right)\phi\left(\frac{x_2 - k_2 2^j}{2^j}\right),\right.$$
$$\left.\frac{1}{2^j}\psi\left(\frac{x_1 - k_1 2^j}{2^j}\right)\psi\left(\frac{x_2 - k_2 2^j}{2^j}\right)\right\}_{j\in\mathbb{Z},k_1\in\mathbb{Z},k_2\in\mathbb{Z}}$$

constitue une base orthonormale d'ondelettes.

Ainsi, l'algorithme JPEG 2000 proposé fin 2000 par le comité JPEG, mais terminé uniquement en 2002, débute par une décomposition de l'image dans une base d'ondelettes. Les coefficients ainsi obtenus sont alors quantifiés et codés par plan de bits à l'aide d'un codage arithmétique utilisant un modèle contextuel de dépendance entre les coefficients. Celui-ci permet une amélioration de la qualité par rapport à JPEG pour un même taux de compression mais surtout offre de nouvelles possibilités pratiques : transmission progressive, région d'intérêt. Bien qu'il s'agisse d'un standard libre de droit, son utilisation reste encore limitée. L'une des raisons est sans doute que pour des taux de compression de moins de 10 la différence avec l'algorithme JPEG n'est pas toujours visible. Pour des images géométriques qui constituent un modèle pour les images naturelles, il est de plus prouvé que ni les bases issues de Fourier ni même les ondelettes ne sont optimales. Elles sont incapables en effet d'exploiter la composante géométrique de celles-ci.

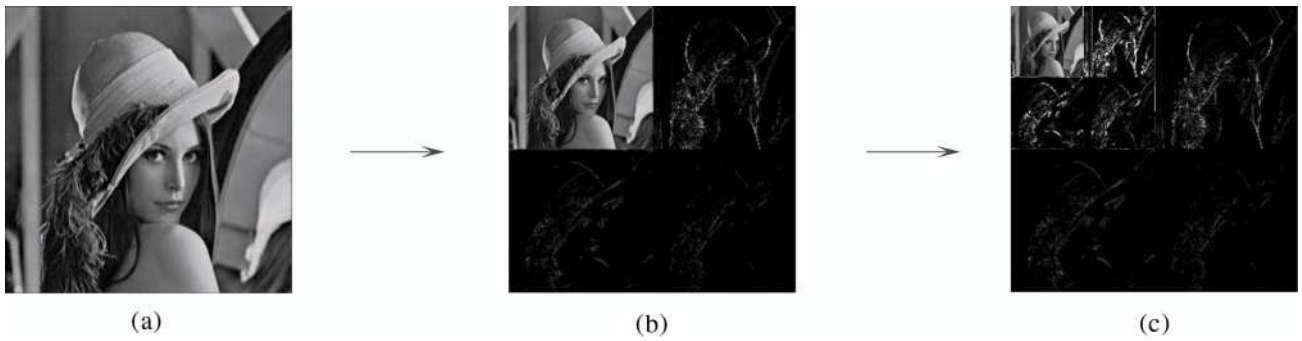


Figure 3 – *Transformation en ondelettes. Deux étapes de la transformation en ondelettes sont représentées ici. L'image (a) est transformée en une image basse résolution (en haut à gauche de l'image (b)) et trois sous images de détails. Chacune de ces images correspond à l'amplitude des coefficients dans une base d'ondelettes. L'image (c) de coefficients est obtenues en répétant ce procédé pour l'image basse résolution de (b). Ceci correspond à la structure multi-résolution des bases d'ondelettes.*

Et la géométrie ?

La géométrie est une des caractéristiques essentielles des images naturelles : elle constitue un élément de régularité qui n'est pas pris en compte par les bases classiques. L'exploitation de cette régularité géométrique est ainsi une direction prometteuse pour la compression d'image et plus généralement pour le traitement des images. L'objectif de cette section est de présenter un bref panorama des directions explorées actuellement. Elle débute cependant par un cas plus simple, celui des signaux monodimensionnels, qui permet de présenter les enjeux de l'exploitation de la régularité.

Signaux monodimensionnels et régularités

Les performances de l'algorithme de compression présenté dans la section 3.3 sont contrôlées par la vitesse de décroissance de l'approximation non linéaire du signal dans la base. L'erreur obtenue en ne conservant que les grands coefficients et en quantifiant les coefficients restants est en effet du même ordre de grandeur que celle obtenue sans la quantification. Cette décroissance dépend à la fois du signal considéré et de la base utilisée.

On étudie classiquement la décroissance de cette erreur pour une base donnée en fonction de l'appartenance à une classe de régularité. Pour les signaux monodimensionnels, la régularité est souvent mesurée par l'appartenance à la classe $\mathbf{C}^\alpha$ des fonctions α fois dérivables et dont la dérivée d'ordre α est continue (cette définition valable pour α entier s'étend au cas α positif). On démontre que si f est $\mathbf{C}^\alpha$ avec $\alpha \geqslant 1$ alors l'erreur d'approximation obtenue en conservant les M plus grands produits scalaires $\|f - f_M\|^2$ décroît comme $M^{-2\alpha}$ que se soit avec la base de Fourier ou avec une base d'ondelettes. Cette décroissance s'avère de plus optimale : il est impossible de trouver une base dans laquelle la décroissance est plus rapide.

Si l'on considère maintenant des signaux qui ne sont plus que $\mathbf{C}^\alpha$ par morceaux, les deux bases induisent des comportements différents. Dans la base de Fourier, la décroissance de l'erreur ne peut être bornée que par M^{-1} tandis que, grâce à la structure de multirésolution de la base d'ondelettes, la décroissance de l'erreur en $M^{-2\alpha}$ est optimale.

La construction d'une nouvelle base, celle des ondelettes, a permis d'exploiter la régularité de manière plus fine que ce que ne permet l'autre base, la base de Fourier. La situation pour les images est similaire.

Modèle géométrique des images

Si l'on considère des images uniformément régulières, $\mathbf{C}^\alpha$, on vérifie que les bases de Fourier ainsi que celles d'ondelettes permettent d'obtenir une décroissance optimale de l'erreur de $M^{-\alpha}$. Ce modèle est cependant trop grossier pour les images puisque celles-ci sont plutôt régulières par morceaux.

Un modèle géométrique d'image simple a été proposé par Tsybakov et Korostelev : une image géométrique est obtenue par discrétisation d'une fonction régulière $\mathbf{C}^\alpha$ en dehors de contours eux-mêmes réguliers $\mathbf{C}^\alpha$. Ce modèle simple néglige la partie texturée des images mais donne un cadre raisonnable pour comparer les différentes méthodes d'un point de vue théorique.

On montre que les bases d'ondelettes, bien que plus efficaces que la base de Fourier, sont incapables de capturer la régularité géométrique des contours : les ondelettes qui touchent les bords donnent de grands coefficients et leur nombre important limite fortement la décroissance de l'erreur qui se comporte comme M^{-1} alors que la décroissance optimale est en $M^{-\alpha}$. Une première approche plus géométrique des images est possible en abandonnant l'idée de base et en recherchant une triangulation adaptée. A l'aide de triangles allongés s'adaptant aux contours, le nombre de paramètres nécessaires est beaucoup plus petit que dans le cas des ondelettes.

La recherche d'une triangulation optimale et son codage efficace reste cependant une question ouverte. L'idéal serait d'exhiber une transformation présentant un comportement similaire en terme de nombre total de paramètres.

Curvelets, Wedgelets, ...

En 1999, Candès et Donoho ont construit une transformation en curvelets réalisant quasiment ce programme. Ses éléments de base sont des structures anisotropes à des échelles, des orientations et des positions variées. Ils forment une structure multirésolution similaire à celle des bases d'ondelettes à laquelle est été ajouté un filtrage orienté et permettent de capturer la régularité des fonctions de type $\mathbf{C}^2 - \mathbf{C}^2$. La décroissance de l'erreur d'approximation pour cette classe de fonction se comporte de manière quasi optimale en $\log M M^{-2}$.

Cependant la structure obtenue n'est pas celle d'une base mais d'un repère oblique (frame) : la famille est génératrice mais elle n'est pas libre. Ceci implique une certaine redondance dans la représentation et constitue un handicap pour un algorithme de compression. Cette construction est de plus continue et en obtenir une discrétisation efficace est un problème difficile.

Baraniuk *et al*. propose une direction différente pour l'utilisation de la géométrie. La transformée utilisée est une transformée en ondelettes classique et la géométrie n'intervient que pour définir un contexte pour le codage des coefficients. Les wedgelets sont des éléments de contours rectilignes placés sur une grille dyadique. Chacun d'entre eux fournit un contexte pour un grand nombre d'ondelettes dans son voisinage. Un algorithme d'optimisation de cette géométrie existe et l'algorithme de compression ainsi obtenu est asymptotiquement plus efficace que l'algorithme classique de compression en ondelette qui n'incorpore pas la géométrie. Il a les mêmes performances qu'un algorithme basé sur une transformée en curvelets.

Bandelettes

Introduites avec S. Mallat, elles s'adaptent aux images utilisées. Les bases de bandelettes sont des bases indicées par une géométrie : leurs vecteurs de bases sont allongés le long de courbes qui sont choisies pour épouser au mieux les contours. Plus précisément, l'image est segmentée en carrés dyadiques et chacun de ces carrés est muni d'une base adaptée. Ces bases locales sont construites à partir d'une base d'ondelettes anisotropes (obtenues par un simple produit tensoriel de deux bases monodimensionnelles) déformée pour suivre le contour principal dans chaque carré. La figure 4 illustre l'intérêt de cette représentation.

Le choix de la segmentation et de sa géométrie, et donc de la base utilisée, dépend alors de l'image à comprimer et du taux de compression. Le surcoût de la spécification de cette base *via* la géométrie est pris en compte dans un algorithme rapide de recherche de meilleure base qui permet d'obtenir automatiquement le meilleure compromis entre l'adaptation de la base à l'image (le bon suivi des contours) et son coût de codage.

La décroissance de l'erreur d'approximation en bandelettes est optimale pour les fonctions régulières en dehors de contours réguliers : $M^{-\alpha}$ pour les fonctions $C^\alpha - C^\alpha$. Ce résultat théorique s'accompagne de bonnes performances pour les images réelles.

Les méthodes géométriques de compression d'images constituent une direction de recherche très dynamique, aussi ce panorama est loin d'être complet. On pourrait mentionner la construction de transformations qui s'adaptent automatiquement à la géométrie sans le besoin de spécifier celle-ci ou encore des approches de géométrie discrètes,… Elles s'accordent cependant toutes sur le fait que la géométrie est la clé pour améliorer significativement les méthodes actuelles de compression.

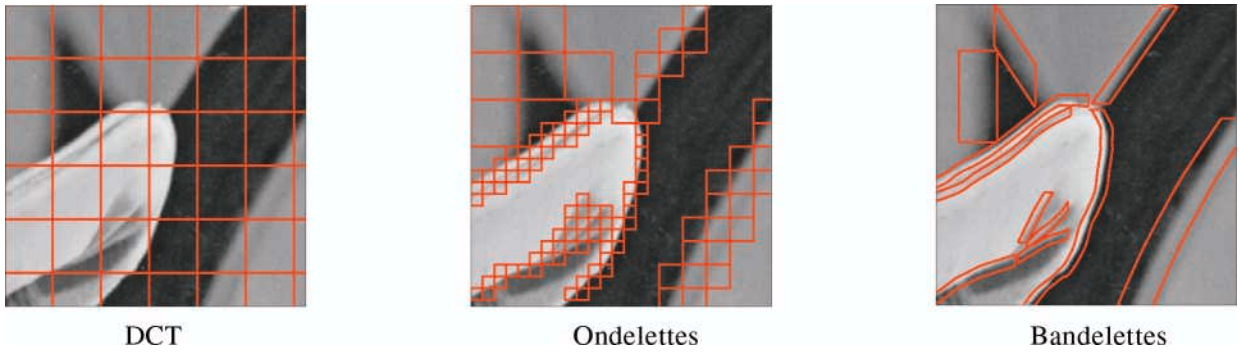


Figure 4 – *DCT, Ondelettes et Bandelettes. Cette figure illustre l'évolution des représentations des images. La DCT correspond à un découpage uniforme de l'image tandis que les ondelettes permettent d'adapter la taille des structures utilisées à celles présentes dans l'image. Enfin, les bandelettes s'adaptent à la composante géométrique des images.*

Pour en savoir plus

[1] Candès (E.), Donoho (D.), Curvelets : A Surprisingly Effective Nonadaptive Representation of Objects with Edges, in Schumaker L. L., Cohen A. and Rabut C., *Curves and Surfaces fittin*g, Vanderbilt University Press (1999).

[2] Do (M. N.), Vetterli (M.), Contourlets, in Stoeckler J. and Welland G. V. *Beyond Wavelets*, Academic Press, (2003).

[3] Le Pennec (E.), Mallat (S.), Sparse Geometrical Image Representation with Bandelets, *IEEE Transaction on Image Processing* (2005).

[4] Mallat (S.), *A wavelet tour of signal processing*, Academic Press, 2nd edition (1998).

[5] Nelson (M.), Gailly (J.-L.), *The data compression book (2nd ed.)*, MIS :Press, New York, NY, USA (1996).

[6] Shannon (C. E.), A mathematical theory of communication, *Bell System Technical Journal* (1948).

[7] Wakin (M.), Romberg (J.), Choi (H.), Baraniuk (R.), Rate-Distortion Optimized Image Compression using Wedgelets, in *IEEE Internat. Conf. on Image Processing* (2002 sep.).

[8] Welch (T.), A Technique for High-Performance Data Compression, *Computer* (1984).

[9] Ziv (J.), Lempel (A.), A Universal Algorithm for Sequential Data Compression, *IEEE Transactions on Information Theory* (1977).

Nombres transcendants et la diagonale de Cantor

Michel MENDÈS FRANCE*

En revisitant de vieilles idées de Cantor, on découvre de nouveaux résultats concernant la construction des nombres transcendants. En particulier on montre comment « engendrer » tous les nombres transcendants de l'intervalle unité à partir de l'ensemble de tous les nombres algébriques de l'intervalle unité (théorème 4).

L'oeil nu

Le but de cet article est double. D'abord présenter des mathématiques que j'espère simples mais belles, ensuite montrer qu'il est possible d'écrire des articles de recherche compréhensibles par des jeunes étudiants.

Le contenu de ce qu'on va lire est en effet extrait d'un récent papier que j'ai écrit conjointement avec S. Brlek, M. Robson et M. Rubey [1]. Il vient de paraître dans la revue « L'Enseignement Mathématique » qui s'adresse à des chercheurs professionnels. Le nom du journal est trompeur car il contient essentiellement des articles de recherche mathématique et qui – on peut le regretter – n'ont rien à voir avec l'enseignement !

Le savant Jean Rostand disait dans les années 60 qu'il y a encore matière à faire de la biologie à l'oeil nu. Je reprendrais volontiers cette philosophie à mon compte : faire des maths à l'oeil nu, c'est-à-dire sans nécessairement mettre en oeuvre des outils mathématiques difficiles et sophistiqués. Tout l'art est alors de découvrir des choses simples et cependant pertinentes. Nul besoin donc d'être compliqué pour être novateur et intéressant.

Nombres algébriques, nombres transcendants

Un nombre α est dit algébrique de degré $d \geqslant 1$, s'il existe un polynôme irréductible sur Q à coefficients entiers.

$$P(X) = a_0 X^d + a_1 X^{d-1} + \cdots + a_{d-1} X + a_d, \ a_0 \neq 0$$

tel que $P(\alpha) = 0$. Ainsi par exemple

$$-3, 5/2, \sqrt{10}, i, \sqrt[3]{2} + i\sqrt{7}$$

sont algébriques de degrés respectifs $1,1,2,2,6$. L'ensemble des nombres algébriques forme un corps pour les opérations usuelles.

Un nombre qui n'est pas algébrique est dit transcendant. En existe t-il ? Il a fallut attendre J. Liouville [4] qui au milieu du XIX$^{\text{ème}}$ siècle a pu en exhiber.

** UFR Math-Info – Université Bordeaux I, CNRS – laboratoire A2X, UMR 5465, 351, Cours de la Libération – 33405 – Talence cedex*

Théorème 1 [Liouville]. *Soit α un nombre irrationnel algébrique réel de degré $d \geqslant 2$. Il existe une constante $c(\alpha) > 0$ telle que pour tout nombre rationnel $p/q, q > 0$ on ait*

$$|\alpha - \frac{p}{q}| \geqslant \frac{c(\alpha)}{q^d}$$

En d'autres termes un nombre algébrique irrationnel ne se laisse pas approcher de trop près par un nombre rationnel !

Preuve. Soit α un zéro réel du polynôme irréductible à coefficients entiers $P(X) = a_0X^d + a_1X^{d-1} + \cdots + a_{d-1}X + a_d, \quad a_0 \neq 0$.

Soit p/q un nombre rationnel arbitraire de l'intervalle $(\alpha - 1, \alpha + 1)$. Le bon vieux théorème des accroissements finis montre que

$$0 - P\left(\frac{p}{q}\right) = P\left(\alpha\right) - P\left(\frac{p}{q}\right) = \left(\alpha - \frac{p}{q}\right)P'(\xi)$$

où ξ est compris entre α et p/q. Par suite

$$\begin{aligned} |P\left(\frac{p}{q}\right)| &\leqslant |\alpha - \frac{p}{q}| \max_{\alpha-1 \leqslant t \leqslant \alpha+1} |P'(t)| \\ &= |\alpha - \frac{p}{q}| M(\alpha). \end{aligned}$$

$M(\alpha)$ est une constante non nulle qui ne dépend que de α puisque α étant donné il lui correspond un polynôme P non constant. Par ailleurs

$$\left|P\left(\frac{p}{q}\right)\right| = \frac{|a_0p^d + a_1p^{d-1}q + \cdots + a_dq^d|}{q^d}$$

Le numérateur est un entier non nul donc supérieur ou égal à 1, d'où

$$\left|P\left(\frac{p}{q}\right)\right| \geqslant \frac{1}{q^d}.$$

Bref, pour tout rationnel p/q de l'intervalle $(\alpha - 1, \quad \alpha + 1)$

$$|\alpha - \frac{p}{q}| \geqslant \frac{1}{M(\alpha)q^d}.$$

Si maintenant p/q est un rationnel hors de l'intervalle $(\alpha - 1, \alpha + 1)$, on a trivialement

$$|\alpha - \frac{p}{q}| \geqslant 1 > \frac{1}{q^d}.$$

En posant

$$c(\alpha) = \min\left\{1, \frac{1}{M(\alpha)}\right\}$$

on voit donc que pour tout rationnel p/q

$$|\alpha - \frac{p}{q}| \geqslant c(\alpha) q^d.$$

C.Q.F.D

Vers le milieu du XX$^{\text{ème}}$ siècle, K.F. Roth a considérablement amélioré le résultat de Liouville en établissant que

$$\forall \varepsilon > 0 \quad \exists\, c_1 > 0 \quad \left|\alpha - \frac{p}{q}\right| > \frac{c_1}{q^{2+\varepsilon}}.$$

Nous ne nous servirons pas de ceci. Le théorème de Liouville suffit à établir l'existence de nombres transcendants comme on va le voir maintenant.

Corollaire 2. *Soit $b \geqslant 2$ un entier donné. Le nombre $\alpha = \sum_{n=1}^{\infty} \frac{1}{b^{n!}}$ est transcendant.*

Ainsi en base b (pensez $b = 10$ par exemple) le nombre α ne contient que des 0 et des 1, ces derniers n'apparaissant que de façon très lacunaire.

Preuve.

$$\alpha - \sum_{n=1}^{N} \frac{1}{b^{n!}} = \sum_{n=N+1}^{\infty} \frac{1}{b^{n!}} \leqslant \frac{1}{b^{(N+1)!}} \left[1 + \frac{1}{b} + \frac{1}{b''} + \cdots\right] = \frac{b}{b-1} \frac{1}{b^{(N+1)!}} \leqslant \frac{2}{b^{(N+1)!}}$$

Les nombres rationnels

$$\frac{p_N}{q_N} = \sum_{n=1}^{N} \frac{1}{b^{n!}} = \frac{p_N}{b^{N!}}$$

vérifient donc

$$0 < \left|\alpha - \frac{p_N}{q_N}\right| < \frac{2}{q_N^{N+1}}$$

ce qui exclut l'existence d'une constante $c(\alpha) > 0$ et d'un exposant d indépendant de p/q tels que

$$\left|\alpha - \frac{p}{q}\right| > \frac{c(\alpha)}{q^d}.$$

Le nombre α n'est donc pas algébrique. CQFD

Cantor et la diagonale

Quelques décennies après Liouville, G. Cantor [2] apporte une preuve très originale de l'existence de nombres transcendants. Voici son argument. Soit encore $b \geqslant 2$ un entier fixé qui servira de base dans l'écriture des nombres réels et soit $B = \{0, 1, \cdots, b-1\}$ l'ensemble des chiffres.

Sur la première ligne d'un tableau infini on écrit le développement en base b d'un nombre algébrique de l'intervalle unité $(0,1)$:

$$0,a_{11}\,a_{12}\,a_{13}\,\cdots\,a_{1n}\cdots$$

Sur la deuxième ligne on écrit le développement d'un autre nombre algébrique de l'intervalle unité :

$$0,a_{21}\,a_{22}\,a_{23}\,\cdots\,a_{2n}\cdots$$

et ainsi de suite pour les lignes $3,4,\cdots$ jusqu'à ce qu'on ait épuisé l'ensemble (dénombrable) des nombres algébriques de l'intervalle unité. On obtient ainsi le tableau

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & \cdots \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & \cdots \\ a_{41} & a_{42} & a_{43} & \cdots & a_{4n} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

où $a_{ij} \in B$. Si un nombre rationnel admet deux développements différents, chacun occupera une ligne distincte.

Ainsi par exemple $1/b$ s'écrit

$$0,\,1\,0\,0\cdots$$

et

$$0,\,0\;\;b-1\;\;b-1\cdots$$

Ces deux développements devront apparaître dans le tableau. Cantor raisonne alors ainsi.

Considérons la diagonale $a_{11}\,a_{22}\,a_{33}\cdots$ qu'on perturbe de la façon suivante. On choisit un chiffre $b_1 \neq a_{11}$, puis $b_2 \neq a_{22},\cdots,b_n \neq a_{nn},\cdots$ Le nombre $0,b_1b_2\cdots b_n\cdots$ est transcendant. En effet, s'il était algébrique, il se trouverait sur une ligne du tableau, disons au rang i. Alors

$$0,\,b_1\,b_2\,b_3\cdots \quad = 0,a_{i1}\,a_{i2}\,a_{i3}\cdots$$

et en particulier $b_i = a_{ii}$ ce qui contredit l'hypothèse $b_n \neq a_{nn}$ pour tout n.

En fait nous allons voir que sans perturbation, la diagonale $0,a_{11}\,a_{22}\,a_{33}\cdots$ est un nombre trascendant.

Théorème 3. *Le nombre diagonal* $0,\,a_{11}\,a_{22}\,a_{33}\cdots$ *est transcendant.*

Preuve. On commence par observer que si

$$x = 0,c_1\,c_2\,c_3\cdots$$

est algébrique, alors il existe x' algébrique

$$x' = 0,c'_1\,c'_2\,c'_3\cdots$$

tel que pour tout j, $c'_j \neq c_j$.

En effet, en base 2, il suffit de choisir $x' = 1 - x$ car alors $c'_j = 1 - c_j$. En base $b \geqslant 3$ remarquant que

$$\frac{1}{b-1} = 0,1\,1\,1\cdots$$

on vérifie que $x' = x + \dfrac{1}{b-1} \pmod 1$ convient.

Ceci étant établi, revenons au tableau de Cantor. Supposons par l'absurde que $0, a_{11}\,a_{22}\cdots$ soit algébrique. Ce développement apparaît donc sur l'une des lignes, disons la i^e. D'après ce qu'on vient de voir, il existe une ligne au rang i' où chacun des chiffres diffère de ceux de la ligne i. Ainsi, d'une part

$$a_{11}a_{22}a_{33}\cdots = a_{i1}\,a_{i2}\,a_{i3}\cdots$$

et en particulier $a_{i'i'} = a_{ii'}$ et d'autre part pour tout $j, a_{i'j} \neq a_{ij}$ ce qui pour $j = i'$ conduit à la contradiction $a_{i'i'} \neq a_{ii'}$. CQFD

On peut apporter une précision à ce dernier théorème.

Théorème 4. *La diagonale* $0, a_{11}\quad a_{22}\quad a_{33}\cdots$ *qui est donc transcendante contient une infinité de fois chacun des chiffres* $0,1,\cdots,b-1$.

Preuve. On s'inspire de l'argument de Cantor. Supposons par l'absurde que le chiffre $a_0 \in \{0,1,\cdots,b-1\} = B$, n'apparaisse qu'un nombre fini de fois dans la diagonale $a_{11}, a_{22}, \cdots$

Soit $\sigma : B \to B$ une application telle que $\sigma(a_0) \neq a_0$ et pour tout $a \neq a_0$; $\sigma(a) = a_0$. L'application est donc sans point fixe.

D'après notre hypothèse, la suite

$$\sigma(a_{11})\quad \sigma(a_{22})\quad \sigma(a_{33})\cdots$$

est constante à partir d'un certain rang : $\sigma(a_{nn}) = a_0$ pour tout grand n. Le nombre

$$0,\sigma(a_{11})\quad \sigma(a_{22})\quad \sigma(a_{33})\cdots$$

est donc rationnel. Il est donc égal à l'une des lignes du tableau, disons le i^e :

$$a_{i1}\,a_{i2}\,a_{i3}\cdots = \sigma(a_{11})\,\sigma(a_{22})\,\sigma(a_{33})\cdots$$

donc $a_{ii} = \sigma(a_{ii})$ ce qui est absurde. CQFD

En base $b = 2$ le théorème 3 est une conséquence triviale du théorème 2 puisqu'un nombre transcendant doit contenir une infinité de 0 et de 1.

Supposons maintenant qu'on permute les lignes d'un tableau de Cantor de toutes les façons possibles. Les diagonales des nouveaux tableaux sont alors toutes transcendantes et constituent un ensemble $T(b)$, sous-ensemble de l'ensemble T de tous les nombres transcendants de l'intervalle unité. Le théorème 3 et le corollaire de Liouville montrent que pour $b \geqslant 3$, $T(b) \neq T$. Que dire de $T(2)$?

Théorème 5. $T(2) = T$.

Ce résultat est essentiellement dû à R. Gray [3] dont les définitions diffèrent des nôtres. Pour lui, une diagonale est toujours perturbée. Mais en base 2 une diagonale perturbée et une diagonale non perturbée ne diffèrent pas essentiellement : substituer le chiffre 0 au chiffre 1 et réciproquement. Convenablement amendée, voici la preuve de Gray adaptée à notre théorème.

Preuve. Soit $t \in T, t = 0, t_1\, t_2\, t_3 \cdots (t_j = 0 \text{ ou } 1)$. On considère un tableau de Cantor dont je note les lignes par $\ell_1, \ell_2, \ell_3, \cdots$ et dont l'ensemble est l'ensemble des nombres algébriques de l'intervalle unité.

Par abus de notation, l_i représentera soit la suite des éléments de la ligne soit le nombre algébrique dont l_i est le développement binaire.

On va montrer qu'en permutant convenablement les lignes du tableau, on obtient un nouveau tableau dont la diagonale est t.

Soit k le plus petit entier tel que $a_{k1} = t_1$. On pose alors $\ell'_1 = \ell_k$; ℓ'_1 est la première ligne du nouveau tableau. On considère maintenant le plus petit $k' \neq k$ tel que $a_{k'2} = t_2$. On pose $\ell'_2 = \ell_{k'}, \cdots$ et ainsi de suite.

Au bout du compte on a un nouveau tableau dont les lignes sont $\ell'_1, \ell'_2, \ell'_3, \cdots$ et dont le diagonale est t.

Il reste à vérifier que $\ell'_1, \ell'_2, \ell'_3, \cdots$ est bien une permutaion de $\ell_1, \ell_2, \ell_3, \cdots$ en d'autres termes, il faut s'assurer que chaque ℓ_i a bien été utilisée. Supposons par l'absurde qu'il existe des ℓ_n non utilisées et soit ℓ_k celle de rang minimale. Les lignes $\ell_1, \cdots, \ell_{k-1}$ ont toutes été utilisées pour obtenir des ℓ'_j. Soit N le j maximal. Alors pour tout $n > N, a_{kn} \neq t_n$, soit en d'autres termes $a_{kn} = 1 - t_n$ (on est en base 2).

Mais alors $\ell_k + t$ ne contient que des 1 à partir du rang $N + 1$ dans son développement. Ce nombre est donc rationnel ce qui est absurde car ℓ_k est algébrique et t est transcendant. CQFD

Victor Kleptsym remarque qu'en s'inspirant du résultat initial de R. Gray on peut établir le théorème qu'on reproduit ci-dessous sans preuve.

Théorème 6. *Considérons l'ensemble de toutes les diagonales obtenues par permutation des lignes d'un tableau de Cantor en base $b \geqslant 2$, et toutes les perturbations possibles à la Cantor de chacune de ces diagonales. On obtient ainsi l'ensemble de tous les nombres transcendants de l'intervalle unité, et ceci quel que soit la base $b \geqslant 2$.*

Le lecteur attentif aura sans doute observé qu'on peut modifier les théorèmes 2, 3, 4, 5 en considérant des tableaux où les lignes sont les nombres algébriques de degré $\leqslant d$. Les diagonales sont alors soit transcendantes soit algébriques de degré $> d$. Bien entendu, d'autres extensions sont possibles !

Conclusion

Depuis la fin du XIX$^{\text{ème}}$ siècle, la théorie des nombres transcendants s'est énormément développée. Ch. Hermite et F. Lindemann ont respectivement établi la transcendance de e et π. Plus tard, on montrait celle de $\log\alpha$ (α algébrique $\neq 0$ ou 1), e^α(α algébrique $\neq 0$), α^β(α algébrique $\neq 0$ ou 1, β algébrique irrationnel).

Ces derniers résultats sont dus à A. O. Gelfond et T. Schneider dans les années 30. Remarquer que e^π est transcendant puisque $e^\pi = i^{-2i} \cdots$

La théorie est aujourd'hui très fleurissante grâce aux travaux tous très profonds de nombreux mathématiciens dont A. Baker, W.D. Brownawell, D. et G. Chudnovski, S. Lang, K. Mahler, D. Masser, K.F. Roth, W. Schmidt, M. Waldschmidt, et de bien d'autres. La théorie de l'approximation diophantienne (mesure de la distance entre nombres irrationnels et nombres rationnels) inaugurée par Liouville reste très à la mode. Elle trouve des applications inattendues en mécanique et en physique. La stabilité du système solaire en dépendrait !

Merci à T. Rivoal pour m'avoir signalé l'article de R. Gray.

Pour en savoir plus

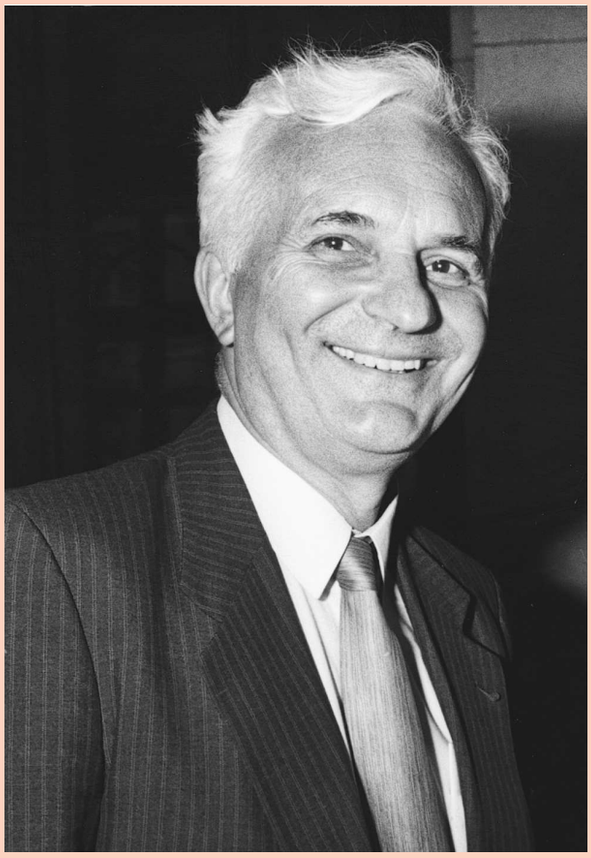
[1] Brlek (S.), Mendès France (M.), M. Robson (M.), M. Rubey (M.), *Cantorian tableaux and permanents* ; L'Enseignement Mathématique, 50, 287–304, (2004).

[2] Cantor (G.), *Über eine elementare Frage der Mannigfaltigkeitslehre*, Jahresbericht der Deutschen Math. Vereinigung 1, 75–78, (1891).

[3] Gray (R.), *Georg Cantor and Transcendental Numbers*, Amer. Math. Monthly, 101, 819-832, (1994).

[4] Liouville (J.), *Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques* ; CRAS, 18, 1844, 883–885, 910-911.

Jacques-Louis Lions



Jacques-Louis Lions (1928-2001) était un homme d'une stature exceptionnelle, chaleureux, profondément humain, pragmatique, doté d'une efficacité et d'un charisme hors du commun. C'était un très grand mathématicien qui a eu, en France et dans le monde, un impact profond sur le développement des mathématiques appliquées, domaine dans lequel il a créé une école qui travaille toujours sur les pistes qu'il a ouvertes.

Son adolescence ne fut pas commune : à quinze ans il s'engagea dans la résistance, dans les Forces Françaises de l'Intérieur (F.F.I.), fait qu'il rappelait souvent et qui a certainement contribué à modeler le caractère de l'homme.

Frais émoulu de l'Ecole Normale Supérieure, il partit avec son condisciple Bernard Malgrange travailler à Nancy sous la direction de Laurent Schwartz qui venait de mettre au point la théorie des distributions et d'obtenir pour cela la médaille Fields. Il y prépara sa thèse d'état et ce fut le début d'une suite impressionnante de travaux mathématiques.

Les recherches de Jacques-Louis Lions furent, au début, de nature théorique, axées sur l'utilisation systématique des distributions dans l'étude et la résolution des équations aux dérivées partielles (edp). La caractéristique fondamentale de ces travaux est l'étude approfondie des espaces fonctionnels adaptés à la résolution d'un problème donné. Pour les problèmes elliptiques, ce sont les espaces de Sobolev, introduits auparavant par Serguei Sobolev et l'école russe, et aussi de façon implicite par Jean Leray en France. Leur étude conduisit Jacques-Louis Lions aux problèmes de traces puis à la théorie de l'interpolation des espaces de Banach où il introduisit plusieurs concepts nouveaux (dont la méthode des traces, la méthode holomorphe et la méthode des moyennes).

Très vite, Jacques-Louis Lions comprit que la méthode théorique de résolution des edp qu'il employait pouvait être étendue à des méthodes d'approximation des solutions permettant de les calculer sur des ordinateurs. La philosophie est d'approcher les espaces fonctionnels adaptés par des sous-espaces de dimension finie et de conserver ainsi la structure du problème, et non plus d'approcher les opérateurs différentiels eux-mêmes (comme le fait la méthode des différences finies qui, en simplifiant, remplace une dérivée par un quotient différentiel). Les problèmes ainsi approchés en dimension finie se prêtent directement au traitement sur ordinateur *via* la résolution de systèmes linéaires de grande taille.

C'est là un exemple fascinant de progressions simultanées de la pensée mathématique théorique d'une part, et du développement et de l'utilisation pratique des ordinateurs d'autre part, rencontre féconde qui permit des avancées impressionnantes sur l'approximation numérique des solutions des edp. Cette approche changea les stratégies de résolution des edp et des méthodes de calcul, et les bouleversa complètement : recherche et étude des espaces fonctionnels adaptés ; recherche des propriétés qualitatives des solutions (existence, unicité, régularité, …) ; recherche (basée sur l'étude théorique) de méthodes d'approximation des espaces fonctionnels débouchant sur des calculs (et sur des méthodes de calcul) en dimension finie.

Une fois cette philosophie adoptée, de vastes chantiers s'ouvraient pour étudier :

– les problèmes d'edp linéaires, de manière systématique ;

– l'approximation numérique de ces problèmes ;

– les problèmes d'edp non linéaires ;

– les problèmes du calcul des variations et les problèmes de contrôle des edp, sous les aspects les plus variés (contrôle optimal, contrôlabilité, …) ;

– des questions comme les perturbations singulières, l'homogénéisation, etc.

Jacques-Louis Lions a laissé un grand nombre de travaux mathématiques de très haute qualité qui ont été publiés dans de grands journaux mathématiques (on pourra par exemple en trouver des présentations dans [1], [2] et [3]). L'essentiel en a été repris récemment dans ses Oeuvres choisies [4], mais figure déjà dans ses livres (une vingtaine) dont la plupart sont devenus de grands classiques connus sous des noms raccourcis : « Lions-Magenes » [5], « Lions non linéaire » [6], « Contrôle optimal » [7], « Duvaut-Lions » [8], « Bensousssan-Lions-Papanicolaou » [9], « Contrôlabilité » [10], … La plupart d'entre eux ont été traduits en anglais, en espagnol et en russe, et certains en chinois !

Outre ses propres travaux, Jacques-Louis Lions a ouvert des pistes particulièrement fructueuses qui ont été suivies ou poursuivies par le « groupe » (comme il aimait l'appeler) de ses très nombreux élèves et collaborateurs. Ses cours lumineux et ouverts, en particulier ses cours de Dea, puis ses cours au Collège de France, ses conférences qui donnaient aux auditeurs le sentiment d'être intelligents, étaient pour lui une occasion d'ouvrir et de développer ces nouvelles voies. Il débutait souvent ses exposés par la présentation d'un exemple modèle qui contenait les difficultés principales et lui permettait d'exposer la méthode de résolution, méthode qu'il cherchait toujours robuste et susceptible de nombre de variations à adapter selon les difficultés propres de chaque cas ; il entraînait ainsi continuellement ses auditeurs sur de nouvelles pistes et vers de nouveaux problèmes. On peut citer ainsi les très nombreux et importants travaux numériques entrepris dans son sillage sans que lui-même ait jamais travaillé directement sur un ordinateur, ou encore par exemple les travaux en mécanique des structures et en mécanique des fluides, sur les problèmes de la physique mathématique, sur la neutronique, sur l'homogénéisation ou la controlabilité exacte ou approchée entrepris à la suite de ses cours et de ses livres.

Jacques-Louis Lions a su reconnaître les formidables nouvelles possibilités qu'offrait le développement des ordinateurs et il s'est aperçu qu'il avait dans les mains (et dans la tête !) un outil mathématique qui lui permettait de tirer partie de l'outil technologique. Enfin, outre les qualités d'un grand scientifique, il avait les qualités d'un grand homme et une personnalité rayonnante et attachante qui a fait l'admiration continuelle de tous ceux qui ont travaillé avec lui ou qui l'ont côtoyé.

François MURAT, murat@ann.jussieu.fr
Laboratoire Jacques-Louis Lions,
Université Pierre et Marie Curie (Paris VI)

Jean-Pierre PUEL, jppuel@cmapx.polytechnique.fr,
Laboratoire de mathématiques appliquées,
Université de Versailles-St Quentin en Yvelines

Pour en savoir plus

[1] CIARLET (P.G.), *Jacques-Louis Lions (1928-2001)*, Matapli (bulletin de la SMAI), 66, pp. 5-16 (octobre 2001).

[2] TEMAM (R.), *Jacques-Louis Lions*, Notices of the AMS, 48 (11), pp. 1315-1320 (December 2001).

[3] MAGENES (E.), Ricordo di Jacques Louis Lions, Notiziario UMI (juin 2001).

[4] BENSOUSSAN (A.), CIARLET (P.G.), GLOWINSKI (R.), TEMAM (R.), éditeurs, MURAT (F.), PUEL (J.-P.), coordinateurs, *Œuvres choisies de Jacques-Louis Lions. Volume I : Equations aux dérivées partielles, Interpolation* (740 pages), *Volume II : Contrôle, Homogénéisation* (874 pages), *Volume III : Analyse numérique, Calcul scientifique, Applications* (828 pages), EDP Sciences, Paris (2003).

[5] LIONS (J.-L.), MAGENES (E.), *Problèmes aux limites non homogènes et applications, volumes 1, 2 et 3*, Dunod, Paris (1968).

[6] LIONS (J.-L.), *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod et Gauthier-Villars, Paris (1969).

[7] LIONS (J.-L.), *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod et Gauthier-Villars, Paris (1968).

[8] DUVAUT (G.), LIONS (J.-L.), *Les inéquations en mécanique et en physique*, Dunod, Paris (1972).

[9] BENSOUSSAN (A.), LIONS (J.-L.), PAPANICOLAOU (G.), *Asymptotic analysis for periodic structures*, North-Holland, Amsterdam (1978).

[10] LIONS (J.-L.), *Contrôlabilité exacte, perturbations et stabilité. Tome 1: contrôlabilité exacte*, Collection RMA, Masson (1988).

Henri Poincaré

Les mathématiciens connaissent surtout de Poincaré ses travaux en topologie, en mécanique céleste, en théorie des équations différentielles, en physique mathématique … Les multiples théorèmes ou formules de Poincaré suffisent à rappeler l'importance des contributions de Poincaré dans presque tous les domaines des mathématiques. Son modèle de la géométrie hyperbolique lui a servi à donner une compréhension géométrique de sa théorie des fonctions fuchsiennes qu'il développe dans les années 1880. La reprise qu'il fait de ce modèle dans le cadre de la discussion du statut des axiomes de la géométrie permet d'illustrer un aspect un peu moins connu de son œuvre.

En effet, Poincaré fut et reste un immense philosophe des sciences. Certes, Poincaré n'est pas le seul scientifique à participer aux débats épistémologiques à la fin du 19e siècle ou au début du 20e siècle. Mais il est un des rares dont les théories épistémologiques continue d'influencer de nombreux courants actuels de philosophie.

Les premières interventions philosophiques de Poincaré concernent la question de la géométrie et de l'espace. A la fin du 19e siècle, avec l'irruption des nouvelles géométries, le problème des liens entre la géométrie et l'espace était particulièrement crucial ; jusqu'alors, la philosophie kantienne répondait de manière assez satisfaisante à la question de l'espace et de la géométrie : l'espace était une intuition *a priori* ce qui justifiait que les axiomes de la géométrie euclidienne aient un caractère d'évidence immédiate. L'apparition de nouvelles géométries dont on dut reconnaître qu'elles avaient la même consistance que la géométrie euclidienne donna des arguments à ceux qui défendaient le caractère empirique des axiomes de la géométrie. Poincaré proposa une solution originale en refusant les points de vue kantien et empiriste en défendant la thèse que l'expérience jouait un rôle dans la genèse de nos conceptions géométriques sans pour autant réduire les jugements géométriques à des vérités empiriques. Pour Poincaré, les axiomes de la géométrie sont des *conventions* au sens où la décision d'utiliser une géométrie plutôt qu'une autre pour représenter les phénomènes physiques ou rapporter notre perception spatiale résulte d'un choix. Pour autant, l'expérience joue un rôle fondamental de guide dans le choix des conventions les plus commodes. Si aucune géométrie n'est imposée par notre esprit comme condition nécessaire de nos expériences, il n'y a pas, non plus, de géométrie imposée par l'expérience ; par contre, parmi toutes les conventions possibles, autrement dit, parmi toutes les géométries possibles, l'expérience nous guide dans le choix d'un cadre commode pour rendre compte de celle-ci. Poincaré conclut que la géométrie euclidienne « est et restera la plus commode » parce que celle-ci est la plus simple d'un point de vue mathématique et « parce qu'elle s'accorde assez bien avec les propriétés des solides naturels, ces corps dont se rapprochent nos membres et notre œil et avec lesquels nous faisons nos instruments de mesure ».

L'essentiel des contributions philosophiques de Poincaré est réuni en cinq volumes : *La science et l'hypothèse* (1902), *La valeur de la science* (1905), *Science et méthode* (1908), *Dernières Pensées* (1910) et *L'opportunisme scientifique* (2002).

Adresse du site des Archives Henri Poincaré (UMR 7117 du CNRS) sur Poincaré et notamment sa bibliographie et sa correspondance :
http://www.univ-nancy2.fr/poincare/

Philippe NABONNAND
Archives Henri Poincaré,
Facultés des Lettres et Sciences Humaines,
23 Bd Albert 1er, BP 3397, F-54015 Nancy Cedex

Surfaces à courbure moyenne constante

Frank PACARD*

Les surfaces à courbure moyenne constante apparaissent de manière naturelle dans la modélisation des interfaces entre fluides de densités différentes ou encore dans l'étude du problème isopérimétrique. Ces 20 dernières années, l'introduction de techniques d'analyse a permis de faire des progrès considérables dans la compréhension de ces objets géométriques.

Problème isopérimétrique

Didon, fondatrice de Carthages, aborda l'Afrique où le roi Jarbas lui accorda la portion de terrain que pourrait contenir la peau d'un bœuf. Didon fit découper cette peau en une bande étroite et s'en servit pour délimiter le bord d'un territoire semi-circulaire centré en un point de la côte, elle obtint ainsi un terrain assez vaste pour y construire une citadelle qui fut ensuite l'acropole de Carthages : Didon avait trouvé la solution du « problème isopérimétrique dans un demi plan ».

Soit (M,g) est une variété Riemannienne compacte de dimension $m+1$, $m \geqslant 1$. Le *problème isopérimétrique* dans (M,g) s'énonce de la manière suivante : étant donnée une constante $0 < \nu < \mathrm{Vol}_{m+1}(M)$, on cherche à déterminer le ou les domaines $\Omega \subset M$ dont la mesure m-dimensionnelle du bord $\mathrm{Vol}_m(\partial\Omega)$ est minimale parmi tous les domaines dont la mesure $(m+1)$-dimensionnelle $\mathrm{Vol}_{m+1}(\Omega)$ est égale à ν.

La *théorie de la mesure géométrique* permet d'apporter une réponse à ce problème et l'on sait qu'il existe (au moins) un domaine $\Omega \subset M$ dont la mesure m-dimensionnelle du bord est minimale parmi tous les domaines dont la mesure $(m+1)$-dimensionnelle est égale à ν. De plus, en dehors d'un ensemble de dimension de Hausdorff $m-7$, le bord de Ω est une hypersurface plongée dont la courbure moyenne est constante. Dans le cas où la variété M est une variété à bord et où $\partial M \cap \partial\Omega \neq \emptyset$, le bord de Ω rencontre ∂M de manière orthogonale. Si ce résultat assure l'existence d'un domaine solution du problème isopérimétrique, la détermination du domaine lui même reste un problème extrêmement compliqué (même dans un cadre très simple comme par exemple le cas où (M,g) est un tore plat de dimension 3). La caractérisation des solutions du problème isopérimétrique reste un domaine de recherche particulièrement actif dans lequel de nombreuses questions restent sans réponse [R-05].

La solution du problème isopérimétrique permet de distinguer une catégorie particulière d'hypersurfaces, celles dont la courbure moyenne est constante.

* *Université Paris 12, UMR-CNRS 8050,*
61, Avenue du Général de Gaulle, 94010 Créteil Cedex.
pacard@univ-paris12.fr

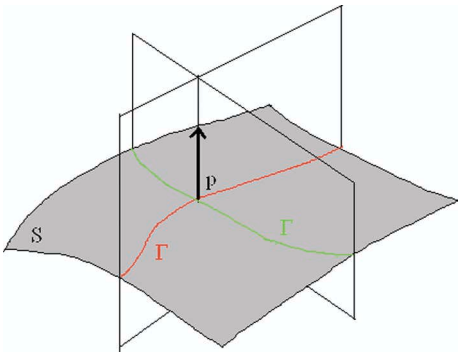


Figure 1 – *Courbure moyenne d'une surface de* $\mathbb{R}^3$.
On considère deux plans orthogonaux qui passent par le point $p \in S$ *et contiennent le vecteur normal* $\mathbf{n}(p)$*, ils coupent la surface* S *en deux courbes* Γ *et* Γ' *dont on calcule* $\mathbf{h}$ *et* $\mathbf{h}'$ *les vecteurs courbures respectifs au point* p*. La courbure moyenne de* S *au point* p *est alors donnée par la formule* $H = (\mathbf{h} + \mathbf{h}') \cdot \mathbf{n}$.

Encadré 1

Courbure moyenne d'une hypersurface

Soit S *une hypersurface compacte orientable, plongée dans une variété orientable* (M,g)*, on note* $\mathbf{n} = \mathbf{n}_S$ *le vecteur normal à* S *compatible avec l'orientation de* S*. Etant donnée* w*, une fonction régulière (suffisamment petite) et définie sur* S*, on peut définir l'hypersurface* S_w *paramétrée par*

$$p \in S \longrightarrow \mathrm{Exp}_p(w(p)\,\mathbf{n}(p)) \in S_w$$

où Exp *désigne l'application exponentielle dans* (M,g)*. Par exemple, dans le cas où* (M,g) *est l'espace Euclidien l'hypersurface* S_w *est simplement paramétrée par*

$$p \in S \longrightarrow p + w(p)\,\mathbf{n}(p)$$

On note alors $\mathcal{A}(w) := \mathrm{Vol}_m(S_w)$ *la mesure* m*-dimensionnelle de l'hypersurface* S_w*. La différentielle de* $\mathcal{A}$*, calculée en* $w = 0$*, est une forme linéaire qui peut s'écrire sous la forme*

$$D\mathcal{A}_{|w=0}(v) = -\int_S H(S)\, v\, dvol_S$$

où $dvol_S$ *désigne la forme volume sur* S*. La fonction* $H(S)$ *qui apparaît dans cette formule est la courbure moyenne de l'hypersurface* S*. On peut définir de manière équivalente* $H(S)$ *comme étant la somme des courbures principales de* S*, i.e. les valeurs propres de l'endomorphisme* $A_S : TS \longrightarrow TS$ *déterminé par la formule*

$$g(A_S X, Y) = g(\nabla_X Y, \mathbf{n}), \qquad \forall\, X, Y \in TS$$

où ∇ *est la dérivée covariante dans* (M,g).

Le cas de l'espace euclidien

A.D. Alexandrov a démontré que les sphères sont les seules hypersurfaces à courbure moyenne constante compactes, plongées dans l'espace euclidien $\mathbb{R}^{m+1}$. La démonstration de ce résultat repose sur un principe de réflection par rapport à des hyperplans. Ce « principe de réflection d'Alexandrov » a par la suite connu de nombreuses généralisations notamment dans le domaine des équations aux dérivées partielles non linéaires grâce aux travaux de J. Serrin, B. Gidas, W.M. Ni et L. Nirenberg.

Pendant longtemps, on a pensé que l'on pouvait affaiblir les hypothèses du résultat d'Alexandrov en supprimant la condition de plongement. En fait il n'en est rien et, en 1984, H. Wente a démontré l'existence de tores (immergés

dans $\mathbb{R}^3$) dont la courbure moyenne est constante (voir Figures 2 et 3). Ce résultat a ensuite donné lieu à de nombreux travaux qui ont mis en évidence le lien entre les tores à courbure moyenne constante de $\mathbb{R}^3$ et les systèmes intégrables. L'existence de surfaces de genre $g \geqslant 2$, immergées, à courbure moyenne constante, est maintenant établie grâce aux travaux de N. Kapouleas [K-05], M. Jleli et F. Pacard, mais les résultats ne sont encore que parcellaires.

On peut aussi s'intéresser aux hypersurfaces à courbure moyenne constante qui sont complètes, non compactes. Par exemple, si S^n_ρ désigne la sphère centrée en 0 et de rayon ρ dans $\mathbb{R}^{n+1}$, les cylindres droits $S^{m-k}_\rho \times \mathbb{R}^k$ sont des hypersurfaces complètes dont la courbure moyenne est constante $H = \dfrac{m-k}{\rho}$. Outre les cylindres droits, il existe dans $\mathbb{R}^{m+1}$, une famille à un paramètre d'hypersurfaces de révolution dont la courbure moyenne est constante si $m \geqslant 2$. En dimension $m = 2$, ces surfaces ont été découvertes au XIX$^{\text{ième}}$ siècle par Delaunay et elles ont pour génératrices des roulettes de coniques (voir Figures 4,…, 7) .

Les surfaces de Delaunay sont à l'origine du développement, dans les années 1990, de nombreux travaux portant sur $\mathcal{M}_{g,k}$, l'ensemble des surfaces de genre g, complètes, non compactes, à courbure moyenne constante, qui ont k bouts asymptotes à des onduloïdes de Delaunay [KMP-96] (voir Figures 8 et 9). Les principaux résultats montrent d'une part que $\mathcal{M}_{g,k}$ a une structure de variété dont la dimension (formelle) est égale à $3\,k$ (donc ne dépend pas du genre g) et d'autre part que $\mathcal{M}_{0,k}$ n'est pas vide dès que $k \geqslant 2$. Enfin, signalons le résultat de K. Grosse-Brauckman, R. Kusner et J. Sullivan [KGBS-03] qui permet de classifier les éléments de $\mathcal{M}_{0,3}$.

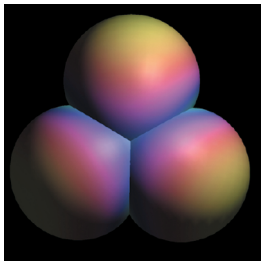


Figure 2 – *Tore de Wente.*

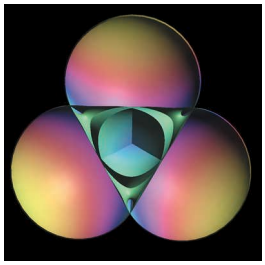


Figure 3 – *Vue en coupe d'un tore de Wente.*

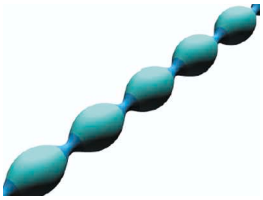


Figure 4 – *Onduloïde : Surface de Delaunay dont la génératrice est une roulette d'ellipse.*

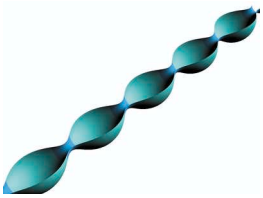


Figure 5 – *Vue en coupe d'un onduloïde.*

Figure 6 – *Nodoïde : Surface de Delaunay dont la génératrice est une roulette d'hyperbole.*

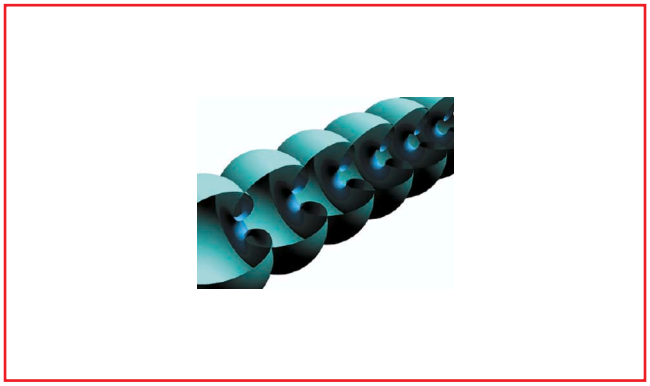


Figure 7 – *Vue en coupe d'un nodoïde.*

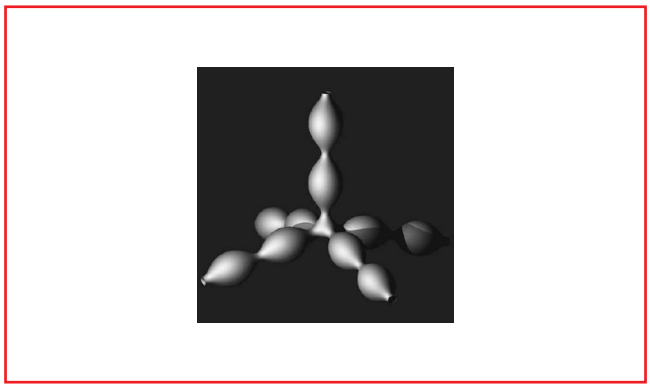


Figure 8 – *Surface à 5 bouts appartenant à* $\mathcal{M}_{0,5}$.

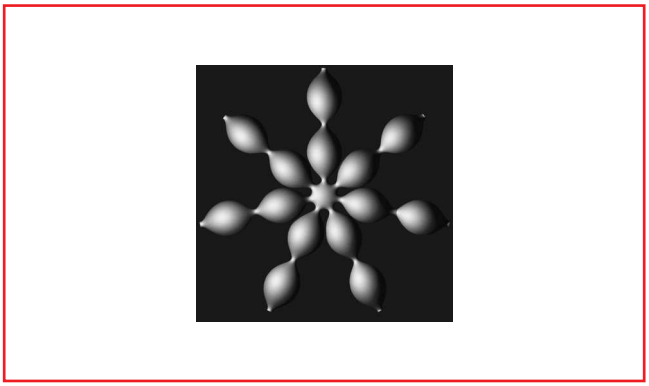


Figure 9 – *Surface à 7 bouts appartenant à* $\mathcal{M}_{0,7}$.

Le cas des variétés Riemanniennes

Définissons $\mathcal{M}(\Sigma, M, g)$ comme étant l'ensemble des hypersurfaces Σ qui sont plongées dans une variété Riemannienne compacte (M, g) et dont la courbure moyenne est constante. Précisons que la topologie des éléments de cet ensemble est fixée par celle de Σ, mais que la valeur de la courbure moyenne elle est une constante qui n'est pas fixée. Cet ensemble s'avère avoir une structure très riche et, pour un choix générique de la métrique g définie sur M, l'ensemble $\mathcal{M}(\Sigma, M, g)$ est une réunion de variétés régulières de dimension 1. Les résultats ci-dessous donnent une description (partielle) de certaines composantes non compactes de $\mathcal{M}(\Sigma, M, g)$.

Supposons que K est un point de M ou bien une sous-variété K de dimension $k \leqslant m-1$ plongée dans M. Définissons le tube géodésique de rayon $\rho > 0$ autour de K par

$$S_\rho(K) := \{p \in M \quad : \quad \mathrm{dist}(p, K) = \rho\}$$

On vérifie que, quand ρ tend vers 0, la courbure moyenne de $S_\rho(K)$ est presque constante au sens où

$$H(S_\rho(K)) = \frac{m-k}{\rho} + \mathcal{O}(1).$$

Il semble alors raisonnable de perturber $S_\rho(K)$ en une hypersurface à courbure moyenne constante, du moins lorsque ρ est assez petit. Il s'avère que des conditions supplémentaires portant sur K sont nécessaires pour pouvoir mettre en œuvre cette stratégie.

Dans le cas où K est un point $p \in M$, R. Ye [Y-91] a démontré le :

Théorème 1 [R. Ye]. *Soit $p \in M$ un point critique non dégénéré de la courbure scalaire R sur (M,g). Alors, il existe $\rho_0 > 0$ et une famille à un paramètre de sphères topologiques $\Sigma(\rho)$, pour $\rho \in (0,\rho_0)$, qui sont obtenues en perturbant $S_\rho(p)$ et dont la courbure moyenne est constante $H(\Sigma(\rho)) = \frac{m}{\rho}$. De plus, ces hypersurfaces constituent un feuilletage d'un voisinage de p par des hypersurfaces à courbure moyenne constante.*

Les solutions du problème isopérimétrique pour des contraintes de volume petites (*i.e.* $\nu \sim 0$) sont proches de sphères géodésiques. Lorsque la courbure scalaire R est une fonction de Morse, il est conjecturé que ces solutions appartiennent à la branche d'hypersurfaces obtenue par R. Ye qui se concentre autour du maximum de la courbure scalaire sur (M,g).

Dans le cas où K est une sous-variété de dimension $k = 1,\ldots,m-1$, la situation est radicalement différente [MMP-05] et nous avons alors le :

Théorème 2 [F. Mahmoudi, R. Mazzeo, F. Pacard]. *Soit K une sous-variété minimale non dégénérée, il existe $I \subset (0,1)$ tel que $\forall \rho \in I$, $S_\rho(K)$ peut être perturbé en une hypersurface $\Sigma(\rho)$ dont la courbure moyenne est constante égale à $H(\Sigma(\rho)) = \frac{m-k}{\rho}$. De plus, pour tout $t \geqslant 2$, il existe $c_t > 0$ tel que $|I \cap (0,r) - r| \leqslant c_t\, r^t$.*

Ce résultat met en évidence le lien entre sous variétés minimales de (M,g) et les branches non compactes de $\mathcal{M}(SNK,M,g)$, où SNK désigne le fibré en sphères associé au fibré normal à la sous-variété K dans la variété (M,g). Cette fois-ci, et contrairement à ce qui se passe dans le cas où K est un point, le résultat ne semble pas être valable pour toutes les valeurs de ρ. Ceci est dû à un phénomène de résonance qui est inhérent à la construction.

Il est intéressant de comprendre dans quelle mesure les conditions suffisantes d'existence énoncées dans les deux théorèmes ci-dessus sont aussi nécessaires. En d'autres termes : est-il possible de caractériser les sous ensembles sur lesquels des familles d'hypersurfaces à courbure moyenne constante se concentrent lorsque leur courbure moyenne tend vers l'infini ? Dans cette direction, mentionnons le résultat :

Théorème 3 [H. Rosenberg]. *Il existe $H_0 > 0$ et $c > 0$ (qui ne dépendent que de la géométrie de (M,g)) telles que, si S est une hypersurface plongée dont la courbure moyenne constante est (en valeur absolue) plus grande que H_0 alors S sépare M en deux composantes connexes. De plus, la distance entre un point p appartenant à la composante de $M - S$ vers laquelle le vecteur courbure moyenne pointe et l'hypersurface S est majorée par c/H.*

Encadré 2

Courbure scalaire

La courbure scalaire apparaît par exemple dans le développement limité, quand ρ tend vers 0, de la mesure m-dimensionnelle de la sphère géodésique $S_\rho(p)$ de centre p et de rayon ρ

$$\mathrm{Vol}_m(S_\rho(p)) = \rho^m\, \omega_m \left(1 - \frac{1}{6\,(m+1)}\, R(p)\, \rho^2 + \mathcal{O}(\rho^4)\right)$$

où ω_m est la mesure m-dimensionnelle de la sphère unité de $\mathbb{R}^{m+1}$.

Encadré 3

Un exemple explicite

Dans le cas particulier où $M^{m+1} = S_1^{m+1}$, la sphère unité de $\mathbb{R}^{m+2}$, et $K = \{0\} \times S_1^k$, on considère pour $r \in (0,1)$, l'hypersurface

$$\Sigma(r) := S_r^{m-k} \times S^k_{\sqrt{1-r^2}}$$

dont la courbure moyenne est constante

$$H(\Sigma(r)) = (m-k)\frac{\sqrt{1-r^2}}{r} - k\frac{r}{\sqrt{1-r^2}}.$$

Nous avons là un exemple explicite d'hypersurfaces dont l'existence est assurée par le théorème ci-dessus. On montre en outre que, lorsque le paramètre r tend vers 0, il existe une infinité de points de bifurcation qui donnent lieu à des hypersurfaces de S_1^{m+1} dont la courbure moyenne est constante mais qui ne sont pas aussi symétriques que $\Sigma(r)$. Ce résultat de bifurcation est en fait à rapprocher du phénomène de résonance mentionné ci-dessus.

Pour en savoir plus

[K-05] Kapouleas (N.), *Construction of Minimal Surfaces by Gluing Minimal Immersions*, Global Theory of Minimal Surfaces, Clay Mathematics Proceedings, D. Hoffman Edt, AMS (2005).

[KMP] Kusner (R.), Mazzeo (R.), Pollack (D.), *The moduli space of complete embedded constant mean curvature surfaces*, Geom. Funct. Anal. 6, 120-137 (1996).

[KGBS-03] Grosse-Brauckmann (K.), Kusner (R.), Sullivan (J.), *Triunduloids : Embedded constant mean curvature surfaces with three ends and genus zero*, J. Reine Angew. Math. 564, 35-61 (2003).

[MMP-05] Mahmoudi (F.), Mazzeo (R.), Pacard (F.), *Constant mean curvature hypersurfaces condensing along a submanifold*, preprint.

[R-05] Ros (A.), *The isoperimetric problem*, Global Theory of Minimal Surfaces, Clay Mathematics Proceedings, D. Hoffman Edt, AMS (2005).

[Y-91] Ye (R.), *Foliation by constant mean curvature spheres*, Pacific J. Math. 147, no. 2, 381-396 (1991).

Je voudrai remercier Nick Schmitt (GANG, University of Massachusetts at Amherst) pour m'avoir autorisé à utiliser ses images de surfaces qui sont disponibles sur le site http://www.gang.umass.edu

La plus grande valeur propre de matrices de covariance empirique

Sandrine PÉCHÉ*

Dans cet article nous expliquons brièvement l'intérêt de l'étude des valeurs propres extrêmes de matrices aléatoires hermitiennes de grande taille. Nous donnons ensuite les grandes lignes des méthodes d'étude de ces propriétés fines du spectre.

Matrices de covariance empirique

Une motivation statistique

Bob est statisticien et travaille dans le département de recherche d'une grande banque. Il veut modéliser l'évolution du marché pour les 10 prochaines années, afin de pouvoir faire des prévisions sur les cours d'un grand nombre d'actions. Pour ce faire, il choisit de représenter dans un tableau les rendements d'un grand nombre N d'actions sur un grand nombre p de jours. Typiquement p et N peuvent être d'ordre 10^4. On peut montrer qu'en utilisant un modèle aléatoire, on obtient une bonne représentation de ces rendements. En effet, les variations des cours sont particulièrement désordonnées, les marchés étant très sensibles à de multiples facteurs, qui peuvent varier suivant les titres et au cours du temps. Ainsi Bob choisit de modéliser l'évolution des rendements dans une matrice de taille $N \times p$, $M = [Z_1, \ldots, Z_N]$, où les Z_i sont des vecteurs de taille p représentant les p valeurs des rendements de chaque action. Il fait les hypothèses suivantes :

– les entrées $Z_{ij}, j = 1, \ldots, p$ des vecteurs Z_i sont des variables aléatoires

– elles sont mutuellement indépendantes.

La matrice $M = [Z_1, \ldots, Z_N]$ est ce que l'on appelle une *matrice aléatoire réelle*, dont l'étude a réellement commencé avec [10] en physique nucléaire.

Le problème posé : Afin de donner à ses supérieurs un compte-rendu fiable des données prévisionnelles, sans entrer dans une lecture longuette de tous les chiffres, Bob doit donc trouver un moyen de résumer au mieux l'information dont il dispose, tout en pouvant préciser quelle erreur il commet en la résumant.

Le problème de Bob est en fait un vieux de problème de statistique. Dès les années 1930, Wishart et James ([2]) sont les premiers à considérer le moyen optimal de résumer l'information statistique recueillie en effectuant p mesures sur une population de taille N. Le but est de minimiser les coûts des calculs numériques sur des très grandes matrices. Ce moyen sera par la suite défini par [1] en 1933, donnant naissance à l'analyse par « composantes prin-

* *Institut Fourier, Université Joseph Fourier,*
UMR 5582, BP 74, 38402 St Martin d'Heres Cedex, France.
sandrine.peche@ujf-grenoble.fr

cipales ». Nous allons rappeler et expliquer les grands principes de cette méthode et l'intérêt de l'étude des plus grandes valeurs propres de certaines matrices de « covariance empirique » pour son application.

Matrices de covariance empirique

Commençons par quelques rappels. Soient p et N des entiers fixés avec $p \geqslant N$. Soit M une matrice complexe de dimension $N \times p$. La matrice M^*, de dimension $p \times N$, est définie par $M^*_{ji} = \overline{M_{ij}}$ pour $i = 1,\dots,N$ et $j = 1,\dots,p$.

Définition 1. *Une matrice$X$ est Hermitienne si $X = X^*$. Une matrice Hermitienne X est positive si $\forall V \in \mathbb{C}^N$, $V^*XV \geqslant 0$.*

Typiquement, la matrice $X = MM^*$ est une matrice Hermitienne positive.

D'après le théorème spectral, on sait que la matrice X est diagonalisable dans une base orthonormée B de $\mathbb{C}^N$, $B = (V_1,\dots,V_N)$, et admet N valeurs propres positives $\lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant 0$. Ceci s'écrit mathématiquement $X = VDV^*$, si V est la matrice $(V_1 \ \cdots \ V_N)$ et $D = \mathrm{diag}(\lambda_1,\dots,\lambda_N)$, avec $XV_i = \lambda_i V_i$, pour $i = 1,\dots,N$, et $VV^* = Id$.

Remarque 2. *On sait aussi que la matrice $W = M^*M$ de taille $p \times p$ admet les mêmes valeurs propres non nulles que la matrice X. Ainsi il existe une matrice U de taille $p \times p$ telle que $UU^* = Id$ et $W = U\mathrm{diag}(\lambda_1,\dots,\lambda_N,0,\dots,0)U^*$.*

Nous donnons maintenant la définition dans un cadre général d'une matrice de covariance empirique complexe.

Soit μ,μ' deux mesures de probabilité sur $\mathbb{R}$.

Définition 3. *Une matrice de covariance empirique complexe est une matrice X_N avec $X_N = \frac{1}{N}MM^*$ si $M = Z + iY$, avec Z,Y de taille $N \times p$ dont les entrées $Z_{i,j}$, (resp.Y_{ij}), $i = 1,\dots,N$, $j = 1,\dots,p$, sont des variables aléatoires indépendantes identiquement distribuées (i.i.d.) de loi μ (resp. μ').*

On supposera toujours que les lois μ et μ' sont centrées et de variance finie indépendante de N. On note alors $\sigma^2 := Var(Z_{11}) + Var(Y_{11})$.

Exemple important : Si les entrées Z_{ij} et Y_{ij} sont des Gaussiennes i.i.d. $\mathcal{N}(0,\sigma^2/2)$, la matrice $X_N = X_N^{LUE}$ est alors dite de l'ensemble de Laguerre unitaire, noté LUE. Le LUE est la loi de cette matrice.

Dans la suite, on note W_N la matrice $\frac{1}{N}M^*M$, naturellement associée à X_N. On note aussi $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N \geqslant 0$ les valeurs propres ordonnées de X_N et V_i (resp. U_i) les vecteurs propres orthonormés de X_N (resp. W_N) associés.

Le problème de Bob : Supposons que Bob a modélisé les cours par une matrice aléatoire réelle M comme dans la définition 3. Cette modélisation est ici trop simpliste, mais elle va nous permettre d'expliquer les principes de l'étude des matrices de covariance. Bob cherche une approximation de la matrice $M_N = N^{-1/2}M$ par une matrice Y de rang plus petit. Il veut aussi maximiser la qualité de l'approximation, mesurée par la quantité

$$Q(Y) = \frac{\sum_{i=1}^N \sum_{j=1}^p |Y_{ij}|^2}{\sum_{i=1}^N \sum_{j=1}^p |(M_N)_{ij}|^2}.$$

L'analyse par composantes principales résout ce problème. La meilleure approximation de M_N par une matrice de rang 1 (par exemple) est alors la matrice $M_1 = \sqrt{\lambda_1} V_1 U_1^*$ et sa qualité est $Q(M_1) = \dfrac{\lambda_1}{Tr(X_N)}$. La meilleure approximation de rang $k > 1$ est aussi connue et sa qualité s'exprime en fonction des k plus grandes valeurs propres. Revenons à l'approximation de rang 1. Afin de contrôler l'erreur commise, Bob doit déterminer les propriétés de la plus grande valeur propre λ_1 et de la trace de la matrice X_N. Calculer toutes les valeurs propres de X_N pour déterminer ensuite la plus grande, et ce pour chaque réalisation aléatoire est numériquement trop coûteux. Nous allons ici donner d'autres méthodes permettant d'étudier λ_1, et $Tr(X_N)$, quand N est grand.

Comportement global des valeurs propres

Nous allons maintenant obtenir une première borne inférieure pour la plus grande valeur propre (et identifier au passage le comportement limite de $Tr(X_N)$). Pour simplifier, on suppose à partir de maintenant que $p - N$ est un entier fixé. Le nombre de données p est donc comparable à la taille de la population N (*cf.* [3] pour une étude plus générale).

La méthode est la suivante. Etant donné un intervalle borné I, nous allons dénombrer le nombre de valeurs propres qui tombent dans cet intervalle.

Théorème 4. *Soit I un intervalle quelconque. La proportion des valeurs propres de X_N dans I, notée $N_N(I)$, est donnée asymptotiquement par*

$$\lim_{N\to\infty} N_N(I) = \int_I \frac{1}{2\pi\sigma^2\sqrt{x}}\sqrt{4\sigma^2 - x}\,1_{[0,4\sigma^2]}(x)dx. \qquad (1)$$

Remarque 5. *La fonction intégrée dans (1) est une densité de probabilité. Elle définit la loi dite de Marchenko-Pastur.*

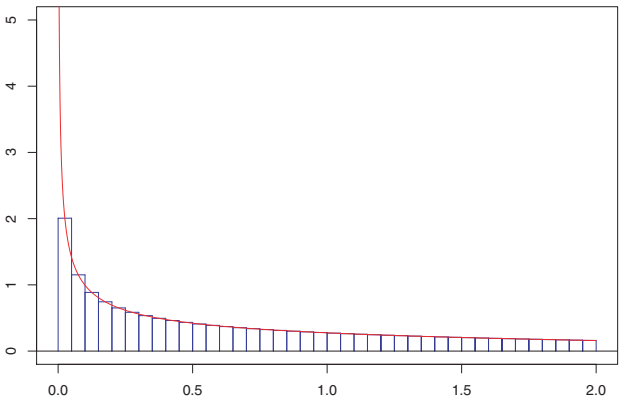


Figure 1 – *Histogramme des valeurs propres et loi de Marchenko-Pastur.*

La Figure 1 montre l'histogramme des valeurs propres d'une matrice de taille 40, et montre l'adéquation avec la densité de la loi de Marchenko et Pastur. Grossièrement, on est « sûr » de trouver, pour N assez grand, des valeurs

propres de X_N dans tout sous intervalle de $I = [0,4\sigma^2]$. On a donc, pour N assez grand, et en dehors d'un ensemble de probabilité nulle, $\limsup_{N\to\infty} \lambda_1 \geqslant 4\sigma^2$.

Remarque 6. *Une preuve consiste à étudier les « moments » d'ordre $k \in \mathbb{N}$ fixé, $\frac{1}{N}\mathbb{E}\left(Tr\left(X_N^k\right)\right) = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_i^k\right)$, et de montrer que ces moments convergent, quand $N \to \infty$ vers le moment d'ordre k de la loi de Marchenko-Pastur, si la variance σ^2 est fixée. En particulier, $\frac{1}{N}Tr(X_N)$ converge vers σ^2.*

Des questions se posent alors naturellement. Par exemple,

– a-t-on $\lim_{N\to\infty} \lambda_1 = 4\sigma^2$?

– Si oui, quel est la vitesse de convergence de λ_1 vers $4\sigma^2$?

– La loi limite de $\lambda_1 - 4\sigma^2$ est-elle centrée ou non ?

Le théorème de Marchenko et Pastur ne nous permet pas de répondre à ces questions. Il concerne en effet un comportement de type global, à savoir les propriétés de toutes les valeurs propres, considérées simultanément. Pour étudier le comportement plus précis de λ_1, nous devons définir des outils qui permettent de différencier mieux les valeurs propres, sans toutefois revenir au calcul de chacune d'entre elles.

L'ensemble de Laguerre unitaire

Pour étudier le comportement de λ_1, nous avons besoin de faire des hypothèses supplémentaires sur la loi des entrées de X_N. La seule connaissance de la variance σ^2 ne semble pas suffire. Nous allons donc nous intéresser plus particulièrement au LUE. Le LUE présente en effet deux particularités, qui ne sont pas vraies en général pour des entrées non Gaussiennes, et qui font que c'est l'ensemble mathématiquement parlant le plus simple.

D'abord, on sait calculer explicitement la densité de probabilité jointe $g : \mathbb{R}_+^N \to \mathbb{R}_+$ des valeurs propres du LUE. Le calcul remonte à [2]. Cette densité est importante car elle permet, *a priori*, de déterminer la fonction de répartition de la plus grande valeur propre. En effet,

$$\begin{aligned}\mathbb{P}(\lambda_1 \leqslant s) &= \mathbb{P}(\text{toutes les valeurs propres sont dans } (-\infty,s]) \\ &= \int_{-\infty}^{s}\cdots\int_{-\infty}^{s} g(x_1,\cdots,x_N)\prod_{i=1}^{N}dx_i. \end{aligned} \tag{2}$$

Reste le calcul de cette intégrale N-dimensionnelle, ce qui n'est pas connu en général. C'est là la deuxième particularité du LUE. On peut explicitement calculer la fonction de répartition, ce qui a été obtenu par Bronk, [4]. Ce calcul utilise une formidable astuce initialement due à Mehta, Gaudin, ([8], chap. 5), qui exprime (2) comme une certaine fonction des polynômes orthogonaux de Laguerre (*cf.* [7]) ! Ceci explique d'ailleurs la dénomination LUE. Ces polynômes étant parfaitement bien connus, on a pu en déduire le comportement asymptotique de λ_1 pour le LUE, que nous donnons maintenant.

On appelle loi de Tracy-Widom, de fonction de répartition F_2, la mesure de probabilité dont la densité est donné par la Figure 2. La définition mathématique de cette loi, plutôt compliquée, est donnée dans [9].

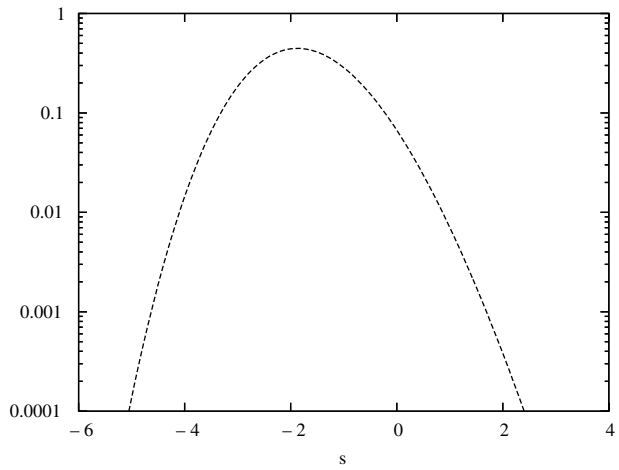


Figure 2 – *Densité de la loi de Tracy-Widom.*

Théorème 7. *[5] Soit $x_o \in \mathbb{R}_-$ fixé et $x \in [x_o, \infty)$. Alors,*

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3} \frac{\lambda_1^{LUE} - 4\sigma^2}{4\sigma^2} \leqslant x \right) = F_2(x).$$

Le Théorème 7 répond ainsi aux questions de la section précédente, dans le cas du LUE. D'abord, quand $N \to \infty$, λ_1^{LUE} converge vers $4\sigma^2$. Elle n'a donc pas tendance à se séparer des autres valeurs propres. De plus, elle fluctue autour de $4\sigma^2$ dans des intervalles de longueur typique d'ordre $N^{-2/3}$. Les fluctuations de cette valeur propre autour de $4\sigma^2$ et dans l'échelle typique sont aléatoires et de loi asymptotiquement donnée par la loi de Tracy-Widom, qui n'est pas centrée.

Modèles de matrices plus généraux

Une fois le comportement de λ_1 identifié pour le LUE, on montre que ce comportement est en fait valable pour des matrices X_N plus générales. L'idée de base est la suivante. Soit $A > 0$ fixé (grand), et k le nombre des valeurs propres $\lambda_i > 4\sigma^2 - AN^{-2/3}$. Ecrivons alors $\lambda_i = 4\sigma^2(1 + \chi_i N^{-2/3})$, où χ_i est une variable aléatoire de loi F_i. On obtient

$$\mathbb{E}\left(Tr\left(\frac{X_N}{4\sigma^2} \right)^{tN^{2/3}} \right) = \mathbb{E}\left(\sum_{i=1}^{N} \left(\frac{\lambda_i}{4\sigma^2} \right)^{tN^{2/3}} \right) \simeq \mathbb{E}\left(\sum_{i=1}^{k} \int_{\mathbb{R}} e^{tx} dF_i(x) \right).$$

Si on montre que, pour tout $j \in \mathbb{N}$ et pour tout réels positifs $t_1, \ldots, t_j$, la limite $\lim_{N\to\infty} \mathbb{E}\left(\prod_{i=1}^{j} Tr\left(\frac{X_N}{4\sigma^2} \right)^{t_i N^{2/3}} \right)$ existe et ne dépend que de σ^2, comme dans le cas du LUE, alors on montre « grossièrement » que les plus grandes valeurs propres de X_N ont le même comportement asymptotique que celles du LUE.

Théorème 8. *[6] Si pour tout entier k, $\mathbb{E}\left((M_{11})^{2k+1}\right) = 0$, et $\mathbb{E}\left(|M_{11}|^{2k}\right) \leqslant (Ck)^k$, alors*

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3} \frac{\lambda_1^{LUE} - 4\sigma^2}{4\sigma^2} \leqslant x \right) = F_2(x).$$

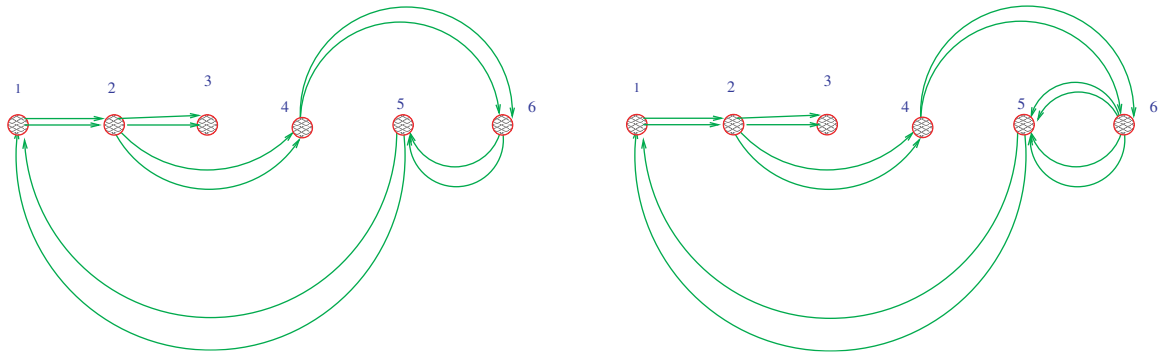
Remarque 9. *Une matrice du LUE satisfait les conditions du Théorème 8.*

Pour esquisser la preuve, supposons que $j = 1$ et que $tN^{2/3} = l_N$ est un entier pair. On développe la trace

$$\mathbb{E}\left(Tr\left(\frac{X_N}{4\sigma^2}\right)^{l_N}\right) = \frac{1}{(4\sigma^2)^{l_N}} \sum_{i_o,\ldots,i_{l_N-1}=1}^{N} \mathbb{E}\left(X_{i_o i_1} X_{i_1 i_2} \cdots X_{i_{l_N-1} i_0}\right)$$

$$= \frac{1}{(4\sigma^2 N)^{l_N}} \sum_{k_1,\ldots,k_{l_N-1}=1}^{p} \sum_{i_o,\ldots,i_{l_N-1}=1}^{N} \mathbb{E}\left(M_{i_o k_1} \overline{M_{i_1 k_1}} \ldots M_{i_{l_N-1} k_{l_N-1}} \overline{M_{i_o k_{l_N-1}}}\right). \tag{3}$$

A chacun des termes de (3), on associe un graphe orienté : on trace les arêtes du sommet i_o vers k_1, de i_1 vers k_1, ..., de i_{l_N-1} vers k_{l_N-1} et de i_o vers k_{l_N-1}. On obtient donc $4s_N$ arêtes reliant des sommets choisis dans $\{1,\ldots,N\}$ ou $\{1,\ldots,p\}$. On regroupe les termes associés à chaque arête orientée et on calcule l'espérance associée. Or, dès qu'une arête apparaît un nombre impair de fois, cette espérance est nulle. On tient donc compte des seuls graphes où chaque arête est parcourue un nombre pair de fois.



(a) Chemin valide (b) Chemin négligeable

Figure 3

L'étape suivante est de montrer que les seuls graphes qui vont avoir une contribution significative sont ceux pour laquelle chaque arête est passée au plus deux fois. Par exemple, le chemin de la Figure 3, où l'arête (6,5) est passée 4 fois, ne sera pas parmi les chemins à comptabiliser. En effet, on choisit beaucoup moins de sommets (puisqu'on en répète) que dans le cas où les arêtes sont « doubles ». La présence du facteur $1/N^{l_N}$ fait alors que ces chemins avec arêtes plus que doubles sont négligeables. Or pour chaque arête passée deux fois, on a $\mathbb{E}\left(M_{ij} M_{ji}\right) = \sigma^2$ ou, les entrées étant complexes non réelles, $\mathbb{E}\left(M_{ij} M_{ij}\right) = 0$. Le résultat final s'exprime donc en fonction de σ^2 seulement, comme dans le cas du LUE. On en déduit ensuite le Théorème 8.

Conclusion

Nous avons donné ici les idées d'une méthode d'étude des plus grandes valeurs propres dans le cas général de matrices aléatoires complexes. Des petites modifications sont à apporter dans le cas de matrices réelles, où les formules sont en fait plus compliquées ! Concluons avec le problème de Bob. En choisissant des modèles de matrices aléatoires un peu plus compliqués (et plus proches de la réalité) que celui présenté ici, par exemple avec des entrées M_{ij} de variances différentes, on peut montrer que les marchés sont en fait très bien représentés par des matrices de très faible rang (au plus 10). C'est un point très interessant pour la gestion d'un portefeuille...

Pour en savoir plus

[1] Hotelling (H.), Analysis of a complex of statistical variables into principal components, *Jour. Educ. Psych.*, 24 : 417–441, (1933).

[2] James (A.), Distributions of matrix variates and latent roots derived from normal samples, *Annals of Mathematical Statistics*, 35 : 475–501, (1964).

[3] Marcenko (V.A.) and Pastur (L.A.), Distribution of eigenvalues for some sets of random matrices, *Math. USSR-Sbornik*, 1 : 457–486, (1967).

[4] Bronk (B.V.), Exponential ensembles for random matrices, *J. Math. Phys,* 6 : 228–237, (1965).

[5] Forrester (P.J.), The spectrum edge of random matrix ensembles, *Nuclear Physics B*, 402 : 709–728, (1993).

[6] Soshnikov (A.), A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, (2001), Preprint, arXiv : math. PR/0104113 v2.

[7] Szego (G.), *Orthogonal polynomials*, American Mathematical Society, Providence, RI, (1967).

[8] Mehta (M.), *Random matrices,* Academic press, San Diego, second edition, (1991).

[9] Tracy (C.) and Widom (H), Level spacing distributions and the Airy kernel, *Comm. Math. Phys*, **159** : 33–72, (1994).

[10] Wigner (E.) Characteristic vectors bordered matrices with infinite dimensions, *Ann. Math.* **62** : 548–564, (1955).

Rôles des figures dans la production et la transmission des mathématiques

Jeanne PEIFFER*

L'histoire des images en mathématiques, balbutiante, est encore à la recherche d'outils appropriés. Deux approches sont ici décrites. La première, située à la conjonction de l'histoire des sciences et de l'histoire des pratiques culturelles, s'intéresse à la matérialité de la communication scientifique et donc aux textes fabriqués pour un public. La seconde se place dans l'espace de la recherche et étudie les premières matérialisations, sous forme de figures, des idées dans le cerveau des mathématiciens. Quelques exemples illustrent brièvement ces approches.

Introduction

Prenant le titre de cette publication au mot, la présente contribution traite d'images et de leurs présences en histoire des mathématiques : figures, diagrammes, illustrations ou autres formes de représentations visuelles. Pour anciens qu'ils soient – certains des dessins trouvés dans des grottes ornées ont été interprétés comme ayant une signification mathématique –, les liens des images avec les mathématiques sont compliqués. Tantôt les figures sont considérées n'en faire qu'un avec le texte mathématique, tantôt elles en sont bannies. Ainsi, Lodovico Cigoli écrivait le 11 août 1611 à son ami Galilée : « *un matematico, sia grande quanto si vole, trovandosi senza disegnio, sia non solo un mezzo matematico, ma anche uno huomo senza occhi* »[1]. Alors qu'en 1788, Lagrange affirme dans l'avertissement à sa *Mechanique analytique* : « On ne trouvera point de Figures dans cet Ouvrage. Les méthodes que j'y expose ne demandent ni construction, ni raisonnemens géométriques, ou mécaniques, mais seulement des opérations algébriques, assujetties à une marche régulière et uniforme ». Pour le premier, savoir dessiner fait partie du métier de mathématicien alors que pour le second le rejet des figures va de pair avec celui de la géométrie en faveur de la régularité des opérations algébriques. Chez Lagrange, le banissement des figures traduit un nouvel équilibre entre deux branches des mathématiques, la prédominance de l'analyse algébrique sur la géométrie. C'est dire que la question des figures se trouve au cœur de certaines représentations que l'on s'est faites des mathématiques dans l'histoire. La juxtaposition des deux citations éclaire par ailleurs l'historicité de ce lien, qui change avec le temps et le lieu. C'est ce que nous allons éclairer dans la suite par quelques exemples.

* *Centre Alexandre Koyré, UMR CNRS-EHESS- MNHN 8560*
27 rue Damesme, 75013 Paris.
peiffer@damesme.cnrs.fr

1. Cité d'après [M1, 282]. En voici une traduction approximative : « un mathématicien, aussi grand soit-il, ne possédant pas le dessin ne sera non seulement à moitié mathématicien, mais encore un homme sans yeux ».

Un objet à construire

De fait, l'objet de cette histoire – images en mathématiques – reste à construire. Jusqu'à tout récemment, les historiens des mathématiques n'y ont guère attaché d'importance. Mais, sous la double impulsion de l'histoire des sciences et de celle du livre, nos collègues ont commencé à étudier les pratiques matérielles de production et de circulation des savoirs, notamment scientifiques. En effet, des articles assemblés en livres – comme par exemple les périodiques qui sont avec les monographies la forme prépondérante de circulation du savoir mathématique après la montée en puissance de l'imprimerie et avant la dématérialisation qui caractérise la publication électronique – sont aussi des objets issus de processus de fabrication et mis en circulation sur un marché. Cette insistance sur la matérialité de la communication scientifique et les savoir-faire qu'elle requiert a mis en avant l'étude de ce que les historiens anglo-saxons appellent les « *inscription devices* » [L1] ou « *representational technologies* » et de leur histoire sociale. Ainsi, Simon Schaffer affirme : « *The work involved in making pictures is a fundamental aspect of the labor process of the sciences* » [L1, 184], mais il souligne aussi le fait que l'étude des images exige une méthodologie spécifique qui doit s'appuyer selon lui sur les outils de l'iconographie et de l'histoire culturelle. En France, Karine Chemla[2] a, quant à elle, fait le choix d'une approche linguistique.

Les programmes de recherche brièvement esquissés ci-dessus prennent pour cible les textes fabriqués à l'intention d'un public, leur transmission et leur réception. Mais on a également vu apparaître des études qui s'intéressent peut-être davantage aux processus même de la recherche. Elles s'appuient sur les archives scientifiques – cahiers de laboratoire, notes manuscrites, brouillons, etc. – et tentent de mieux comprendre les mécanismes d'écriture et finalement les processus intellectuels de création scientifique dont ceux-là sont l'indice[3].

Questions nouvelles concernant les mathématiques grecques

En histoire des mathématiques, on commence à peine à se pencher sur ce type de questions. Comme souvent, les innovations viennent des mathématiques anciennes dont le corpus très réduit et lacunaire pousse à renouveler les problématiques capables de l'éclairer. Ainsi l'importance des figures pour la constitution et la transmission des textes de l'Antiquité a été soulignée et leur étude entreprise notamment de la part des linguistes et sémiologues (comme Reviel Netz et Micheline Decorps). Chez Platon ou Aristote, la distance est nettement établie entre l'objet géométrique que la proposition vise à construire et la figure matérielle qui le représente. Les géomètres utilisaient des mots différents pour désigner la figure au sens d'« objet géométrique » et la figure comme dessin (D1, 65). Texte et figure forment pourtant un tout dans la pratique de la démonstration en Grèce ancienne. Analysant les outils mathématiques à l'œuvre dans les *Eléments* d'Euclide, Reviel Netz, [N1], place le renvoi à des figures lettrées au premier rang. Pour lui la figure est indispensable, car des énoncés sont déduits directement des figures sans pouvoir l'être du texte. Au contraire, Micheline Decorps, dans son examen minutieux des figures d'Apollonius puis de celles du commentateur Eutocius, voit dans la figure un support visuel à une démonstration achevée, sans que le texte d'Apollonius établisse un lien explicite avec la figure. Eutocius fait de celle-ci un outil pédagogique lorsqu'il met sans cesse une figure sous les yeux du lecteur, soit pour expliciter le propos d'Apollonius soit pour le compléter ou l'enrichir. Dans les manuscrits médiévaux, les coniques tracées par les copistes à la règle et au compas sont composées d'arcs de cercle, et permettent seulement un repérage des points. La représentation de ces courbes relève donc de la convention. C'est aussi la conclusion à laquelle arrive De Young (D2) qui vient de retrouver quelques figures géométriques d'une des traditions arabes de transmission des *Eléments*. Pour lui, les figures techniques sont imprégnées de postulats et conventions spécifiques à la culture dans laquelle elles ont été tracées, mais il doit aussi les considérer, dans l'état actuel de nos connaissances, comme « *an incomplete fossil record from which we attempt to reconstruct an organism and its relations to its environment* [D2,162] ».

Le langage visuel de la géométrie du 16e siècle

Les travaux sur la géométrie grecque s'interrogent sur le statut et le rôle des figures, leur relation au texte, leur transmission, l'existence de traditions et leur stabilité au cours du temps. Que tous ces éléments varient devient manifeste lorsqu'on formule ces mêmes questions pour les figures d'une autre époque, celles des textes mathématiques du 16e siècle par exemple. Il est alors difficile de séparer la géométrie des arts entendus en un sens très large, allant des

2. Karine Chemla anime depuis des années à Paris un séminaire intitulé « Histoire des sciences, histoire du texte ». Voir [C1] pour son programme.
3. La revue *Genesis* a consacré son numéro 20 (2003) à l'*écriture scientifique*.

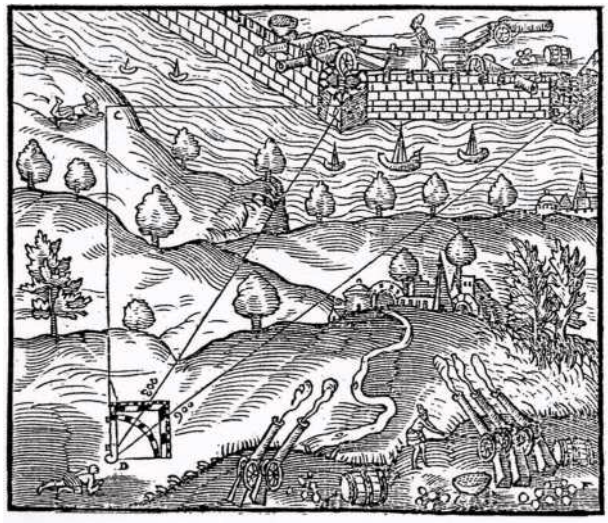


Figure 1 – *Extraite de Leonard Digges,* Pantometria.

pratiques graphiques des métiers aux pratiques artisanales et artistiques. Il suffit pour s'en rendre compte d'ouvrir un ouvrage de géométrie pratique, comme la *Pantometria* de Leonard Digges publiée par son fils Thomas en 1571.

La figure géométrique est superposée à un paysage représentant de façon suggestive, même si rudimentaire, le siège d'une ville. Le quadrant dont elle met en scène l'usage n'est cependant pas à l'échelle, au contraire il apparaît bien trop grand par rapport aux autres éléments de la gravure sur bois. Loin de servir de support visuel à une démonstration, comme en Grèce, les figures des traités la Renaissance mettent souvent l'accent sur l'utilité pratique de la géométrie, dont les instruments servent à observer, mesurer et maîtriser le terrain.

La distinction grecque entre objet géométrique et sa représentation tend à disparaître dans certains textes de la Renaissance, comme dans l'*Underweysung der messung* (Nuremberg 1525) d'Albrecht Dürer, où les figures sont de fait très souvent des patrons à découper. L'objet géométrique non seulement coïncide avec la figure, mais découpée elle sert de modèle ou d'outil dans les ateliers. Certaines de ces figures se sont transmises sur la longue durée. Ainsi, on retrouve le patron du dodécaèdre dessiné par Dürer (figure 33 du Livre IV de son *Underweysung*) dans Simon Stevin, et jusque dans un manuscrit intitulé « Introduction Géométrique à l'Etude de la Geographie … », conservé à la British Library de Londres et longtemps faussement attribué à d'Alembert.

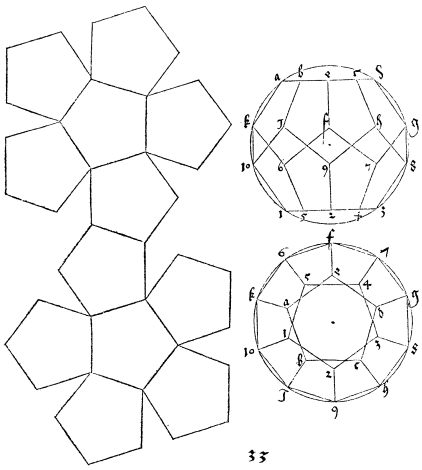


Figure 2 – *Albrecht Dürer,* Underweysung der messung, *Livre IV, figure 33.*

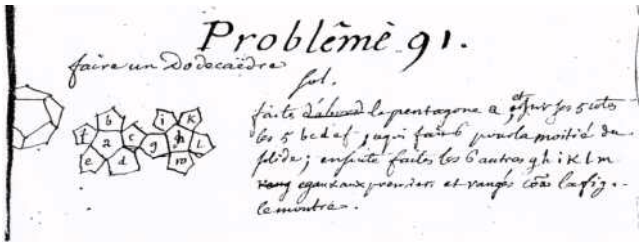


Figure 3 – *Dodécaèdre dans un manuscrit anonyme (British Library 22758).*

Préfigurations mathématiques

La figure, si elle peut être simple support visuel d'une démonstration, outil pédagogique, élément central de l'argumentation, simple ornement, etc., peut aussi aller bien au-delà du texte et dépasser l'information mathématique qui y est contenue. Il arrive au lecteur d'aujourd'hui de découvrir, dans les textes mathématiques du passé – et même sur des objets ornés – des représentations visuelles, des figures, qui donnent accès à des réalités mathématiques dont les mathématiciens ne s'étaient pas encore saisis et qui ne faisaient donc pas partie du corpus mathématique de l'époque. Dürer excelle dans ce type de constructions inspirées de pratiques graphiques. Federico Amodeo, puis René Taton, ont naguère attiré l'attention sur la présence d'épures de géométrie descriptive dans l'*Underweysung der messung*, alors que cette discipline n'a été élaborée par Gaspard Monge que près de trois siècles plus tard. Dans sa figure 38 du Livre I, on voit apparaître une parabole comme enveloppe de ses tangentes. Dürer l'engendre point par point en plaçant l'extrémité d'une règle de longueur fixe *ab* successivement sur les points de l'axe horizontal (dont une partie est divisée par 16 points en 16 intervalles égaux) et en la faisant passer par les points de même nom de l'axe vertical issu du point 13. L'autre extrémité désigne les points successifs de la courbe.

La présence de ces figures pose le problème proche de celui, actuellement très étudié en ethnomathématique, de la reconnaissance d'activités mathématiques non identifiées comme telles par ceux qui les pratiquent. Nous reconnaissons dans la figure de Dürer l'enveloppe d'une famille de courbes, mais Dürer n'avait aucune connaissance de cette notion qu'il a pourtant su représenter dans un cas particulier. Sa figure a un contenu mathématique plus riche que le texte accompagnateur. La représentation précède l'élaboration de l'objet mathématique.

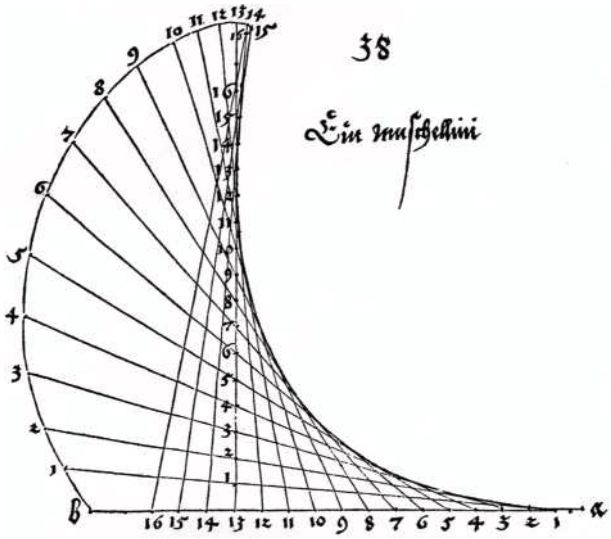


Figure 4 – *Ligne en forme de coquille (Dürer,* Underweysung der messung*, Livre I, figure 38).*

Dans l'espace de la recherche : le dessin comme expression de la pensée

Dans les exemples décrits, nous avons considéré des figures prises dans un processus de transmission, qu'il soit manuscrit ou passe par l'imprimé. Or, les archives scientifiques recèlent elles aussi parfois des trésors d'images, des figures gribouillées par les mathématiciens pour eux-mêmes dans le processus de création. L'exemple que je vais présenter ici est emprunté à Henk Bos [B1] et concerne Christiaan Huygens et la chaînette, c'est-à-dire la courbe qu'épouse une « corde librement suspendue entre deux points fixes » (dans la formulation de Jean Bernoulli).

Dans un manuscrit de 1646, le jeune Huygens – il a alors 17 ans – utilise l'approximation suivante : une corde supposée sans poids est chargée de poids égaux suspendus à des distances égales. Grâce à la statique, Huygens détermine ce qui se passe pour chacun des poids, puis extrapole au cas continu, où les poids sont répartis uniformément tout au long de la courbe. Cette approximation, qui permet un passage à la limite, du cas discret au cas continu, lui permet de conclure que la courbe cherchée ne peut être une parabole, comme Galilée l'avait suggeré. Huygens communiqua ce résultat au Père Mersenne dans une des premières lettres qu'il lui adressa. Mais il ne détermina la nature de la courbe qu'en 1690 lorsque Jakob Bernoulli lança son défi concernant la chaînette dans les *Acta eruditorum*. Huygens semble avoir eu besoin d'une impulsion de l'extérieur, d'un problème posé et à resoudre, pour se mettre à réfléchir en gribouillant. Selon Bos, ces figures aident Huygens à ordonner une information spatiale complexe.

Michael Mahoney et Henk Bos ont l'un et l'autre examiné quelques figures excessivement complexes de ce que Joella Yoder a appelé la « cinématique géométrique » [Y1,52] de Huygens. Sans qu'il soit possible de faire justice dans un texte aussi bref que celui-ci à la richesse de leur analyse, résumons-en deux conclusions intéressantes pour notre propos. D'abord les figures permettent à Huygens de représenter l'irreprésentable, comme les paramètres physiques du mouvement : la vitesse, l'accélération, le temps et leurs relations mutuelles. Puis, Mahoney a pu montrer que dans les figures de Huygens trois strates différentes se superposent : un espace physique, l'espace mécanique des vitesses, temps, etc, et finalement l'espace mathématique. C'est par un va et vient entre ces strates que Huygens parvient à imaginer des solutions hautement techniques et singulières. Mahoney comme Bos sont d'accord pour voir la figure fonctionner chez Huygens comme un modèle géométrique d'un phénomène naturel complexe. L'art du dessin scientifique lui permet d'étudier ces phénomènes. La figure est plus proche de la pensée mathématique de Huygens que les équations et les formules qu'il note et publie ensuite, même si on retrouve dans celles-ci les éléments que le modèle exhibe.

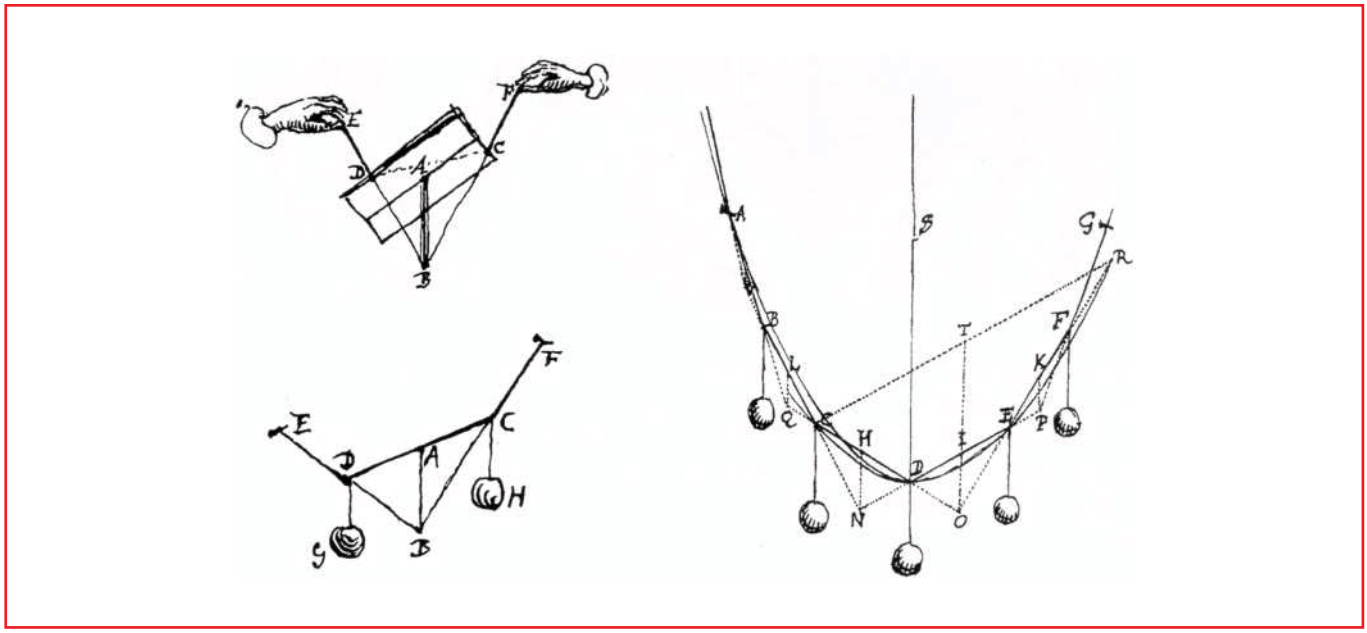


Figure 5 – *Recherches sur la chaînette, 1646, Christiaan Huygens,* Œuvres complètes *XI, 37-40.*

Pour conclure

Les exemples cités dans ce bref survol concernent tous une époque dont la citation de Lagrange annonce la fin. Mais ses modes graphiques de pensée ont marqué l'analyse et la mécanique naissantes. Les figures ont pu par moments s'éclipser des textes mathématiques, mais des mathématiciens continuent à dessiner et à réfléchir en dessinant. En témoigne une mathématicienne contemporaine : « … je fais beaucoup de dessins, mais de grands dessins,

représentant le même concept, avec ce que je crois être des angles de vue différents. Puis je les jette, quelques jours ou quelques mois plus tard, quand je pense avoir trouvé la ou les bonnes représentations, ou, parce qu'ils ne me satisfont pas » [C2, 105].

Pour en savoir plus

[B1] Baigrie (B. S.), ed., Picturing Knowledge. Historical and Philosophical Problems Concerning the Use of Art in Science, *Toronto University Press* (1996).

[B2] Bos (H.J.M.), Huygens and Mathematics in Titan-From Discovery to Encounter, ed. by the *European Space Agency*, 67-80 (2004).

[C1] Chemla (K.), Histoire des sciences et matérialité des textes, *Enquête*, 1, 167-180 (1995).

[C2] Chotteau (T.) *et al*., Rencontres entre artistes et mathématiciennes, *L'Harmattan* (2001).

[D1] Decorps-Foulquier (M.), Sur les figures du traité des coniques d'Apollonios de Pergé édité par Eutocius d'Ascalon, *Revue d'histoire des mathématiques*, 5, 61-82 (1999).

[D2] De Young (G.), Diagrams in the Arabic Euclidean tradition: a preliminary assessment, *Historia mathematica*, 32, 129-179 (2005).

[D3] Dürer (A.), Géométrie. Présentation et traduction de J. Peiffer, *Le Seuil* (1995).

[F1] Ford (B.J.), Images of Science. A History of Scientific Illustration, *The British Library* (1992).

[G1] Gross (A.G.), Harmon (J.E.), Reidy (M.), Communicating Science. The Scientific Article from the 17th Century to the Present, *Oxford University Press* (2002).

[L1] Lenoir (T.), ed., Inscribing Science. Scientific Texts and the Materiality of Communication, *Stanford University Press* (1998).

[M1] Mahoney (M. S.), Drawing mechanics in Picturing Machines 1400-1700, ed. by Wolfgang Lefèvre, *The MIT Press*, 281-306 (2004).

[N1] Netz (R.), The Shaping of Deduction in Greek Mathematics. A Study in Cognitive History, *Cambridge University Press* (1999).

[Y1] Yoder (J.G.), Unrolling Time. Christiaan Huygens and the Mathematization of Nature, *Cambridge University Press* (1988).

A propos de la description des gaz parfaits

Laure SAINT-RAYMOND*

Le sixième problème posé par Hilbert au Congrès International des Mathématiciens en 1900 appelait une compréhension globale de la dynamique des gaz. Une analyse fine de l'équation de Boltzmann permet aujourd'hui d'obtenir rigoureusement une description multi-échelles complète des gaz parfaits en régime visqueux.

De la dynamique moléculaire aux modèles cinétiques

Au niveau microscopique, un gaz est constitué d'un grand nombre de particules élémentaires en interaction, dont la dynamique est régie par le principe de Newton. Dans le cas d'un gaz parfait, par exemple pour l'air dans les condi-

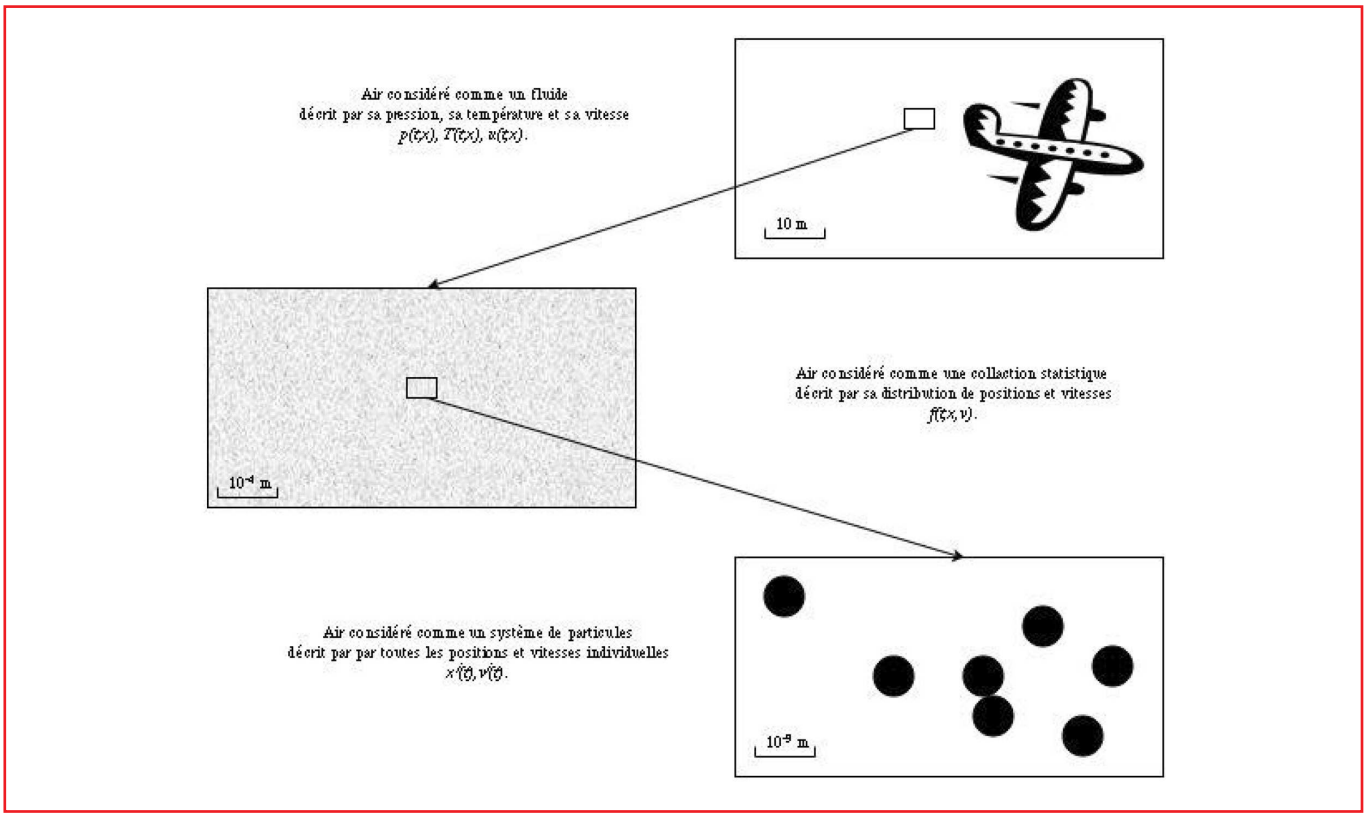


Figure 1 – *Echelles de description du gaz.*

* *Laboratoire Jacques-Louis Lions UMR 7598*
Université Paris VI, 175 rue de Chevaleret, 75013 Paris.
saintray@ann.jussieu.fr

tions normales de température et de pression, le volume occupé par les particules est négligeable par rapport au volume total du gaz, ce qui signifie que les collisions mettant en jeu plus de deux particules ou des particules ayant déjà interagi sont fortement improbables : seules les collisions binaires entres particules non corrélées ont un rôle déterminant dans l'évolution du gaz. La dynamique d'un gaz parfait peut alors être décrite par une approche statistique, à l'aide des fonctions de distribution de chaque espèce de particules (qui donnent le nombre instantanné de particules de position et vitesse quelconques fixées).

Pour un gaz parfait monoatomique, la fonction de distribution est régie par une équation aux dérivées partielles de type Boltzmann

$$Ma\partial_t f + v \cdot \nabla_x f = \frac{1}{Kn} Q(f),$$

qui prend en compte d'une part le transport des particules (membre de gauche) et d'autre part la modification des vitesses par les collisions (membre de droite) qui sont supposées instantannées et élastiques. Les propriétés de symétrie sur l'opérateur Q ainsi obtenu (opérateur intégral par rapport à la variable v) impliquent en particulier les deux principes fondamentaux de la thermodynamique, c'est-à-dire la conservation locale de la masse, de l'impulsion et de l'énergie, ainsi que la croissance locale d'une certaine quantité, appelée entropie. Les maximiseurs de l'entropie à masse, implusion et énergie fixées (qui sont aussi les annulateurs de l'opérateur Q) sont les distributions gaussiennes, conformément à la prédiction statistique de Boltzmann.

Approximations hydrodynamiques

Si les collisions sont suffisamment fréquentes, l'entropie du gaz croit rapidement et la distribution de vitesses en tout point de l'espace relaxe rapidement vers une Gaussienne. L'état du gaz est alors complètement déterminé par ses grandeurs thermodynamiques locales, à savoir sa température, sa pression et sa vitesse macroscopique d'écoulement. Des modèles hydrodynamiques permettent donc d'obtenir des approximations de l'équation de Boltzmann dans la limite de relaxation rapide, c'est-à-dire quand le nombre de Knudsen Kn (mesurant le rapport entre le libre parcours moyen et l'échelle de longueur considérée) est très petit.

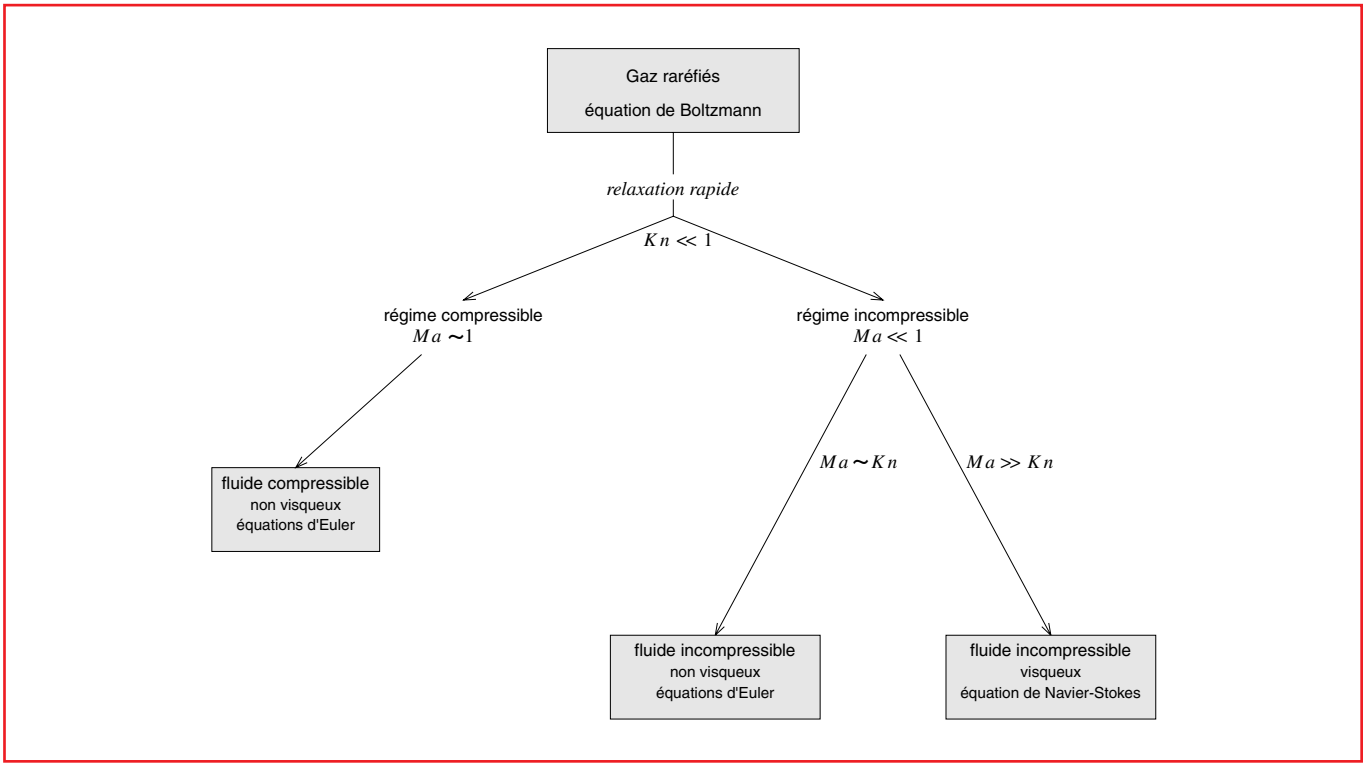


Figure 2 – *Approximations hydrodynamiques de l'équation de Boltzmann.*

Ces modèles hydrodynamiques dépendent d'une autre caractéristique du gaz, sa compressibilité. Pour mesurer la capacité du gaz à propager les variations de pression, on introduit un autre nombre sans dimension, le nombre de Mach Ma, qui est défini comme le rapport entre la vitesse moyenne d'écoulement et la vitesse d'agitation thermique (appelée aussi vitesse du son). Si le nombre de Mach est très petit, les variations de pression sont rapidement compensées par l'agitation thermique et le gaz est presque incompressible. Dans le cas où le nombre de Mach est aussi petit que le nombre de Knudsen, l'agitation thermique est tellement importante qu'elle induit des corrélations à l'échelle de longueur d'observation : l'écoulement est dissipatif, comme le prédit la relation de Von Karmann qui relie le nombre de Reynolds Re (inversement proportionnel à la viscosité cinématique) aux nombres de Mach et de Knudsen :

$$Re = \frac{Ma}{Kn}.$$

Ainsi les modèles compressibles visqueux ne sont pas obtenus comme limites fluides de l'équation de Boltzmann : pour les gaz parfaits, le volume des particules est négligeable devant le volume total occupé par le gaz, de sorte qu'il n'y a pas de terme de volume exclu dans la relation d'état, et pas de dissipation visqueuse.

Phénomènes de relaxation et d'oscillation

Le mouvement « moyen » décrit par les modèles hydrodynamiques dans la limite de relaxation rapide est associé à la notion de convergence faible, cela signifie que l'on néglige tous les phénomènes physiques qui se passent sur des échelles spatio-temporelles plus petites que l'échelle d'observation, pourvu qu'ils ne perturbent pas le mouvement d'ensemble (même sur des temps longs).

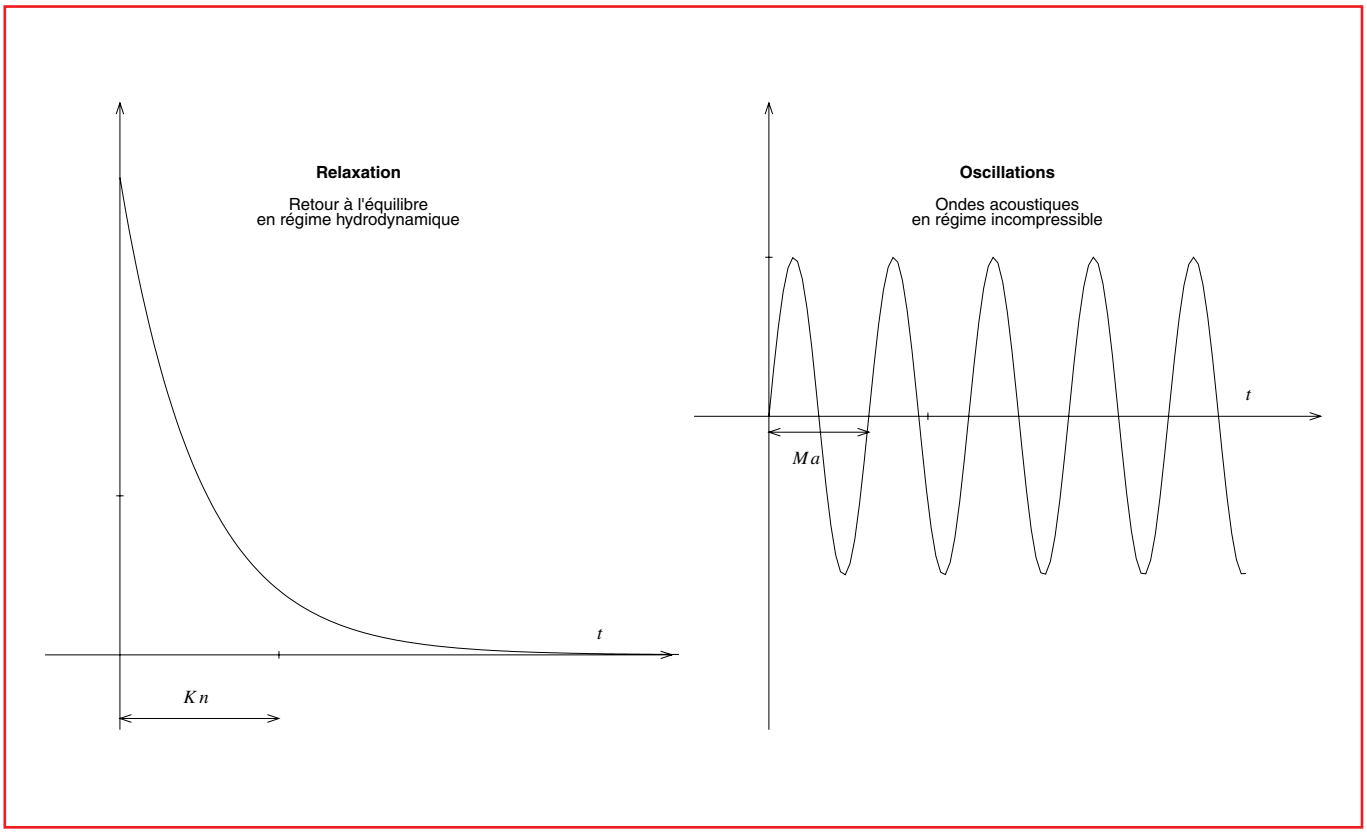


Figure 3 – *Corrections à l'approximation hydrodynamique.*

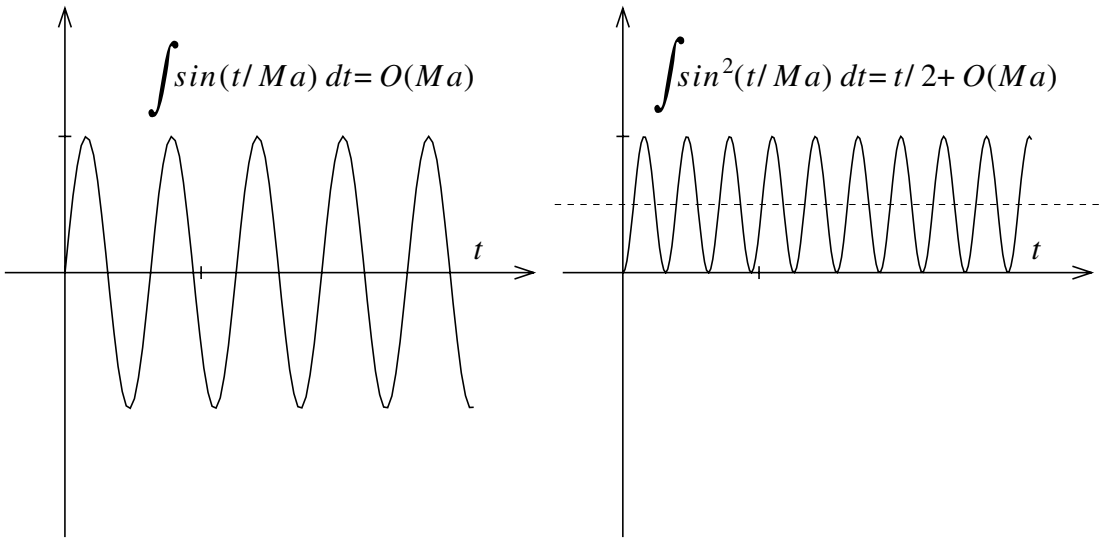


Figure 4 – *Modification du mouvement moyen par couplage d'ondes haute fréquence.*

La relaxation est un phénomène très localisé en temps et en espace : en remplaçant les distributions de vitesses par les équilibres thermodynamiques locaux correspondants, on commet essentiellement une erreur sur de fines couches spatiales ou temporelles, appelées couches limites. Toutefois cela peut entrainer une modification des données initiales ou des conditions aux bords à prendre en compte pour le mouvement macroscopique.

Les phénomènes oscillatoires haute fréquence ou à petite longueur d'onde n'apparaissent pas non plus dans l'approximation hydrodynamique car, en moyenne, ils n'induisent aucun déplacement du gaz. Ce point est beaucoup plus délicat à vérifier car les modèles fluides obtenus asymptotiquement ne sont pas linéaires, les différentes ondes sont donc couplées et il faut alors s'assurer qu'elles ne produisent pas d'interférences constructives. De plus, même dans le cas où les oscillations se découplent totalement du mouvement moyen, on peut être amené à en garder une description précise à cause de leurs propriétés de propagation : c'est par exemple le cas si on veut étudier les répercussions sonores du décollage d'un avion (étant bien entendu que les ondes acoustiques ne modifient pas le mouvement de l'avion !).

Vers une justification mathématique des développements multi-échelles

Etant donné les enjeux tant du point de vue de la modélisation que du point de vue de l'analyse des équations aux dérivées partielles, le sixième problème de Hilbert [4] (qui consiste à obtenir de façon rigoureuse une description multi-échelle de la dynamique des gaz) a suscité de nombreux travaux. Aujourd'hui le seul régime qui donne lieu à une étude asymptotique à peu près complète est celui qui conduit aux équations de Navier-Stokes incompressibles, dont la théorie est la mieux comprise.

Le schéma de preuve introduit par Bardos, Golse et Levermore [1] consiste à obtenir les équations du mouvement moyen, en passant à la limite dans les conservations locales de masse, impulsion et énergie associées à l'équation de Boltzmann. Le développement asymptotique de la distribution de vitesses $f \sim M_f + Kn\ (v \cdot \nabla_x) M_f + o(Kn) + o(Ma)$ (où M_f désigne la Gaussienne de mêmes moments que f) permet alors de décomposer chacun des termes de flux en un terme de convection (qui s'exprime comme une fonction non linéaire des grandeurs thermodynamiques locales), un terme de dissipation et des termes de reste.

Les principales difficultés consistent alors à obtenir un contrôle sur les particules de grandes vitesses (assurant qu'il n'y a pas de déperdition d'énergie asymptotiquement), et surtout à décrire précisément les oscillations et leurs éventuels couplages.

Lions et Masmoudi [5] ont montré que les oscillations temporelles, bien connues sous le nom d'ondes acoustiques, n'avaient pas de contribution au mouvement moyen par un argument de type compacité par compensation. Golse et l'auteur [3] ont ensuite établi un résultat de régularité spatiale sur les grandeurs thermodynamiques qui repose sur un argument de dispersion et un lemme de moyenne, et qui écarte toute possibilité d'oscillations spatiales. Les autres difficultés, beaucoup plus techniques, sont liées à la théorie de l'équation de Boltzmann, et notamment au concept très faible de solution introduit par DiPerna et Lions [2].

Pour en savoir plus

[1] Bardos (C.), Golse (F.) & Levermore (D.), Fluid Dynamic Limits of Kinetic Equations II : Convergence Proofs for the Boltzmann Equation, *Comm. Pure Appl. Math.*, **46** (1993), 667–753.

[2] DiPerna (R.J.) & Lions (P.L.), On the Cauchy Problem for the Boltzmann Equation : Global Existence and Weak Stability Results, *Annals of Math.*, **130** (1990), 321–366.

[3] Golse (F.) & Saint-Raymond (L.) The Navier-Stokes Limit of the Boltzmann Equation for Bounded Collision Kernels, *Invent. Math.*, **55** (2004), 81–161.

[4] Hilbert (D.) Sur les problèmes futurs des Mathématiques, *Congrès intern. des math., Paris 1900*, (1902), 58–114.

[5] Lions (P.L.) & Masmoudi (N.), From Boltzmann Equations to Navier-Stokes Equations I, *Arch. Ration. Mech. Anal.,* **158** (2001), 173–193.

Le charme *discret* des mathématiques

András SEBŐ*

Après une brève introduction aux mathématiques discrètes, le lecteur sera invité à suivre la solution simple mais surprenante d'un problème issu des télécommunications qui est resté ouvert pendant plus de vingt ans, et qui l'amènera peut-être à saisir certaines idées discrètes. D'autres exemples confirmeront que les mathématiques discrètes sont ancrées à la fois dans les mathématiques et les applications contemporaines. Nous espérons que les preuves inclues illustreront comment cette jeune discipline se forge ses propres méthodes pour servir à la fois la Beauté et l'Utilité.

Le monde est-il discret ?

« Discret » est ici le contraire de continu. C'est à peu près tout ce qu'on peut dire du terme lui-même. Avouons-le, ce mot ne correspond à aucune notion mathématique précise. Les nombres réels forment un ensemble « continu ». Les entiers, bien qu'infinis, sont *discrets*, c'est à dire « loins » les uns des autres.

Et le monde, est-il discret ? Oui ! Depuis que l'on sait que la matière est faite d'un nombre fini d'atomes (discrets) ... Et les mouvements continus peuvent être remplacés par de très petits mouvements discrets ...

N'est-il pas étonnant, quand on est écolier, que même l'eau soit constituée d'un nombre fini d'atomes ? Et pourtant, la plupart des théories mathématiques classiques traitent seulement « le continu », et sont souvent vues comme les sciences de l'infini. Les mathématiques discrètes veulent traiter de ce qui est discret (fini ?), mais trop complexe pour être traité « à la main », sans outils mathématiques.

On peut arriver au discret par l'approximation du continu d'une part. Mais la science moderne, et en particulier l'industrie automatisée, l'informatique, l'économie etc., produisent des problèmes qui sont dès le départ « discrets ».

Les problèmes à caractère discret peuvent souvent se modéliser d'une manière naturelle en terme de graphes. Mathématiques discrètes, théorie des graphes, ou combinatoire sont des synonymes pour certains, couvrent des notions différentes d'après d'autres, mais la partie commune de ces domaines est certainement très grande : peut-on rédiger un problème combinatoire qui ne peut pas être reformulé en termes de graphes ? Il est vrai aussi qu'il existe des méthodes plus « graphiques » que d'autres.

Rappelons tout d'abord qu'un graphe est une paire d'ensembles $G = (V, E)$ où V est un ensemble quelconque, et E est un sous-ensemble des paires d'éléments de V. Les éléments de V s'appellent les *sommets* du graphe et les éléments de E s'appellent *arêtes*. On dira aussi que $\{u, v\} \in E$ est une arête *entre* u *et* v : on peut représenter les sommets d'un graphe par des points et ses arêtes par des lignes qui les relient. Deux sommets $u, v \in V$ sont *voisins* si $\{u, v\} \in E$.

Voyons maintenant, à titre d'exemple, un des problèmes les plus classiques : le couplage. Il s'agit de marier (coupler) un nombre maximum de filles et de garçons en respectant leurs sympathies mutuelles, données à l'avance sous

* *Laboratoire Leibniz, UMR 5522 CNRS, 46 Avenue Félix Viallet, 38000 Grenoble.*
Andras.Sebo@imag.fr

forme d'un graphe. Quelles peuvent être les raisons pour lesquelles on ne peut pas marier plus de 184 filles lorsqu'il y en a 276 en tout ? Autrement dit, pourquoi peut-on être obligé de renoncer au mariage de 92 filles ? Une des raisons possibles : un sous-ensemble de filles rassemblant par exemple 231 d'entre elles, ne sympatisent en tout qu'avec 139 garçons. Déjà dans un tel sous-ensemble de filles, il y en aura au moins 92 qui ne seront pas mariées !

Appelons *déficit* d'un ensemble de filles la différence entre le nombre de ses éléments et le nombre de ses sympatisants (92 dans l'exemple) ; si ce nombre est négatif, le déficit est défini comme étant 0. Le déficit pourrait être considéré comme une mesure de « difficulté » d'une communauté de filles à se marier.

A priori, on pourrait imaginer qu'il y a d'autres raisons que le déficit pour empêcher le mariage des filles. La clef de ce problème est l'un des premiers résultats de la théorie des graphes, et son message est qu'il n'y a pas d'autre raison !

Théorème 1 [König-Hall, 1931]. *Le nombre minimum de filles non-mariées est égal au maximum du déficit des sous-ensembles de filles.*

Grâce à ce théorème, on peut facilement certifier à quelqu'un qu'un couplage est le plus grand possible : on lui montre un sous-ensemble de filles dont le déficit est égal au nombre de filles non mariées.

D'après Edmonds, l'existence d'un certificat d'optimalité est un bon indice pour que l'optimalité puisse être décidée d'une manière efficace. Malgré ce pari – dont la formulation précise est « $NP \cap coNP = P$? », similaire à la célèbre question « $P = NP$? » –, certains résultats laissent à penser que *trouver* pourrait être plus difficile que certifier. Par exemple, écrire un entier positif donné comme produit de deux entiers supérieurs à 1 pourrait s'avérer plus difficile que certifier l'existence d'une telle décomposition. En ce qui concerne les couplages max, on peut décider s'ils existent ou non, et on peut même les trouver. Ce théorème et l'algorithme efficace qui le montre (la « méthode hongroise », en train de fêter son 50.-ème anniversaire[1], s'exécute en temps « polynomial », voir LL, AS) fournissent à la fois le couplage optimal et un ensemble de filles de déficit maximum.

Le problème des mariages, ou couplages, est l'un des problèmes de base de la théorie des graphes, dont les variantes, spécialisations ou généralisations apparaissent partout en recherche opérationnelle et dans de nombreuses autres disciplines. Voici par exemple comment les couplages apparaissent dans un problème « d'ordonnancement de tâches » classique :

Etant données n tâches devant être exécutées sur deux machines identiques, déterminer le temps d'exécution minimum, si

– l'exécution de chaque tâche demande une minute ;

– des « contraintes de précédence » sont données entre les tâches.

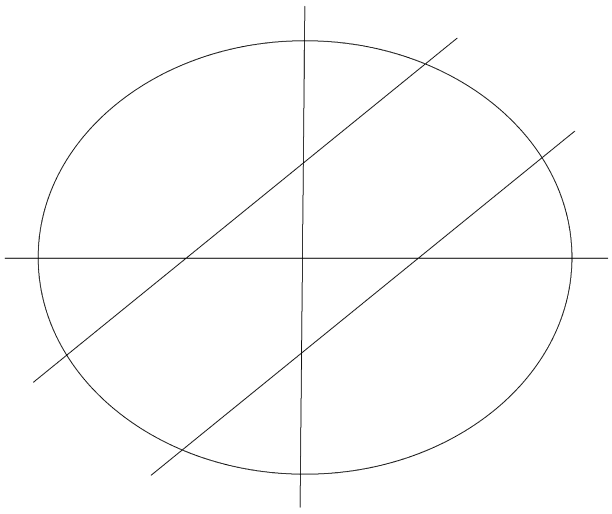


Figure 1 – *VLSI, lignes aériens, etc.*

1. www.cs.elte.hu/HungarianMethod50.

Est-ce que des chemins de précédence entre deux tâches forment le seul obstacle, c'est-à-dire la seule raison de ne pas pouvoir exécuter les tâches en moins de temps ?

On peut essayer de coupler le plus possible de tâches exécutables en même temps. Autrement dit « marier » les tâches, mais c'est un peu plus difficile, car on n'a pas une division naturelle en deux ensembles, « garçons » et « filles ». Le maximum k de paires de tâches exécutables en même temps peut quand même être déterminé (algorithme d'Edmonds). Comme on ne peut pas exécuter plus de 2 tâches en même temps on peut être sûr qu'au moins $n - k$ minutes sont nécessaires pour exécuter toutes les tâches. Miracle : il est vrai aussi (montré par Fujii, Kasami, Ninomiya en 1969), que $n - k$ minutes suffisent ! Pour trois machines déjà, ce problème est ouvert.

Encore quelques exemples – qui mènent d'une manière moins évidente aux couplages – juste pour montrer comment l'abstraction mathématique unifie des modèles qui ont l'air d'être très différents à un plus bas niveau. Les lignes (4 lignes droites et une circulaire) que vous voyez sur la Figure 1 peuvent représenter des connexions électriques. Le dessin doit être réalisé avec des fils électriques (en métal) en utilisant les deux côtés d'une plaque (chaque point sur exactement un des deux côtés). On doit faire attention à une et une seule chose : deux lignes ne peuvent pas se toucher, c'est-à-dire que deux fils qui se croisent doivent être sur des côtés différents de la plaque au point d'intersection.

Pour bien arranger correctement les lignes, on peut creuser des trous dans la plaque, et faire passer les fils d'un côté à l'autre, mais bien sûr, pas dans les croisements, car ceci causerait des faux contacts. Il est facile de voir, qu'on peut toujours réaliser un plan de connexion sans croisements en faisant suffisamment de trous. Mais les trous ne sont pas gratuits à faire, et rendent le système plus compliqué . . . Comment trouver le nombre minimum de trous ?

Les lignes peuvent aussi représenter des trajectoires d'avion qui doivent toujours voler à l'un des niveaux possibles parmi deux, leurs itinéraires ne pouvant pas se croiser (pas même à des temps différents). On doit alors minimiser le nombre de changements de niveaux (qui sont coûteux, inconfortables pour les passagers, dangereux pour la marchandise . . .).

Cette même Figure 1 peut encore être interprétée de multiples façons. Toute discipline mathématique essaiera de trouver des modèles généraux qui synthétisent les méthodes et regroupent les problèmes autour de ces méthodes. Les interprétations de la Figure 1 se rejoignent à un niveau plus général : ce sont des cas particuliers du problème de coupe maximum (coupe max) dans un graphe planaire, qui se résout avec des méthodes de la « théorie des couplages ».

« L'univers discret » est structuré, et les mathématiciens ont bien avancé dans la hiérarchie et les interrelations des classes de problèmes. Les résultats théoriques peuvent devenir de bons conseillers pratiques : ils arrivent souvent à cerner où se situent de nouveaux problèmes dans cette jungle ; ce faisant ils orientent les chercheurs vers des méthodes appropriées.

Notre pari est maintenant d'expliquer la solution complète d'un problème difficile.

La pluie

Que peut-on demander à des enveloppes mouillées ? Peut-on éviter, sinon avec des parapluies, que l'information qu'elles contiennent ne soit perdue ?

Il s'agit du problème de Shannon : trouver un code pour transmettre des messages sans ambiguëté, bien que certaines paires de lettres puissent être confondues.

Imaginez que vous deviez envoyer un message par l'intermédiaire d'un médium qui connait seulement les lettres (« a », « c », « u », « v », « w »). Quand les enveloppes sont mouillées, chaque lettre peut être confondue avec son prédécesseur et successeur dans cet ordre cyclique (le a avec le c et le w, le c avec le a et le u, etc.). Que devraient faire les facteurs sous la pluie ?

Juste ouvrir leur parapluie !

Les paires de lettres qu'on peut confondre sont représentées par les arêtes d'un graphe (Figure 2), qu'on va appeler *graphe des confusions*.

On supposera qu'on envoie des messages ayant tous la même longueur k fixée (nombre de caractères). Si par exemple « a » est l'espace, on peut toujours compléter à k caractères les messages plus courts. Deux messages peuvent être confondus, si toutes les paires de lettres qui sont à la même position dans les deux se confondent. On ne peut pas les confondre, *s'il existe deux lettres aux mêmes positions qui ne peuvent pas être confondues.*

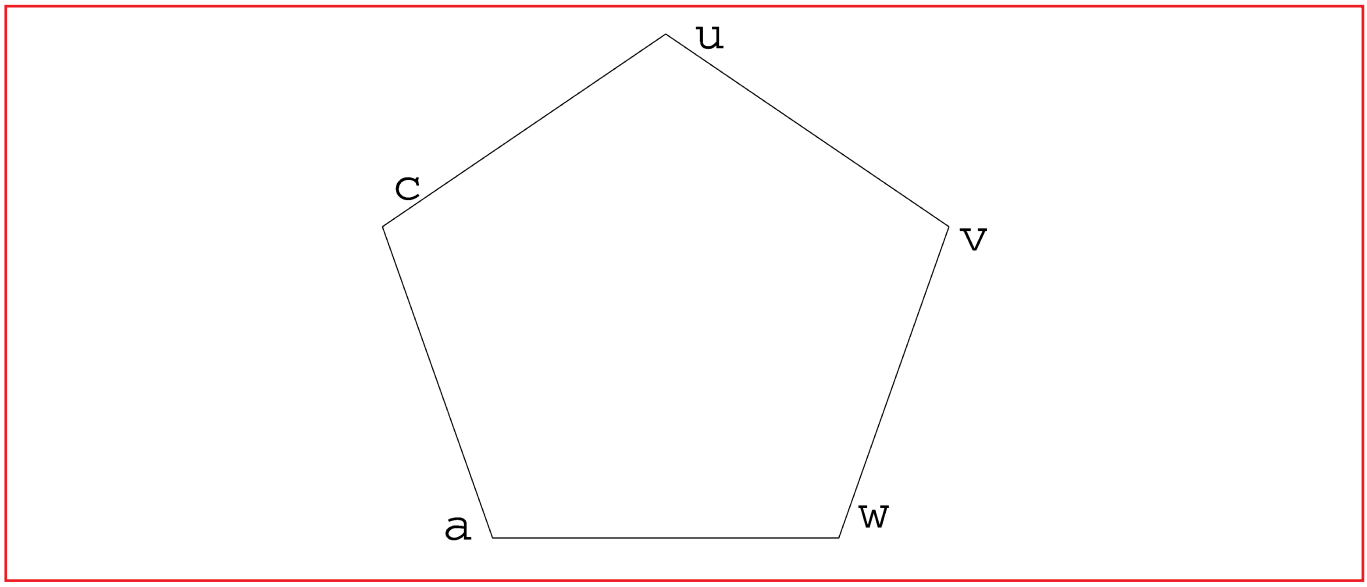


Figure 2 – *Graphe de confusion* C_5.

Si l'on pouvait confondre toutes les lettres les unes avec les autres, alors on ne pourrait pas envoyer plus d'un seul message quelle que soit sa longueur fixée. Ce ne serait même pas la peine de l'envoyer ! La « capacité de Shannon » vaudra 0. Quand on aura deux lettres qu'on ne peut pas confondre elle vaudra 1. Dans un alphabet de 32 lettres (comme l'alphabet hongrois) si aucune paire ne peut jamais être confondue (le graphe des confusions n'a pas d'arêtes), la « capacité de Shannon » vaudra 5. Ceci voudra dire que cet alphabet a la même capacité que les mots binaires de longueur 5. Pour coder n messages différents avec des mots binaires de longueur k, il faut avoir $2^k \geqslant n$, c'est-à-dire $k \geqslant \log_2 n$. En général, quand le graphe G des confusions est donné, la capacité de Shannon[2] va mesurer le nombre de caractères binaires (bits) équivalents à un de nos caractères ; autrement dit, c'est la longueur du mot binaire équivalent, divisé par la longueur de notre mot. La définition formelle :

La *capacité de Shannon* $\mathrm{cs}(G)$ d'un graphe G est définie comme la limite (k tend vers l'infini) du *logarithme en base de 2 du nombre maximum de messages de longueur k qu'on peut envoyer sans confusions, divisé par k.* La limite existe, car la série est (essentiellement) non-décroissante et bornée.

Dans notre premier exemple (Figure 2) le nombre de messages différents de longueur 1 parmi lesquels il n'y a pas de confusion est 2. Comme si on avait l'alphabet binaire à deux lettres, 0 et 1, sans confusion. Un « bit utile » par lettre. Mais on peut faire mieux avec des mots de deux lettres : en les représentant par un tableau 5×5 (Figure 3) on peut confondre deux mots si et seulement si leur représentation sur cet échiquier torique 5×5 est voisine par un côté ou par une diagonale. (La première et la dernière lignes sont voisines, idem pour les colonnes.)

Peut-on faire encore mieux (en moyenne) en augmentant la longueur des mots ? C'était la question posée par Shannon dans un article de 1956. Il a fallu attendre vingt deux ans, jusqu'en 1978 pour une réponse. (Par Lovász, [RR]). La question a beaucoup contribué au développement d'une portion très riche de la théorie des graphes.

Le parapluie

En d'autre termes, les deux lettres – par exemple a et v – qu'on ne peut pas confondre nous donnent 2^k mots de longueur k sans aucun risque de confusion, ce qui nous offre seulement 1 bit utile par lettre. D'après Figure 3 on a mieux : 5^k mots de longueur $2k$ (à la place de 4^k). En effet, pour $k = 1$ on a 5 mots (de longueur $2k$), et avec k répétitions on aura 5^k mots de longueur $2k$, ce qui équivaut à un message binaire de longueur $k\log_2 5$. En moyenne une de nos lettres équivaudra donc à $1/2\log_2 5 > 1$ lettres binaires (bits) ! Si on démontre qu'on ne peut pas faire mieux on aura :

2. La définition habituelle appelle « capacité de Shannon » 2^x où x est la capacité de Shannon d'après notre définition. Nous nous permettons cet abus ici car nous trouvons que cette terminologie plus logique rend les explications plus agréables.

Théorème 2 [Lovász, 1978]. $cs(C_5) = 1/2 \log_2 5$.

La méthode utilisée par Lovász a été à l'origine de méthodes fondamentales en optimisation combinatoire, et a eu des rebondissements récents. C'était la première application importante de la « programmation semidéfinie » en combinatoire :

La *représentation orthonormale* des sommets d'un graphe G, est une affectation de vecteurs unité de dimension d (d arbitraire) aux sommets, de façon à ce que pour toute paire de sommets non-adjacents de G, les vecteurs correspondants aux deux sommets soient perpendiculaires (produit scalaire égal à 0). On dit d'un sous-ensemble $S \subseteq V(G)$ des sommets d'un graphe qu'il est *stable*, si aucune paire d'éléments de S n'est une arête ; $\alpha(G)$ est la cardinalité d'un stable maximum de G.

Lemme 1. *Etant donnée une représentation orthonormale $r : V(G) \to \mathbb{R}^d$ ($r = (r_1,\ldots,r_d)$) d'un graphe G (où d est un entier quelconque), on a pour tout stable $S \subseteq V(G)$:*

$$\sum_{s \in S} r_1^2(s) \leqslant 1$$

En effet, les vecteurs associés aux sommets de S sont orthonormaux, donc on peut les compléter en une base orthonormale.

Dans cette base orthonormale, la somme des carrés des i-ièmes coordonnées des vecteurs pour tout $i = 1,\ldots,d$ est égal à 1 (la transposée d'une matrice orthonormale est aussi orthonormale). Si on restreint la somme à S on a l'inégalité voulue.

La Figure 4 montre la représentation 3-dimensionnelle de C_5 qui s'avérera optimale pour le Lemme.

Le parapluie du dessin est ouvert de façon à ce que les paires de baleines non-voisines soient perpendiculaires. Le lecteur se convaincra facilement que ceci est possible. On choisit la longueur des cinq arêtes égale à 1chacun. Il est clair, que les 5 vecteurs oc, ou, ov, ow, oa fournissent une représentation orthonormale de C_5.

En substituant le résultat de l'encadré dans le lemme, on obtient que la somme sur un stable quelconque de C_5 de la fonction constante $1/\sqrt{5}$ est inférieure ou égale à 1, et par conséquent, $\alpha(C_5) \leqslant \sqrt{5}$.

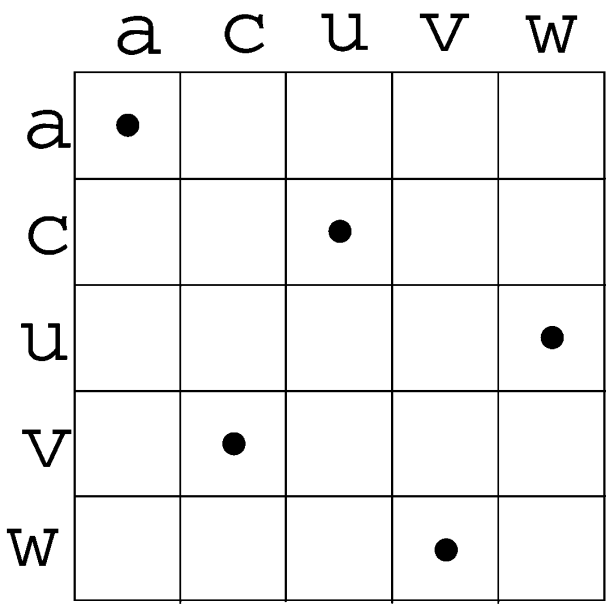


Figure 3 – *Cinq mots de deux lettres qu'on ne peut pas confondre.*

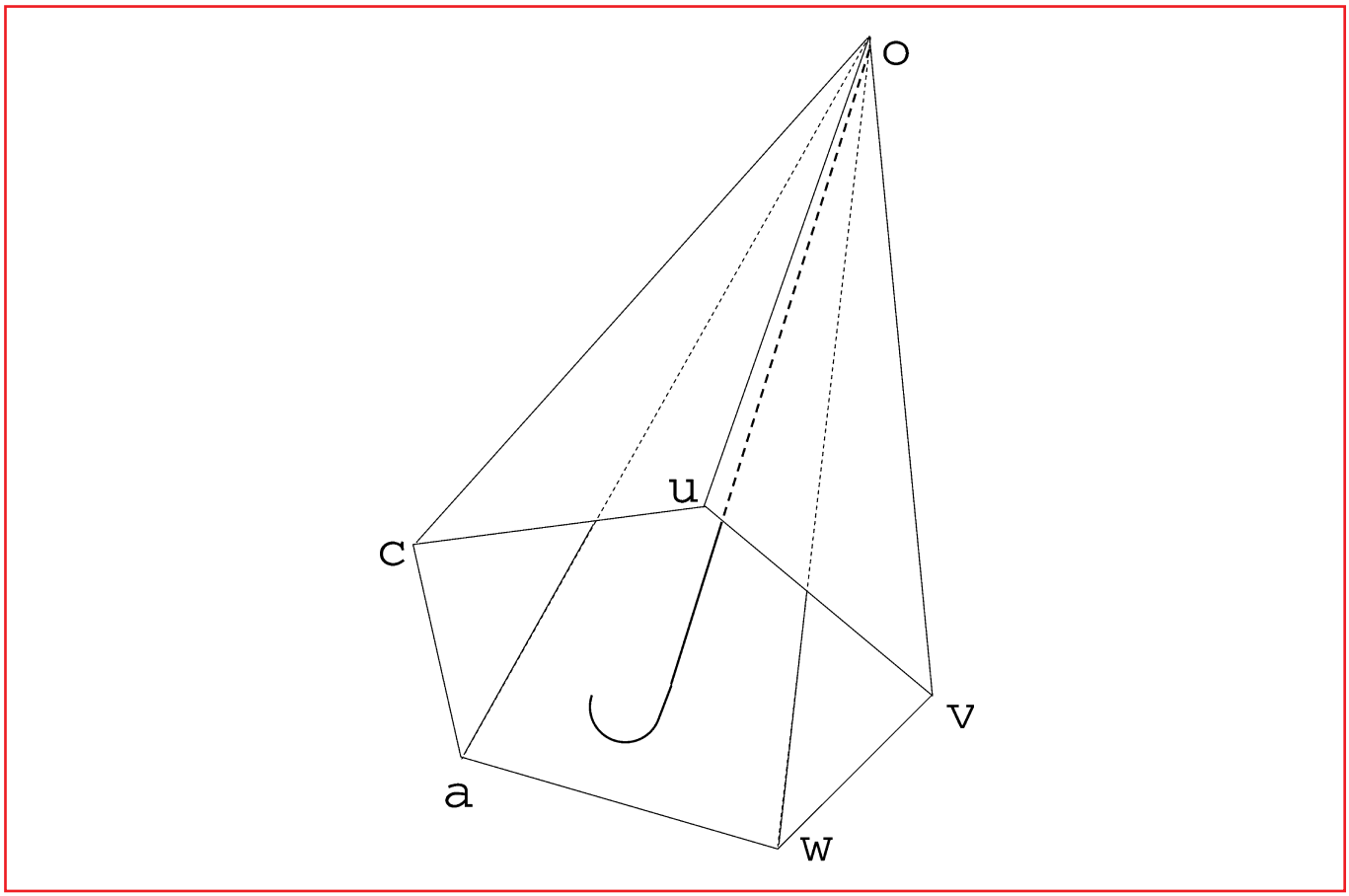


Figure 4 – *Le parapluie magique de Lovász.*

Pourquoi avoir travaillé tant pour aussi peu ? On sait très bien, que dans cette inégalité là, on pourrait remplacer $\sqrt{5}$ par 2 ! Parce que cette borne, même si inexacte pour 2, s'étendra facilement à des messages *arbitrairement longs*, et il se trouve qu'elle donnera la meilleure borne supérieure pour la moyenne du nombre de bits !

Définissons le graphe G^k dont les sommets sont les mots de longueur k construits à partir des sommets du graphe des confusions G, et avec une arête entre deux mots si on peut les confondre. Les stables de ce graphe sont donc les ensembles de k-tuples qui ne peuvent pas être confondus. On va utiliser le lemme pour démontrer que

$$\alpha(C_5^{2k}) \leqslant 5^k \qquad (*)$$

et ceci finit évidemment la preuve du théorème. On va définir une représentation orthonormale de C_5^k pour tout k. On en a déjà une pour C_5, et pour passer de C_5^k à C_5^{k+1}, on va appliquer la méthode suivante :

Si on a déjà une représentation de G^r et G^s, on représentera l'ensemble de leurs concaténations (les mots qu'on obtient en mettant l'un à côté de l'autre) – exactement G^{r+s} –, par le produit tensoriel des vecteurs associés. Le produit tensoriel $x \otimes y$ de $x \in \mathbb{R}^r$, $y \in \mathbb{R}^s$ n'est rien d'autre qu'une table de multiplication $r \times s$: une matrice $r \times s$ dont l'élément ij est $x_i y_j$, et que l'on regarde comme un vecteur, pour que par exemple le produit scalaire de deux produits tensoriels soit bien défini (comme la somme des produits des éléments correspondants).

Alors l'équation élémentaire

$$(x \otimes y)^T (w \otimes z) = (x^T w)(y^T z).$$

est vraie pour tout $x, w \in \mathbb{R}^r$, $y, z \in \mathbb{R}^s$. En effet, les deux côtés somment tous les termes de la forme $x_i y_j w_i z_j$ $i = 1, \ldots, r$, $j = 1, \ldots, s$.

Encadré 1

Calcul pour le parapluie

Calculons la représentation orthonormale de C_5 (Figure 4), pour la substituer dans le Lemme, afin d'obtenir la borne supérieure de la capacité de Shannon d'une manière élémentaire.

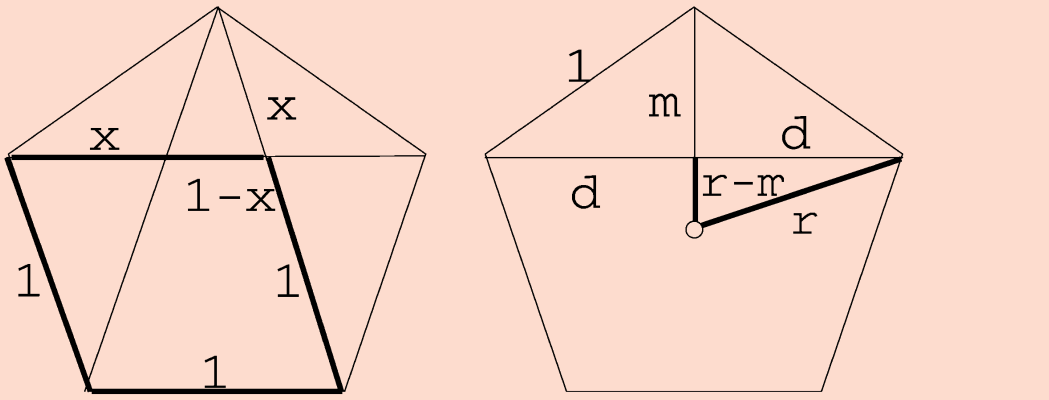
On fait ce calcul de lycée (il y a plus court, mais y a-t-il plus élémentaire) pour aller au bout de l'expérience : un calcul élémentaire avec un petit peu d'algèbre linéaire et de réflexion combinatoire peuvent collaborer d'une manière originale dans la solution d'un problème difficile ! Faites vérifier les calculs par vos enfants lycéens !

On effectue d'abord des calculs sur le pentagone régulier dont le côté est de longueur 1.

Le losange en gras de la Figure 5 a.) nous permet de trouver la longueur de la diagonale qui vaut $1+x$ *où* x *vérifie l'équation en dessous de la figure, d'où* $x = \frac{-1+\sqrt{5}}{2}$ *; la longueur de la diagonale est donc* $\frac{1+\sqrt{5}}{2}$*, et on va noter la longueur de la demi-diagonale par* $d = \frac{1+\sqrt{5}}{4}$*. La Figure 5 b.) montre que le rayon* r *peut aussi être calculé en appliquant à nouveau le théorème de Pythagore :* $m = \sqrt{1-d^2} = 1/4\sqrt{10-2\sqrt{5}}$*. Donc* $r = 1/(2m) = \frac{2}{\sqrt{10-2\sqrt{5}}} = \frac{\sqrt{2}}{\sqrt{\sqrt{5}}\sqrt{\sqrt{5}-1}}$.

Et maintenant passons au parapluie dont les baleines sont de longueur 1*, et qui est ouvert de façon à ce que chaque paire de baleines non-voisines forment un angle droit. Ces deux baleines forment alors un triangle rectangle isocèle dont les deux côtés perpendiculaires sont de longueur* 1*. La longueur de la diagonale du pentagone régulier formé par les extrémités des arêtes à l'ouverture appropriée du parapluie est donc de* $\sqrt{2}$ *au lieu de* $2d$ *et le rayon du cercle circonscrit de ce pentagone régulier est de* $\sqrt{2}/(2d)$ *fois la valeur* $\frac{\sqrt{2}}{\sqrt{\sqrt{5}}\sqrt{\sqrt{5}-1}}$ *du rayon calculé pour le pentagone régulier de la Figure 5 a.) :* $\frac{\sqrt{\sqrt{5}-1}}{\sqrt{\sqrt{5}}}$.

On obtient donc, par une application de plus du théorème de Pythagore, que la projection orthogonale des baleines du parapluie sur le manche est de longueur $1/\sqrt{\sqrt{5}}$*. Si on choisit alors cet axe comme « axe x », la première coordonnée de toutes les arêtes est* $1/\sqrt{\sqrt{5}}$.



a.) Thales : $(1-x)\ /\ 1$ b.) Pythagore : $r^2 - (r-m)^2 = d^2$

$2rm = d^2 + m^2 = 1$

Figure 5 – *Calculs sur* C_5.

Cette équation exprime exactement ce qui nous manque : *si on a une représentation orthonormale de* G^r *et de* G^s *et qu'on représente la concaténation de deux vecteurs par le produit tensoriel de leur représentation, alors on obtient une représentation orthonormale de* G^{r+s} !

Nous avons vu (encadré) une représentation orthonormale de C_5, où la première coordonnée de chaque représentant est $1/\sqrt{\sqrt{5}}$. Donc pour tous les vecteurs de la représentation de $G = C_5^k$ obtenue par produit tensoriel on a $r_1^2 = (1/\sqrt{5})^k$. En substituant ce résultat dans le lemme, on obtient que le nombre maximum de mots de longueur k qu'on ne peut pas confondre est inférieur ou égal à $(\sqrt{5})^k$, c'est à dire pour $2k$ la borne est 5^k, d'où (*). La conjecture de Shannon qui a été une cible bien connue des chercheurs pendant 22 ans, est démontré !

La perfection et au-delà

Bien avant que la capacité de Shannon de C_5 ait pu être calculée, le problème posé par Shannon a stimulé un chapitre de la théorie des graphes qui a été pendant longtemps l'un de ses moteurs et a eu un rayonnement bien au-delà. Ce chapitre est celui des *graphes parfaits*, définis par Claude Berge, qui a aussi énoncé deux conjectures devenues célèbres les concernant, sous les noms de conjecture faible et forte des graphes parfaits. (Voir les notes historiques de Berge, dans son dernier article [RR].)

Un ensemble de sommets mutuellement non-voisins d'un graphe est dit *stable* ; un ensemble de sommets mutuellement voisins est un graphe *complet*. Nous avons vu que $\alpha(G^k) \geqslant (\alpha(G))^k$. Rappelons la source de difficulté exprimée par $\alpha(C_5) = 2$, $\alpha(C_5^2) = 5 > 2^2$.

Quels sont les graphes pour lesquels ce phénomène ne se présente pas, c'est-à-dire $\alpha(G^k) = (\alpha(G))^k$? Pour ceux-là la capacité de Shannon est égal au nombre de stabilité. Nous laissons au lecteur le plaisir de montrer que $\alpha(G) = \chi(\bar{G})$ est une condition suffisante pour avoir cette propriété. On note par $\bar{G}$ le *complémentaire* d'un graphe G, c'est-à-dire le graphe dont les sommets u et v sont voisins si et seulement si ils ne sont pas voisins dans le graphe original ; $\chi(G)$ est le *nombre chromatique* de G, c'est-à-dire le nombre minimum de classes d'une partition des sommets en stables.

On a vu que C_5 n'a pas cette propriété. Il est facile de voir que les graphes « cycles impairs » et leurs complémentaires ne l'ont pas non plus. La conjecture célèbre de Berge [LL], [RR], [AS] énonce : *si à partir d'un graphe G on ne peut pas arriver à un de ces graphes en supprimant des sommets (et toutes les arêtes qui les contiennent), alors* $\alpha(G) = \chi(\bar{G})$. Une variante plus faible a été démontré par Lovász en 1972, et cette conjecture « forte » par Chudnovsky, Robertson, Seymour et Thomas en 2002. Colorier les sommets d'un graphe, ou reconnaître une sous-classe par un algorithme efficace (polynomial) reste un problème utile et intéressant : les recherches sur les graphes parfaits et leurs sous-classes gardent encore beaucoup de secrets [RR], dont Chudnovsky et Seymour ou Maffray et Trotignon continuent à révéler certains, même après la grande percée que représente le théorème fort. Un grand problème reste ouvert : pour les graphes parfaits le stable max, le nombre chromatique (et par conséquent la capacité de Shannon) se laissent calculer en temps polynomial en utilisant la programmation semidéfinie (faisant appel à la méthode des ellipsoïdes) mais il n'y a pas d'algorithme combinatoire !

Et qu'y a-t-il au-delà de la perfection ? L'imperfection ! Les graphes imparfaits ont été étudiés en grande partie pour résoudre la conjecture de Berge [RR], mais la preuve utilise peu ces résultats.

Cependant le monde discret n'est pas toujours complètement lisse, et il faut faire avec. On peut souvent remédier à l'imperfection – dans un sens plus général avec des méthodes non moins profondes.

Le problème du partitionnement des sommets d'un graphe en deux parties, A et B, de façon à maximiser le nombre, ou plus généralement la somme des poids, des arêtes qui sont « à cheval » entre les deux parties, est un des problèmes les plus utilisés de la théorie des graphes. On l'appelle *coupe max*. C'est un problème difficile (NP-difficile) et donc on ne peut pas espérer le résoudre « parfaitement ». Par une idée extrêmement originale de Goemans et Williamson, ce problème admet une solution « imparfaite », mais intéressante, et similaire aux parapluies. Nous allons l'esquisser pour montrer comment des idées similaires apparaissent dans des situations différentes.

Etant donnés des poids w_{ij} $(i,j = 1,\ldots,n)$ sur les paires de sommets d'un graphe (qu'on peut supposer être complet, c'est-à-dire toutes les paires de sommets sont des arêtes) on peut résoudre le problème de représenter les points sur la sphère de dimension n et maximiser

$$\sum_{i<j=1}^{n} \mathrm{dist}(x_i,x_j)^2 w_{ij}, \qquad (*)$$

où x_i, x_j représentent les sommets i, j, et $\mathrm{dist}(a,b)$ est la distance euclidienne entre a et b.

Ceci est un « programme semidéfini ». Le lecteur intéressé par l'algèbre linéaire pourra facilement démontrer que ce problème d'optimisation est équivalent à déterminer la matrice symétrique semidéfinie $X_{n\times n}$ dont les entrées sur la diagonale sont égales à 1, et pour laquelle

$$\sum_{i<j=1}^{n} X_{ij} w_{ij}$$

est minimal.

Un programme semidéfini est traitable (par des algorithmes polynomiaux) et ceci découle du fait que « la contrainte qu'une matrice soit semidéfinie » est une contrainte qui peut être traitée similairement à la non-négativité de la programmation linéaire.

Si nous choisissons la sphère de diamètre 1, *le maximum S de la somme dans (*) nous donnera une borne supérieure à la valeur maximum M d'une coupe,* (on peut placer les sommets dans deux points antipodaux de la sphère), d'autre part, *si dans (*) on remplace* $\mathrm{dist}(i,j)$ par l'angle $x_i O x_j$ où x_i, y_j représentent les sommets i et j, O est le centre de la sphère, et on mesure les angles en demi-tours (la mesure de l'angle de 180° est 1 et les autres angles ont une mesure proportionnelle) – on notera cette quantité $\mathrm{angle}(x_i,x_j)$ –, alors S devient S' et il se trouve qu'on peut trouver une coupe de taille S', de plus, S' n'est pas beaucoup plus petit que S, que nous prouvons :

Un hyperplan aléatoire qui passe par l'origine sépare les sommets en deux, définissant une partition. (La probabilité d'avoir des sommets sur l'hyperplan, est 0.) Cette partition sépare x_i et x_j avec probabilité égale à $\mathrm{angle}(x_i,x_j)$. Et donc l'espérance mathématique de la contribution de l'arc ij est $\mathrm{angle}(x_i,x_j)w_{ij}$. Par conséquent l'espérance mathématique de la coupe est juste S', donc il existe une coupe de taille au moins S' qu'on peut déterminer en temps polynomial avec la technique de la dérandomisation ! Il ne reste plus qu'à vérifier $0{,}878S \leqslant S' \leqslant S$, où on sait déjà $S' \leqslant S$, et le lecteur pourra à nouveau vérifier ceci avec le théorème de Pythagore et de la trigonométrie de lycée : on est de nouveau dans le plan de dimension 2 – défini par le centre de la sphère et x_i, x_j –, et on peut calculer $\mathrm{dist}(i,j)$ en fonction du sinus et cosinus de l'angle. Il est surprenant mais vrai que *l'angle divisé par la distance au carré reste toujours supérieur ou égal à 0,878*, vérifiez le avec votre calculette !

Une place ⟨trop⟩ discrète ?

La méthode des représentations géométriques similaires et l'application de la programmation semidéfinie sont maintenant choses courantes en optimisation combinatoire voir [RR]. D'autres idées géométriques ou algébriques ont montré leur puissance en théorie des graphes. Par exemple le nombre de Colin de Verdière – la multiplicité de la plus petite valeur propre positive d'une certaine matrice associée à un graphe – qui permet d'exprimer des représentations importantes et profondes (dans le plan ou d'une manière avantageuse en trois dimension) du graphe. La méthode utilise de l'algèbre linéaire et des arguments analytiques. Encore un autre des multiples exemples : Lovász, Bárány, Matoušek et récemment un jeune étudiant grenoblois Frédéric Meunier, utilisent des résultats de la topologie pour démontrer des théorèmes combinatoires qui ne se laissent pas aborder d'une manière élémentaire, (voir une introduction à ce sujet dans le joli livre de Matoušek).

A l'inverse, on voit de plus en plus que les autres domaines des mathématiques utilisent des méthodes ou des résultats combinatoires. Quelques exemples : la topologie (la théorie des noeuds utilise la théorie des graphes et des résultats de la combinatoire polyédrale), la géométrie algébrique (relations proches avec la programmation en nombres entiers), et les statistiques mathématiques (qui utilisent la théorie des structures régulières comme les « bloc design »), etc. Les relations tissent les liens entre mathématiques discrètes et les autres domaines des mathématiques. Beaucoup de jolis problèmes viennent aussi de l'informatique, d'autres sciences, de l'industrie, ou de l'économie (recherche opérationnelle), etc. Le domaine reste proche de ses applications, qui demeurent une ressource importante de problèmes.

Un autre domaine d'application mérite d'être mentionné : l'enseignement des mathématiques. Le discret est ludique, et fait bien travailler le cerveau même à un niveau de débutant : on y trouve des mini-théories prêtes a être enseignées dans le secondaire ; elles pourraient remplacer ou compléter certains enseignements classiques. Des collègues y travaillent (« maths à modeler », « math-en-jeans », etc), et avec beaucoup de succès auprès des élèves. Espérons que ce bruit d'enfants joyeux qui aiment les maths sans être nécessairement « matheux » arrivera bientôt aux adultes qui déterminent le programme des écoles.

Quelle est la place des mathématiques discrètes dans le paysage scientifique ? En Allemagne[3], en Hongrie[4], ou encore dans de nombreuses universités d'Amérique du Nord[5], elles ont une place consolidée, leur importance dans les formations universitaires modernes a été reconnue. En France cette place reste encore discrète. Le besoin de son enseignement – pour les informaticiens, ingénieurs, économistes, biologistes, etc. – est identifié, mais la nécessité d'une expertise importante n'est pas partout réalisée. Pourtant plusieurs équipes font une recherche appréciée au niveau international avec un nombre croissant de jeunes, dans le cadre de collaborations florissantes (exemple : le projet européen « Marie Curie training network »[6]), et pourraient fournir des enseignants spécialisés.

Il y a un quart de siècle, Lovász [LL] a écrit : « … It is often forcefully stated that Combinatorics is a collection of problems, which may be interesting in themselves but are not linked and do not constitute a theory … In my opinion, Combinatorics is growing out of this early stage. There are techniques to learn … There are branches which consist of theorems forming a hierarchy and which contain central structure theorems forming the backbone of study ... There are notions abstracted to many non-trivial results, which unify large parts of the theory … » Ceci s'est confirmé depuis, le domaine est devenu adulte, bien que toujours jeune et dynamique ; nous espérons que cette mini-photo condensée en donne une image déchiffrable.

Pour en savoir plus

[LL] Lovász (L.), *Combinatorial problems and exercises,* North Holland and Akadémiai Kiadó (1979).

[RA] Perfect Graphs, recueil d'articles édité par Ramirez-Alfonsin J., Reed B., Springer (2001).

[AS] Schrijver (A.), *Combinatorial Optimization*, 1-3, Springer (2003).

Je voudrais remercier Zoli Szigeti pour ses remarques pédagogiques, Nicolas Trotignon pour son aide professionnelle et linguistique consciencieuse et Sophie Sebő pour son aide stylistique. Le travail a été effectué dans le cadre du réseau ADONET, qui est un « Marie Curie Training Network ».

3. www.or.uni-bonn.de/index.eng.html.
4. www.cs.elte.hu/egres/
5. par exemple www.math.uwaterloo.ca/ CandO_Dept/
6. www.ads.tuwien.ac.at/adonet/sites.html

Images des formes, formes des images

Alain TROUVÉ*

Modéliser la variabilité des formes et faire des comparaisons quantitatives entre formes est indispensable pour tirer tout le partie des techniques modernes d'imagerie médicale. La participation des mathématiciens au vaste projet de l'anatomie numérique est sans doute une nécessité. Quelques jolis problèmes en perspective..

Un pionnier : D'Arcy Thompson

D'Arcy Thompson, naturaliste et mathématicien Ecossais, développe en 1917 dans son ouvrage « Growth and Forms » [Tho42] les bases d'une véritable « science de la forme » pour appréhender l'apparente infinie diversité des formes du vivant. Son idée centrale est de placer la physique et la géométrie (et donc les mathématiques !) au centre du dispositif pour réduire la variabilité des formes du vivant à quelques modèles simples sur lesquels agissent *des transformations géométriques* contrôlées par une certain nombre de contraintes physiques.

"In a very large part of morphology, our essential task lies in the comparison of related forms rather in the precise definition of each; and the deformation of a complicated figure may be a phenomenon easy of comprehension, though the figure itself may have to be left unanalyed and undefined. This process of comparison, of recognizing in one form a definite permutation or deformation of another, apart altogether from a precise and adequate understanding of the original 'type' or standard of comparison, lies within the immediate province of mathematics and finds its solution in the elementary use of a certain method of the mathematician. This method is the Method of Cordinates, on which is based the Theory of Transformations." D'arcy Thompson.

Cent ans plus tard, les idées de d'Arcy Thompson, après les formidables avancées de la biologie et de la génétique au XX^{eme} siecle, continuent d'être fécondes. La puissance des ordinateurs actuels rend maintenant possible le calcul explicite de telles transformations en 3D, leurs comparaisons quantitatives avec des retombées concrètes : un même organe sur des patients différents, ou sur un seul patient à des instants différents ou dans des positions différentes, est soumis à des variations de formes qu'il convient de modéliser et de caractériser, parfois parce qu'elles interviendraient dans un diagnostic (lorsque certaines variations sont susceptibles de correspondre à une pathologie) ou bien simplement parce qu'elles doivent être corrigées pour pouvoir analyser d'autres aspects des images dans une représentation normalisée. Le contexte de l'imagerie médicale n'est pas le seul domaine d'applications : la prise en compte de la variabilité géométrique des objets perçus est au coeur des problématiques de la vision artificielle.

D'autre part, les problèmes soulevés par la comparaison quantitative de formes et plus généralement par les espaces de formes ne cessent de stimuler l'imagination et les efforts des mathématiciens. Les premiers outils effectif sont venus avec des méthodes relativement simples mais efficaces venant de la théorie des splines et proposées par Bookstein

* *CMLA, 61, Avenue du Président Wilson*
94235 Cachan Cedex.
Itrouve@cmla.ens-cachan.fr

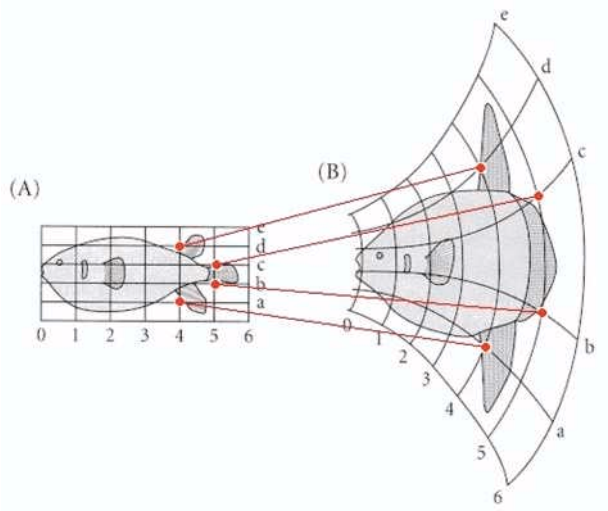


Figure 1 – *(D'après d'Arcy Thompson (1917)). Les formes en apparence différentes (A) et (B) n'en sont pas moins très similaires après application d'une transformation. Des points caractéristiques peuvent être mis en correspondance à travers une transformation simple.*

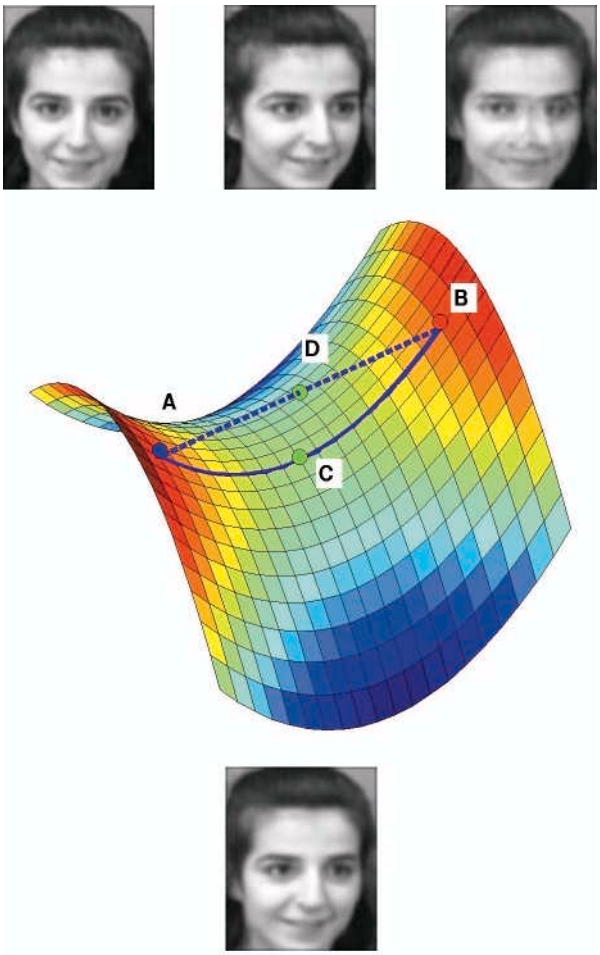


Figure 2 – *Qu'est-ce qu'une image moyenne entre A et B ? La moyenne arithmétique des intensités donne la réponse D. Une approche riemannienne des espaces de formes donne la réponse C. L'image C est issu d'un* calcul *d'un point milieu pour une certaine métrique dite des métamorphoses.*

[Boo91] : dans la comparaison de deux formes (A) et (B), on sélectionne des points caractéristiques $(M_i)_{1\leqslant i\leqslant n}$ sur (A) et $(M'_i)_{1\leqslant i\leqslant n}$ sur (B) puis l'on cherche une transformation permettant de faire coïncider les M_i avec les M'_i en déformant le support de (A) modélisé comme un matériau élastique. En termes mathématiques, il s'agit de résoudre le problème sous contraintes :

$$\left|\begin{array}{l}\inf |v|_V \\ \text{satisfaisant} \\ M_i + v(M_i) = M'_i, \text{ pour tout } i\end{array}\right. \qquad (1)$$

où v est un champ continu de déplacements et $|v|_V$ une norme hilbertienne donnant le « coût » du champ de déplacement v, le plus souvent la racine carrée d'une énergie de déformation inspirée des modèles d'élasticité linéarisée. Ceci peut être fait en résolvant un simple système linéaire.

Espace de formes

Les problèmes mathématiques posés par le codage et la comparaison de formes, la modélisation de la variabilité par des modèles aléatoires appropriés, s'avèrent être tous d'une très grande richesse et demande un véritable travail de conceptualisation pour pouvoir être abordés correctement [Mum03]. Par exemple, la question d'une forme de référence ou d'une forme moyenne entre deux formes oblige à envisager les espaces de formes comme des variétés riemanniennes mais encore faut-il savoir quelles sont les métriques naturelles (*cf* fig 2) et comment tenir compte simultanément des variations géométriques et photométriques. La construction de modèles aléatoires cohérents et les techniques d'estimation adaptées sont des problèmes peut-être encore plus redoutables. Beaucoup d'équipes de mathématiciens et d'informaticiens travaillent sur ces problèmes, plus ou moins proches des applications.

Actions de groupes

Pour illustrer plus concrètement les difficultés que l'on peut rencontrer, et les nouveaux objets que l'on est amené à manipuler, prenons un exemple sur lequel nous avons pas mal travaillé : les méthodes linéaires de types splines

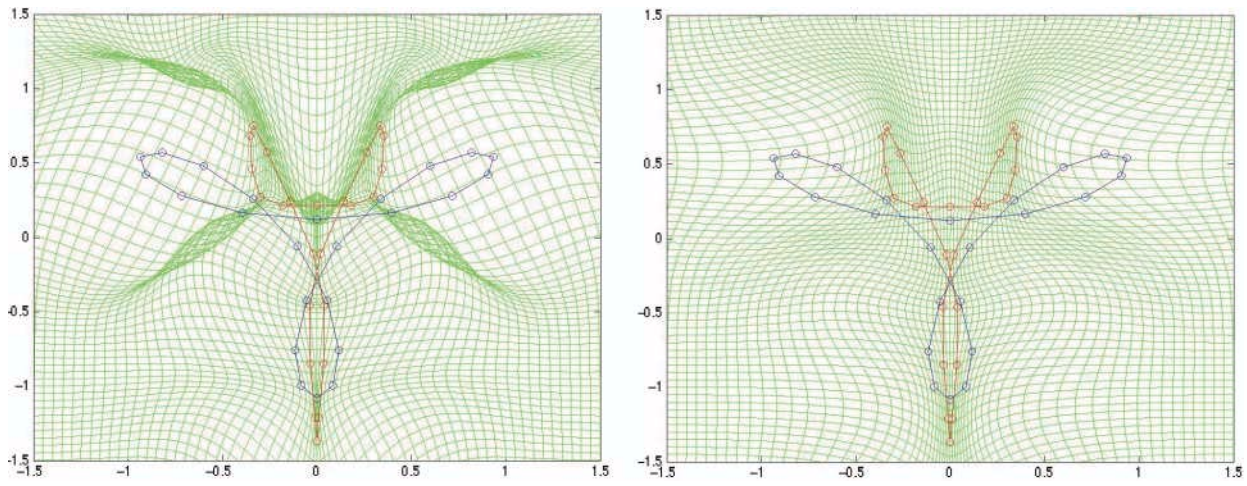


Figure 3 – *A gauche, l'interpolation par splines « à la Bookstein ». La forme (A) correspond à la courbe bleue et la forme (B) à la courbe rouge. Les points caractéristiques sont signalés par des petits cercles. Pour visualiser la transformation, on affiche en vert l'image par la transformation estimée d'une grille régulière. On note des zones de repliements avec perte d'inversibilité de $x + v(x)$ (ces zones apparaissent beaucoup plus foncées sur l'image de gauche). A droite, même problématique, mais on affiche la solution géodésique dans le groupe de difféomorphismes associé (voir plus bas). Les repliements sont éliminés et l'inversibilité préservée.*

ont connues et connaissent toujours un très grand succès mais rencontrent des problèmes lorsque les déformations deviennent importantes car les transformations générées ne sont plus inversibles (fig 3). Il devient alors impossible d'apparier de façon bijective les points de (A) et (B) et donc de passer d'une forme à l'autre. Pour résoudre ce problème technique de non inversibilité, il faut revenir aux idées de D'arcy Thompson : le coeur de son approche est de comparer les formes en considérant les transformations agissant sur quelques modèles simples. Les transformations ont naturellement une structure de groupe agissant (ou sens mathématique) sur les formes ou une représentation de celles-ci (par des points caractéristiques par exemple ou une image en niveau de gris). Cette vision des formes définie dans un langage moderne par des actions de groupes sur des éléments primitifs ou des générateurs est au coeur de la théorie de la reconnaissance des formes de Ulf Grenander [Gre93]. Les besoins de la comparaison de formes anatomiques demandent de considérer des groupes de difféomorphismes c'est-à-dire des groupes de transformations régulières et inversibles. Plus encore, il faut être capable de « facturer » la taille d'un difféomorphisme à la manière de $|v|_V$ qui donne le coût de la transformation $x \to x + v(x)$. L'enjeu est alors de construire des structures riemanniennes sur des groupes de difféomorphismes, et de pouvoir travailler numériquement avec, pour mesurer des distances entre difféomorphismes (et en déduire des distances entre formes) ou pour construire des géodésiques dans ces groupes reliant deux d'entre eux.

En reprenant la théorie des splines, si v est un « petit » champ de déplacement, $\psi(x) = x + v(x)$ est un difféomorphisme proche de l'identité (autrement dit, les champs de vecteurs définissent l'algèbre de Lie). En composant suffisamment de petits difféomorphismes, $\psi_i(x) = x + v_i(x)$, on construit par composition $\phi_n = \psi_n \circ \cdots \psi_1$ qui peut être loin de l'identité pour un coût $\sum |v_i|_V$. Cette idée est mathématiquement plus naturelle après passage à la limite pour laquelle la longueur d'une « courbe » indexée par $t \in [0,1]$ de difféomorphismes définie par l'équation d'évolution

$$\begin{cases} \dfrac{\partial \phi}{\partial t}(t,x) = v(t,\phi(t,x)),\ t \in [0,1] \\ \phi(0,x) = x \end{cases} \tag{2}$$

a pour longueur $\int_0^1 |v(t,.)|_V dt$. Les difféomorphismes sont donc définis à partir de flots de champs de vecteurs et $v(t,x)$ est la vitesse d'un point se trouvant en x à l'instant t tandis que $\phi(t,x)$ donne la position au temps t d'une particule en x au temps 0. L'ancienne métrique des splines $|\ |_V$ est réutilisée pour mesurer des coûts instantanés. Le problème de la construction d'une transformation inversible entre deux formes (A) et (B) à partir de points caractéristiques se traduit maintenant comme le problème

$$\begin{array}{l} \inf \int_0^1 |v(t,.)|_V \\ \text{satisfaisant} \\ \phi(1,M_i) = M_i' \ \forall i \text{ où } \phi \text{ solution de (2).} \end{array} \tag{3}$$

Cette fois, la solution sera un difféomorphisme tout en respectant les contraintes comme on peut le voir fig 3. De nombreuses interprétations géométriques sont possibles : La courbe de difféomorphismes $\phi(t,.)$ obtenue en résolvant (3) est une géodésique particulière entre l'identité et $x \mapsto \phi(1,x)$ pour la métrique riemannienne associée au choix initial de $|\ |_V$. On est alors conceptuellement très proche du cadre géométrique d'Arnold pour Euler incompressible puisqu'il s'agit ici aussi d'étudier les géodésiques sur le groupe des difféomorphismes pour une métrique invariante à droite. Dans le cas d'Arnold, la métrique de départ sur les champs de vecteurs est $|v|_V = ||v||_2$ (métrique L^2) avec la contrainte supplémentaire d'incompressibilité (div $(v) = 0$). Dans les cas utiles pour l'analyse des formes, la condition d'incompressibilité n'est pas naturelle et les métriques intéressantes doivent être plus « rigides » que la métrique L^2 (et même H^1 qui conduit à l'équation de Camassa-Holm) et surtout *dépendent* des objets d'intérêt. Cette rigidité qui évite l'apparition de solutions singulières rend sans doute les modèles plus simples, mais la nécessité de considérer une grande variété de métriques et de construire explicitement des géodé-

siques particulières solutions de nouveaux problèmes variationnels ouvrent de nombreux champs d'exploration à côté du cadre habituel de la mécanique des fluides [HRTY04].

La problématiques précédente correspond à l'action des difféomorphismes sur des n-uplets de points caractéristiques, mais l'extension à d'autres actions sont possibles comme celles sur les images : $(\phi, I) \rightarrow I \circ \phi^{-1}$, ou sur des sous-variétés de $\mathbb{R}^3$ grâce à la représentation des surfaces par mesures ou mieux encore par courants sur lesquels il existe une action géométrique naturelle [Gla05].

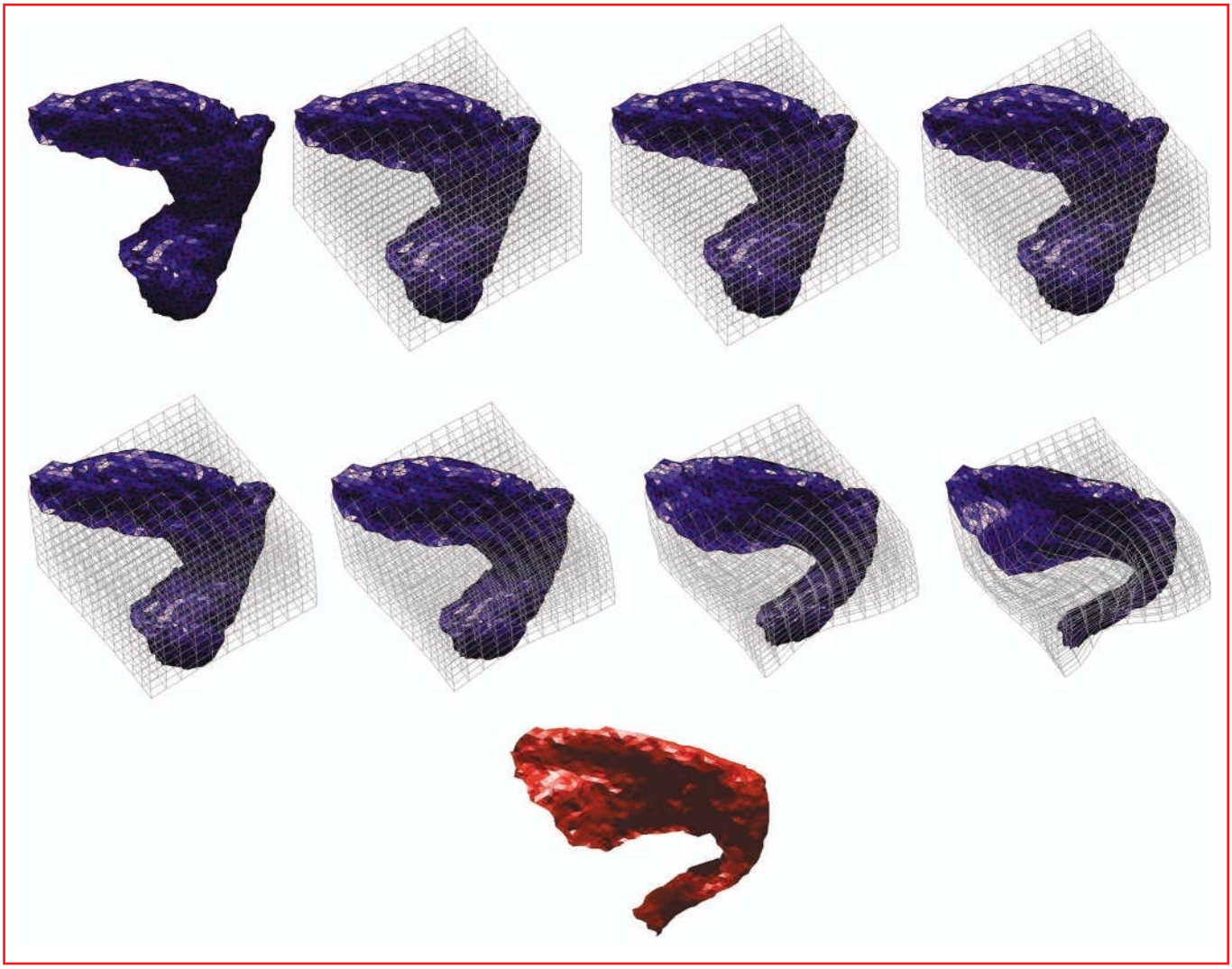


Figure 4 – *Application des techniques d'appariement géodésique pour la comparaison de surfaces corticales (ici le planum temporale qui est une structure bilatérale impliquée dans le langage et l'audition chez l'homme). De telles surfaces 3D sont appariées par une transformation inversible 3D de l'espace. La forme (A) est en haut à gauche, la forme (B) (en rouge) en bas à droite. La séquence représente des étapes intermédiaires dans le processus de transformation.*

Pour en savoir plus

[Boo91] BOOKSTEIN (F.L.), *Morphometric Tools for Landmark Data*, Cambridge University Press, (1991).

[Gla05] GLAUNES (J.), *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique*, PhD thesis, Université Paris 13, (2005).

[Gre93] GRENANDER (U.) *General Pattern Theory*, Oxford Science Publications, (1993).

[HRTY04] HOLM (D.D.), RATNANATHER (J.T.), TROUVÉ (A.) and YOUNES (L.), Soliton dynamics in computational anatomy, *Neuroimage*, 23 : 170–178, (2004).

[Mum03] MUMFORD (D.) The shape of objects in two and three dimensions : Mathematics meets computer vision, (2003), AMS Josiah Willard Gibbs Lecture.

[Tho42] D'ARCY THOMPSON, *On growth and Form*, Cambridge University Press, (1942), Première édition (1917).

[TY05] TROUVÉ (A.) and YOUNES (L.), Metamorphoses through lie group action, *Foundations of Computational Mathematics*, 5(2) : 173–198, (2005).

Géométrie et Dynamique des Surfaces Plates

Marcelo VIANA*

L'étude des surfaces plates est pleine de beaux objets et de belles idées et, malgré son caractère élémentaire, possède des relations profondes avec plusieurs autres domaines des Mathématiques. Cet article est une introduction rapide au sujet et à quelques résultats récents.

Surfaces plates

Le sujet de cet article est l'étude de la géométrie des surfaces lorsqu'elles sont munies d'une *métrique plate*. Commençons par expliquer cette notion à partir d'un cas concret : le cube (Figure 1). D'autres exemples, plus intéressants, apparaîtront par la suite. Du point de vue topologique, le cube est équivalent (homéomorphe) à la sphère « ronde » représentée dans la Figure 2. Mais du point de vue géométrique ces deux surfaces sont très différentes.

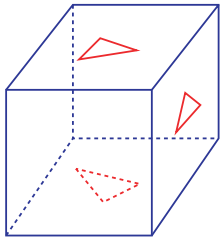


Figure 1 – *Le cube est un modèle plat de la sphère.*

Encadré 1

Surfaces

Par surface *on entend ici une variété compacte sans bord de dimension* 2. *On ne considère d'ailleurs que des surfaces orientables. Rappelons que ces objets sont classifiés par leur genre : deux surfaces compactes orientables sont homéomorphes si et seulement si elles ont le même genre. C'est le cas du cube et de la sphère (genre égal à zero).*

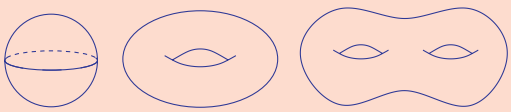


Figure 2 – *Surfaces (non-plates) de genres* $g = 0,1,2$.

* *Unité Mixte Internationale CNRS-IMPA (UMI 2924).*
Est. Dona Castorina 110, 22460-320 Rio de Janeiro, Brésil.
viana@impa.br

Il est clair que la courbe la plus courte (*géodésique*) reliant deux points sur une même face du cube est le segment de droite défini par ces points. De plus, la notion de droite a un sens même pour des courbes qui traversent des arêtes : il suffit de ramener ce cas à une situation plate en « dépliant » l'arête, comme décrit dans la Figure 3. Le fait que les géodésiques soient les segments de droite reste alors vrai aussi pour les points appartenant à des faces différentes.

On appelle *triangle* sur une surface un domaine borné par trois géodésiques. Ce qui caractérise le cube comme surface plate est le fait que la somme des angles internes de ses triangles est, généralement, égale a π. Ceci est, clairement, vrai pour les triangles contenus dans une face, mais aussi pour ceux qui traversent une arête du cube : il suffit de déplier l'arête comme nous l'avons expliqué avant. Par contre, il est bien connu que sur la sphère ronde la somme des angles internes des triangles est toujours supérieure à π, correspondant au fait que la courbure de la surface est positive partout.

Cela ne veut pas dire qu'une surface plate comme le cube soit dépourvue de courbure. En effet, le théorème de Gauss-Bonnet (voir encadré) implique que quand on déforme une sphère ronde en un cube, la courbure totale reste constante : elle est juste concentrée sur certaines régions qui, à la limite, donnent lieu aux sommets du cube. Voir Figure 4.

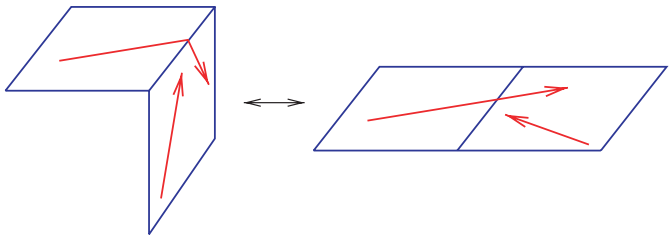


Figure 3 – *Une géodésique qui traverse une arête devient une droite quand on déplie l'arête.*

Encadré 2

Théorème de Gauss-Bonnet

Le fameux théorème de Gauss-Bonnet affirme que la courbure totale d'une surface lisse S ne dépend que de son genre, et pas de sa métrique : si on note par κ la courbure, alors

$$\int_S \kappa = 2\pi\chi(S)$$

où $\chi(S) = 2 - 2g(S)$ est la caractéristique d'Euler de la surface. Dans le cas de la sphère $\chi(S) = 2$ et donc, la courbure totale est 4π. Il y a une version du théorème de Gauss-Bonnet pour les surface plates, qui sera utile par la suite :

$$\sum_{i=1}^{N} \big(2\pi - \mathrm{ang}(V_i)\big) = 2\pi\chi(S), \qquad (1)$$

où $V_1, \ldots, V_N$ sont les sommets de la surface. C'est-à-dire que $2\pi - \mathrm{ang}(V_i)$ mesure la courbure qui est concentrée à chaque V_i. Par exemple, le cube a $N = 8$ sommets, dont les angles sont toujours $3\pi/2$, alors que sa caractéristique d'Euler est égale à 2.

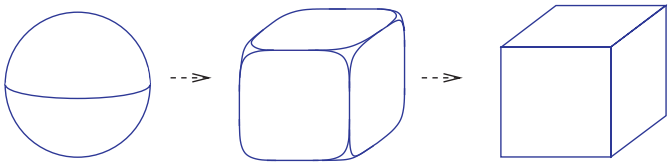


Figure 4 – *Les sommets sont des singularités de la métrique correspondant à des concentrations infinies de la courbure.*

Ceci suggère que pour les triangles qui contiennent une singularité la somme des angles internes doit être différente de π, et c'est effectivement le cas. On appelle *angle* d'une singularité V d'une surface plate, noté $\mathrm{ang}(V)$, la somme des angles de faces qui lui sont adjacents. Par exemple, l'angle de chacun des sommets du cube est $3\pi/2$. La construction dans la Figure 5 (on aplatit un voisinage de la singularité quitte à le « déchirer » le long d'une arête) montre alors que la somme des angles internes d'un triangle contenant V est donnée par

$$\alpha + \beta + \gamma = 3\pi - \mathrm{ang}(V). \qquad (2)$$

Dans le cas présent, cela donne $3\pi/2$. Mais on vérifie aisément que la relation (2) est valable en général, pas seulement pour le cube.

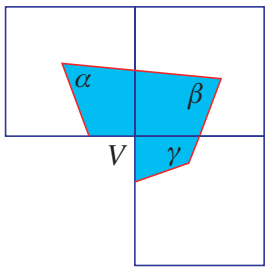


Figure 5 – *Angles internes :* $\alpha + \beta + \gamma + \mathrm{ang}(V) + \pi = 4\pi$.

Flots géodésiques

On s'interesse alors au problème suivant. Considérons des segments géodésiques partant d'un point générique sur la surface plate, dans une direction fixée. Quel est le comportement de ces segments quand la longueur va vers l'infini, comment se déploient-ils autour de la surface ? En particulier : *Quand les géodésiques sont-elles des courbes fermées ? Quand sont-elles denses dans la surface ? Peut-on décrire leur comportement asymptotique de façon quantitative ?*

Bien que formulées dans le language de la géométrie différentielle, ces questions sont aussi motivées par des problèmes dans plusieurs autres domaines des Mathématiques : l'analyse complexe (différentielles quadratiques, espaces et flots de Teichmüller), la topologie (feuilletages mesurés), la théorie des nombres (développements en fractions continues) et, bien sûr, la dynamique (échanges d'intervalles, billards polygonaux, exposants de Lyapounov). Il ne nous est pas possible dans ce court article d'exploiter ces importantes connections. Mais le lecteur pourra en trouver des présentations détaillées dans les beaux travaux d'A. Zorich mentionnés dans notre liste de références.

Tel que nous l'avons formulé, le problème du flot géodésique est à présent trop général pour qu'on puisse lui donner une réponse satisfaisante. Par la suite, nous allons restreindre un peu notre classe de surfaces plates. Cette restriction sert, essentiellement, à garantir que les géodésiques qui commencent dans une même direction restent

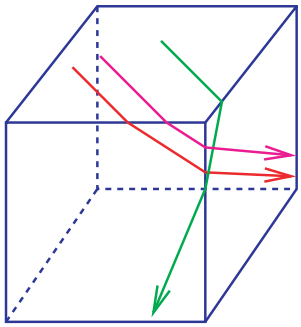


Figure 6 – *Les singularités peuvent rendre le flot géodésique très « chaotique ».*

toujours parallèles ; la Figure 6 montre que ce n'est pas le cas pour le cube, par exemple. Sous cette condition, nous verrons que le problème admet une réponse remarquablement précise. De plus, une bonne partie des motivations mentionnées ci-dessus ne nécessite que ce cadre un peu plus restreint.

Surfaces de translation

On considère un polygone dans le plan dont les côtés sont arrangés en paires telles que les deux segments de chaque paire soit parallèlle et ait la même longueur. On obtient alors une surface plate en identifiant les deux côtés dans chacune de ces paires. Les géodésiques sur cette surface sont des segments de droite ; à chaque fois qu'un segment atteint un coté du polygone, on le prolonge dans la même direction et le même sens à partir du point correspondant dans le coté dual du polygone (Figure 7).

Encadré 3

Tore plat et bitore plat

L'exemple le plus simple correspond au cas où le polygone est un carré. En identifiant les côtés opposés du carré, on obtient alors un tore plat. Le comportement des géodésiques sur cette surface est bien connu : celles dont la pente est rationnelle sont des courbes fermées ; celles dont la pente est irrationnelle sont denses et, même, uniformément distribuées dans le tore.

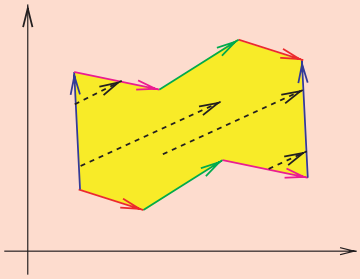


Figure 7 – *Surface de translation définie par un octogone.*

La Figure 7 décrit un autre exemple de cette construction. Il est facile de voir que tous les sommets de cet octogone donnent lieu à une seule singularité V de la surface plate quand on fait les identifications des différents côtés. Il est également clair que l'angle de cette singularité est égal à la somme des angles internes de l'octogone : $\mathrm{ang}(V) = 6\pi$. *On peut alors utiliser le théorème de Gauss-Bonnet (2) pour vérifier que le genre de cette surface est* $g = 2$ *: il s'agit donc d'un bitore plat.*

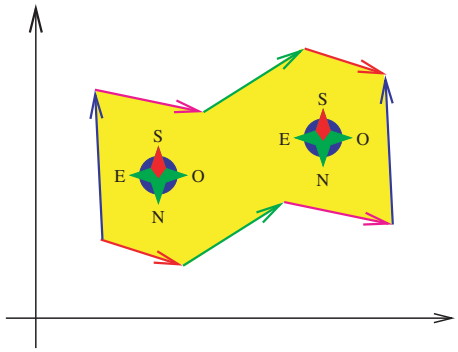


Figure 8 – *La rose-des-vents est définie sur la toute la surface de translation*

Les surfaces plates obtenues de la façon que nous venons de présenter, que l'on appelle *surfaces de translation*, ont la propriété additionnelle suivante : on peut définir sur toute la surface un champ de vecteurs unitaire localement constant[1] : la direction « Haut » (ou « Sud ») de la Figure 8. Ce champ de vecteurs se prolonge même aux singularités, en général de façon multivaluée.

Cycles assymptotiques

Le premier pas pour décrire le comportement assymptotique des géodésiques est une représentation des segments géodésiques sous la forme de vecteurs à coefficients entiers. Formellement, étant donné un long segment géodésique, on le referme en reliant le point final au point initial par une courbe, plus ou moins arbitraire, de longueur bornée. On interprète alors la courbe fermée ainsi obtenue comme un élément du premier groupe d'homologie de la surface. Mais cette procédure peut aussi être décrite de la façon géométrique qui suit.

Supposons que le polygone soit borné par $d \geqslant 2$ paires de côtés, numerotées $1, 2, \ldots, d$. Etant donné un (long) segment géodésique γ, on considère les sommets A et B du polygone les plus proches des points initial et final de γ et qui soient identifiés à une même singularité de la surface quotient. On relie alors A et B au segment γ par des segments géodésiques de longueur bornée. La courbe ainsi obtenue se projette donc sur une courbe fermée dans la surface de translation. Ensuite, on considère un chemin $\hat{\gamma}$ en « zigzag » reliant A à B le long des côtés du polygone, comme dans la Figure 9. On définit alors le vecteur $H(\gamma) = (h_1, \ldots, h_d)$ où h_i est le nombre de fois (avec orientation) qu'un côté portant le numéro i est parcouru par le chemin $\hat{\gamma}$. La définition ne dépend pas du choix de $\hat{\gamma}$.

Il est clair que quand on prend des segments géodésiques γ de plus en plus grands le vecteur $H(\gamma)$ croit aussi. Il est alors naturel de normaliser et de passer à la limite

$$c_1 = \lim_{|\gamma| \to \infty} \frac{1}{|\gamma|} H(\gamma) \tag{3}$$

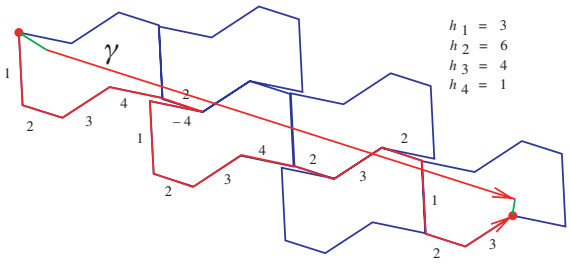


Figure 9 – *Représentation vectorielle d'un segment géodésique.*

1. Champ de vecteurs *parallèle*, dans le language de la Géométrie Riemanniene.

quand la longueur $|\gamma|$ va vers l'infini. Cette notion est due à S. Schwartzman, qui l'introduisit et appela *cycle asymptotique*, voilà un demi-siècle.

Le résultat spectaculaire qui suit donne alors une description très précise, au niveau qualitatif/topologique, du comportement des géodésiques typiques sur toute surface de translation :

Théorème 1 [S. Kerckhoff, H. Masur, J. Smillie]. *Pour toute surface de translation et pour presque toute direction, les géodésiques sont denses et, même, uniformément distribuées dans la surface, et le cycle asymptotique est bien défini et ne dépend pas du point de départ.*

Une version un peu plus faible, valable pour *presque toute surface* avait été obtenue quelques années auparavant, dans des travaux indépendants de H. Masur et W. Veech. Il vaut la peine d'ajouter que la conclusion n'est pas valable pour *toute* direction ; en fait, pour un ensemble dense de directions il y a des géodésiques fermées. Les géodésiques fermées des surfaces de translation furent étudiées récemment par A. Eskin, H. Masur, A. Zorich.

Conjecture du drapeau assymptotique

Vers le début des années 90, A. Zorich décida d'étudier la convergence (3), à l'aide d'un ordinateur. Il découvrit ainsi que les déviations des vecteurs $H(\gamma)$ par rapport à c_1 ont un comportement assez surprenant : *la composante de $H(\gamma)$ dans la direction orthogonale au cycle asymptotique se distribue plutôt dans une direction favorite c_2, et son amplitude maximale est une puissance $|\gamma|^{\nu_2}$ de la longueur*, avec $\nu_2 < 1$. Ce comportement est illustré dans la Figure 10.

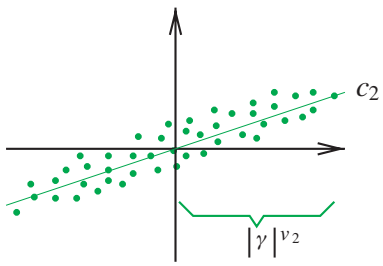


Figure 10 – *Phénomène de Zorich.*

De plus, les déviations de second ordre, c'est-à-dire, les composantes de $H(\gamma)$ dans la direction orthogonale au plan défini par c_1 et c_2 ont le même type de comportement : elles se distribuent dans une direction favorite c_3, et leur amplitude maximale est $|\gamma|^{\nu_3}$ avec $\nu_3 < \nu_2$. De même pour toutes les déviations jusqu'a l'ordre $g = g(S)$: finalement, la composante de $H(\gamma)$ dans la direction orthogonale au sous-espace défini par $c_1, \ldots, c_g$ est bornée, indépendamment de la longueur du segment géodésique γ. Ces observations furent formalisées en la

Conjecture [Zorich-Kontsevich]. *Il existe des vecteurs linéairement indépendants $c_1, c_2, \ldots, c_g$ et des nombres $1 > \nu_2 > \cdots > \nu_g > 0$ tels que*

1. *la composante de $H(\gamma)$ dans la direction orthogonale au sous-espace L_g engendré par $c_1, \ldots, c_g$ est bornée ;*
2. *l'amplitude de la composante de $H(\gamma)$ orthogonale au sous-espace L_i engendré par $c_1, \ldots, c_i$ est $|\gamma|^{\nu_{i+1}}$, pour tout $i = 1, \ldots, g-1$.*

Le flot de Teichmüller

Les travaux de Kontsevich et Zorich ont proposé une explication de ce phénomène surprenant, en termes du comportement d'un système dynamique qui agit dans l'espace des surfaces de translations : le flot de Teichmüller.

Encadré 4

Flot de Teichmüller

Ce flot est très facile à décrire au niveau des polygones : le temps-t du flot est l'opération (illustrée dans la Figure 11) qui consiste à dilater le polygone dans la direction horizontale et à le contracter dans la direction verticale, d'un même facteur e^t. Mais il faut garder en tête que ce flot est défini dans l'espace des surfaces de translation : la distinction est importante parce qu'une même surface de translation peut-être représentée par différents polygones. En fait, alors que l'action sur les polygones est triviale, le flot dans l'espace des surfaces de translation a une dynamique très riche. En particulier, d'après H. Masur et W. Veech, il est ergodique par rapport à une mesure de volume naturelle. Veech a même montré que cette mesure est uniformément hyperbolique : dans notre language, ceci revient à dire que $\nu_2 < 1$.

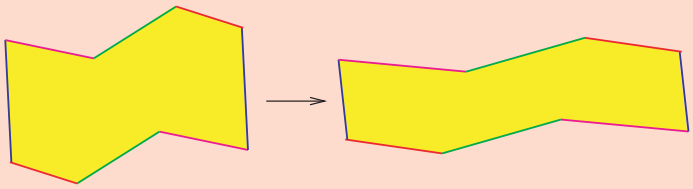


Figure 11 – *Flot de Teichmüller.*

Plus généralement, A. Zorich et M. Kontsevich ont montré que $\nu_2, \ldots, \nu_g$ sont directement liés aux exposants de Lyapounov du flot de Teichmüller. La preuve de la conjecture revenait alors a démontrer que le spectre de Lyapounov est simple, c'est-à-dire que $1 > \nu_2 > \cdots > \nu_g > 0$.

Récemment, G. Forni développa des méthodes analytiques et géométriques puissantes pour montrer que $\nu_g > 0$. Ceci contient le cas $g = 2$ de la conjecture et prouve aussi l'existence du sous-espace L_g dans le cas général. Encore plus récemment, des méthodes issues des systèmes dynamiques et de la théorie ergodique nous ont permis, à A. Avila et moi-même de démontrer le contenu complet de la conjecture.

Théorème 2 [Avila, Viana]. *La conjecture de Zorich-Kontsevich est vraie.*

En guise d'épilogue, mentionnons que les exposants $\nu_2, \ldots, \nu_g$ demeurent entourés de mystères. Des calculs réalisés par M. Kontsevich et A. Zorich les ont amenés à conjecturer que la somme

$$1 + \nu_2 + \cdots + \nu_g$$

est toujours un nombre rationnel. M. Kontsevich a même obtenu des formules analytiques pour ces sommes. Des progrès plus récents, surtout par M. Kontsevich et A. Zorich, ont conduit à des formules plus explicites, à travers lesquelles on peut espérer résoudre cette conjecture. A suivre.

Pour en savoir plus

[AV05a] Avila (A.),. Viana (M.), Simplicity of Lyapunov spectra: A general criterion, Pre-publication IMPA et Jussieu (2006).

[AV05b] Avila (A.),. Viana (M.), Simplicity of Lyapunov spectra: Proof of the Zorich-Kontsevich conjecture, Pre-publication IMPA et Jussieu (2005).

[EMZ03] Eskin (A.), Masur (H.), Zorich (A.), Moduli spaces of abelian differentials: the principal boundary, counting problems, and the Siegel-Veech constants, *Publ. Math. Inst. Hautes Etudes Sci.*, 97:61-179 (2003).

[For02] Forni (G.), Deviation of ergodic averages for area-preserving flows on surfaces of higher genus, *Ann. of Math.*, 155: 1-103 (2002).

[KMS86] Kerckhoff (S.), Masur (H.), Smillie (J.), Ergodicity of billiard flows and quadratic differentials, *Ann. of Math.*, 124: 293-311 (1986).

[KZ01] Kontsevich (M.), Zorich (A.), Connected components of the moduli spaces of Abelian differentials with prescribed singularities, *Invent. Math.*, 153:631-678 (2003).

[Mas82] Masur (H.), Interval exchange transformations and measured foliations, *Ann. of Math*, 115:169-200 (1982).

[Vee82] Veech (W.), Gauss measures for transformations on the space of interval exchange maps, *Ann. of Math.*, 115:201-242 (1982).

[Yoc05] Yoccoz (J.-C.), Continued fraction algorithms for interval exchange maps: an introduction, In *Frontiers in Number Theory, Physics and Geometry*, Vol 1: On random matrices, zeta functions and dynamical systems, Ecole de Physique des Houches, France, 2003, Springer-Verlag (2006).

[Zor99] Zorich (A.), How do the leaves of a closed 1-form wind around a surface ? *Pseudoperiodic Topology*, volume 197 of *Amer. Math. Soc. Transl. Ser. 2*, pages 135-178. Amer. Math. Soc., (1999).

[Zor05] Zorich (A.), Flat surfaces, In *Frontiers in Number Theory, Physics and Geometry*, Vol 1: On random matrices, zeta functions and dynamical systems, Ecole de Physique des Houches, France, 2003, Springer-Verlag (2006).

Je remercie Pierre Py et Anton Zorich d'avoir lu cet article et fait plusieurs suggestions, et Etienne Ghys de m'avoir invité à joindre cette édition des Images des Maths.

Enumération de fractions rationnelles réelles

Jean-Yves WELSCHINGER*

Le corps des réels n'est pas algébriquement clos, et par conséquent le nombre de solutions d'un système d'équations polynomiales à coefficients réels dépend en général fortement du choix des coefficients. Toutefois, lorsque ce système s'interprète géométriquement comme un problème de géométrie énumérative réelle, il est parfois possible de compter ses racines en fonction d'un signe ± 1 de façon à extraire un invariant à valeurs entières, indépendant des données du problème. Après avoir rappelé quelques problèmes classiques de géométrie énumérative, nous allons mettre ce phénomène en évidence.

Introduction

Chacun sait que par deux points distincts du plan passe une droite et une seule. Choisissez cinq points dans le plan, il y passera une conique. Une seule même, pour peu que ces cinq points ne soient pas en position trop spéciale. En voici la raison. Les coniques sont les lieux d'annulation des polynômes de deux variables X, Y – correspondant aux deux coordonnées x, y du plan – et de degré deux. Ces polynômes sont combinaisons linéaires des cinq monômes X^2, Y^2, XY, X, Y et du monôme constant unité. Notons $(x_1, y_1), \ldots, (x_5, y_5)$ les coordonnées des cinq points choisis dans le plan. Un polynôme $P(X,Y) = aX^2 + bY^2 + cXY + dX + eY + f$ s'annule en ces cinq points dès que ses six coefficients satisfont les cinq équations $ax_i^2 + by_i^2 + cx_iy_i + dx_i + ey_i + f = 0, 1 \leqslant i \leqslant 5$. Un tel système linéaire homogène a toujours une solution non nulle, unique même, à multiplication par un scalaire près, dès que ces cinq équations sont indépendantes, ce qui est le cas pour presque tous les quintuplets du plan. D'où le résultat ! Résultat bien plus général d'ailleurs.

Définition 1. *On appelle courbe algébrique plane de degré $d \in \mathbb{N}^*$ le lieu d'annulation d'un polynôme de deux variables X, Y de degré d.*

Un tel polynôme est combinaison linéaire des $\dfrac{d(d+3)}{2}$ monômes $X^k Y^l$, $1 \leqslant k + l \leqslant d$, et du monôme constant unité. De la même manière donc, par $\dfrac{d(d+3)}{2}$ points du plan passe toujours une courbe algébrique plane de degré d qui est en général unique. De plus, lorsque cette dernière n'est pas unique, ce sont une infinité de courbes algé-

* *Ecole normale supérieure de Lyon, Unité de mathématiques pures et appliquées*
UMR CNRS 5669
46, allée d'Italie, 69364, Lyon cedex 07.
jwelschi@umpa.ens-lyon.fr

briques planes de degré d qui relient les $\frac{d(d+3)}{2}$ points donnés. Ainsi, l'énumération de courbes algébriques planes de degré d contraintes à relier un nombre fini de points du plan que l'on vient de réaliser s'est ramenée à un problème linéaire, facilement résolu. Le fait que ces courbes sont définies implicitement par des équations polynomiales a joué un rôle important. S'il s'agit par contre d'énumérer des courbes définies explicitement par des polynômes ou fractions rationnelles, le problème est tout autre, comme nous allons le voir dans ce qui suit...

Enumération de courbes rationnelles complexes

Prenez trois polynômes complexes P, Q, R d'une variable et de degré $d \in \mathbb{N}^*$. Ils définissent une application dans le plan complexe par les relations $x = \frac{P}{R}(z)$, $y = \frac{Q}{R}(z)$, application définie en dehors des racines de R dans la droite complexe.

Définition 2. *L'image d'une telle application est appelée courbe rationnelle plane de degré $d \in \mathbb{N}^*$.*

D'avoir choisi le même dénominateur R dans les fractions rationnelles a pour effet que chaque droite du plan intersecte la courbe en au plus d points, à moins d'être incluse dans cette courbe. Une droite générique intersecte même cette courbe en exactement d points, ce qui justifie qu'on lui attribue un « degré d ». Les fractions rationnelles $\frac{P}{R}$ et $\frac{Q}{R}$ fournissent un paramétrage de la courbe rationnelle. Ce dernier n'est pas unique, puisqu'il peut être composé à droite par n'importe quelle homographie de la droite complexe, c'est-à-dire n'importe quelle application définie comme le quotient de deux polynômes de degré un. Ainsi, si les fractions rationnelles $\frac{P}{R}$ et $\frac{Q}{R}$ dépendent de $3d+2$ paramètres, les courbes rationnelles planes de degré d, elles, n'en dépendent qu'au plus de $3d-1$. En fait, elles dépendent exactement de $3d-1$ paramètres, ce qui amène le problème énumératif suivant. Choisissez $3d-1$ points génériques dans le plan complexe, il n'y a qu'un nombre fini de courbes rationnelles planes de degré d qui passent par ces points. Ce nombre N_d ne dépend pas du choix générique des $3d-1$ points, essentiellement parce que $\mathbb{C}$ est algébriquement clos. Quelle est la valeur de N_d ? En degrés un et deux, ce problème énumératif ne diffère pas du précédent, de sorte que $N_1 = N_2 = 1$. En effet, les droites et les coniques sont toutes rationnelles. Un paramétrage rationnel de ces dernières peut être obtenu comme application inverse d'une projection stéréographique (voir l'encadré 1). Par contre, dès le degré 3, ces problèmes diffèrent, et la dimension $3d-1$ des courbes rationnelles planes est d'ailleurs plus petite que la dimension $\frac{d(d+3)}{2}$ des courbes algébriques de degré d. Une manifestation géométrique de ce fait est également présentée dans l'encadré 1. La valeur $N_3 = 12$ s'obtient sans trop d'effort. La valeur $N_4 = 620$ fut obtenue par Zeuthen au XIX$^{\text{ème}}$ siecle. Il aura fallu attendre le début des années 90 pour connaître la valeur de N_5 et de toute la suite $(N_d)_{d\in\mathbb{N}^*}$, grâce à la découverte par Maxim Kontsevich de la formule de récurrence suivante.

Théorème 3.

$$N_d = \sum_{k+l=d} N_k N_l \left(k^2 l^2 C_{3d-4}^{3k-2} - k^3 l C_{3d-4}^{3k-1}\right), \quad d \geqslant 2$$

En particulier, $N_5 = 87304$, $N_6 = 26312976$, cette suite croît de façon extrêmement rapide de sorte qu'asymptotiquement, $\log(N_d)$ soit équivalent à $3d\log(d)$. Tous ces invariants énumératifs se déduisent donc finalement du seul $N_1 = 1$. Ce fait remarquable a été établi dans le cadre général de la théorie des invariants de Gromov-Witten dont je ne présente ici qu'un aspect très particulier, voir [KM].

Enumération de courbes rationnelles réelles

Supposons à présent les polynômes P,Q,R à coefficients réels. Les relations $x = \frac{P}{R}(z)$, $y = \frac{Q}{R}(z)$ définissent une application de la droite réelle privée des racines de R dans le plan réel. Elles définissent également comme précédemment une application u de la droite complexe privée des racines complexes de R dans le plan complexe, application qui commute cette fois-ci avec les conjugaisons complexes de la droite et du plan. L'image réciproque $u^{-1}\mathbb{R}^2)$ contient $\mathbb{R}$ privé des racines de R bien entendu, mais également en général un nombre fini de paires de points complexes conjugués. Si $u(z) = (z^2, z^3 + \epsilon z)$ avec $\epsilon > 0$ par exemple, alors $u^{-1}(\mathbb{R}^2) = \mathbb{R} \cup \{\pm i\sqrt{\epsilon}\}$. Ces paires de points complexes conjugués apparaissent à l'image comme des points isolés dans le plan réel. En fait, la complexifiée de la courbe rationnelle vient intersecter transversalement le plan réel en ces points. Notons $m(C)$ ce nombre de paires de points que l'on appelera *masse* de la courbe rationnelle réelle. A nouveau, par $3d-1$ points en position générale dans le plan réel passe un nombre fini de courbes rationnelles réelles. On note $\underline{x} = (x_1, \ldots, x_{3d-1})$ la configuration de $3d-1$ points choisie, et $\mathcal{R}_d(\underline{x})$ l'ensemble fini de courbes rationnelles réelles associé. Cette fois-ci toutefois, ce nombre de courbes rationnelles réelles dépend en général du choix, même générique, de la configuration de points $\underline{x}$. En effet, interpoler $3d-1$ points du plan s'interprète comme $3d-1$ équations à coefficients réels que doivent satisfaire les coefficients des polynômes P,Q,R. Si l'on perturbe la configuration de points, cela entraîne une perturbation des équations, et puisque $\mathbb{R}$ n'est pas algébriquement clos, cela entraîne en général un changement du nombre de racines réelles de ces équations. Néanmoins, on a le (voir [W1])

Théorème 4.

La quantité $\chi^d(\underline{x}) = \sum_{C \in \mathcal{R}_d(\underline{x})} (-1)^{m(C)}$ *ne dépend pas du choix générique de la configuration de points* $\underline{x}$.

Ici donc, le simple fait de compter ces courbes réelles en fonction d'un signe ± 1 permet de dégager un invariant à valeurs entières, noté χ^d, de ce problème de géométrie énumérative réelle. Le cardinal de $\mathcal{R}_d(\underline{x})$ se trouve alors borné supérieurement et inférieurement, d'après le

Corollaire 5.

$$|\chi^d| \leqslant \#\mathcal{R}_d(\underline{x}) \leqslant N_d$$

Ce cardinal peut-il prendre toutes les valeurs comprises entre ces deux bornes et de la même parité que χ^d et N_d ? En degré 3, oui, en degré supérieur, personne ne le sait ! D'ailleurs, si $\chi^3 = 8$, $\chi^4 = 240$ et $\chi^5 = 18\,264$, la valeur de χ^d

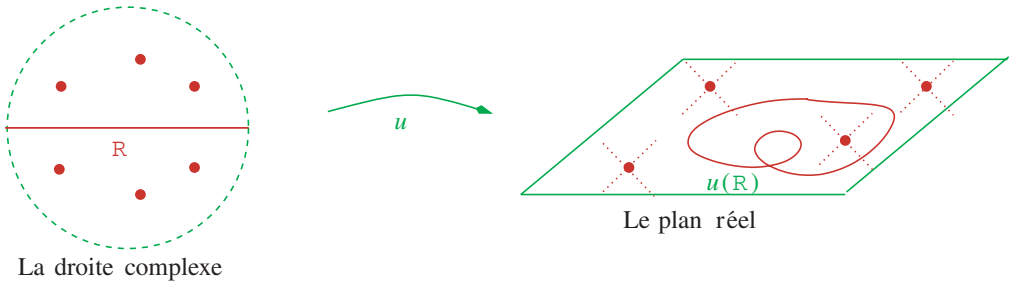


Figure 1 – *Une courbe rationnelle réelle.*

n'est pas connue en degré supérieur. Toutefois, outre ces premières valeurs, la minoration $\chi^d \geqslant \frac{1}{2}d!$ ainsi que l'asymptotique $3d\log d$ de $\log|\chi^d|$ ont été obtenues par Ilia Itenberg, Viatcheslav Kharlamov et Evgenii Shustin par l'intermédiaire de ce que l'on appelle « la géométrie tropicale », voir [IKS1], [IKS2]. Remarquons que dans le choix de la configuration $\underline{x}$, des paires de points réels peuvent être échangées par des paires de points complexes conjugués. Il s'agit alors d'imposer à la complexifiée de la courbe rationnelle d'interpoler les $3d-1$ points $x_1,\dots,x_{3d-1}$. L'entier $\chi^d_r(\underline{x}) = \sum_{C\in\mathcal{R}_d(\underline{x})}(-1)^{m(C)}$ est à nouveau indépendant du choix de la configuration $\underline{x}$, dès lors que l'on fixe le nombre total de points réels r de la configuration. L'entier précédemment noté χ^d est donc ici noté χ^d_{3d-1}, et il est agréable d'introduire la fonction génératrice $\chi^d(T) = \sum_{r=0}^{3d-1}\chi^d_r T^r \in \mathbb{Z}[T]$, où l'on a posé $\chi^d_r = 0$ lorsque r n'a pas la parité de $3d-1$. Cette apparition d'invariants entiers dans des problèmes de géométrie énumérative réelle est une découverte récente, et peu de telles bornes inférieures ont été mises à jour dans de tels problèmes. Toutefois, des résultats analogues ont été obtenus lorsqu'au lieu d'imposer aux courbes rationnelles réelles de passer par le dernier point x_{3d-1}, on les contraint à être tangentes à une courbe lisse générique du plan $\mathbb{R}^2$. Un autre exemple analogue s'obtient dans le problème classique de géométrie énumérative suivant : combien de coniques sont tangentes à cinq coniques génériques du plan ? Dans le cas complexe, 3 264, comme fut établi par de Joncquières en 1859. Dans le cas réel, le cardinal de l'ensemble $\mathcal{C}on(L)$ des coniques réelles tangentes à cinq coniques réelles génériques $L_1,\dots,L_5$ dépend à nouveau du choix de $L = L_1\cup\ldots\cup L_5$. Il est néanmoins possible d'extraire comme précédemment un invariant entier de ce problème de géométrie énumérative réelle, et d'en déduire par exemple les bornes inférieures $32 \leqslant \#\mathcal{C}on(L) \leqslant 3\,264$ lorsque les cinq coniques $L_1,\dots,L_5$ bordent cinq convexes disjoints du plan, voir [W2]. Si je ne mentionne pas ici de résultats plus précis dans ces deux problèmes, c'est qu'il ne suffit plus de compter les solutions du problème en fonction d'un signe ±1 pour obtenir un invariant à valeurs entières. Il faut cette fois-ci étudier simultanément plusieurs problèmes de géométrie énumérative réelle, et c'est seulement en combinant les comptages dans ces différents problèmes que l'on définit un entier invariant. Dans le cas des coniques par exemple, il s'agit de compter avec signe également les couples de droites tangentes à quatre de ces cinq coniques $L_1,\dots,L_5$. Je ne souhaite pas aller plus avant dans ces résultats et renvoie le lecteur intéressé aux articles donnés en référence. Signalons à propos que l'existence de ces invariants entiers, et notamment $\chi^d(T)$, a été établie dans le cadre des variétés symplectiques réelles de dimension quatre. Je n'ai présenté ici qu'un aspect particulier du résultat.

Enumération de fractions rationnelles réelles

Je voudrais conclure cette présentation par un autre aspect particulier de ce même résultat. Considérons cette fois-ci une seule fraction rationnelle $u = \frac{P}{R}$ obtenue comme quotient de deux polynômes à coefficients complexes de degré d. Elle définit une application de $\mathbb{C}$ privé des racines de R dans $\mathbb{C}$ qui a en général $2d-2$ points critiques. Ces derniers correspondent aux racines du polynôme $P'R - R'P$ qui n'est qu'au plus de degré $2d-2$ puisque les coefficients dominants de $P'R$ et $R'P$ se compensent. Remarquons de plus qu'aucune de ces racines n'est réelle en général, de sorte que $u(\mathbb{R})$ est une courbe immergée dans $\mathbb{C}$. Cette dernière reste invariante lorsque l'on compose à droite u par une homographie réelle. Puisque les fractions rationnelles complexes dépendent de $2d+1$ paramètres complexes, l'espace de ces courbes réelles $u(\mathbb{R})$ ne dépend que de $4d-1$ paramètres réels. De fait, si l'on fixe une configuration générique $\underline{y} = (y_1,\dots,y_{4d-1})$ de $4d-1$ points distincts de $\mathbb{C}$, seules un nombre fini de courbes $u(\mathbb{R})$ relient $y_1,\dots,y_{4d-1}$. Notons $\mathcal{F}rac(\underline{y})$ cet ensemble fini dont le cardinal dépend de $\underline{y}$ en général. Soit $[u]$ un élément de $\mathcal{F}rac(\underline{y})$, c'est-à-dire une fraction rationnelle modulo l'action du groupe des homographies réelles. Les nombres de points critiques de parties imaginaires positives et négatives de u ont même parité, puisque leur somme est paire. De là découle un signe $p(u) = +1$ ou -1 selon que ce nombre soit pair ou impair respectivement.

Théorème 6. *L'entier $\theta^d(\underline{y}) = \sum_{u\in\mathcal{F}rac(\underline{y})} p(u)$ ne dépend pas du choix de $\underline{y}$.*

De ce problème de géométrie énumérative réelle se dégage donc à nouveau un invariant à valeurs entières. La valeur absolue de cet invariant borne inférieurement le cardinal de $\mathcal{F}rac(\underline{y})$. Quels sont les problèmes de géométrie énumérative réelle qui cachent de tels invariants entiers ? C'est une question à laquelle il faut à mon avis répondre. Ces invariants se calculent-ils en fonction d'invariants primitifs comme leurs analogues complexes (voir le Théorème 3) ? Plusieurs travaux en cours tendent à le montrer. Enfin, l'existence même de ces invariants doit avoir des implications autres que les seules bornes inférieures présentées dans le Corollaire 5. Il reste à les découvrir !

Encadré 1

La projection stéréographique

Soit $C_2 \subset \mathbb{C}^2$ une conique lisse, p un point de C_2 et L une droite affine ne contenant pas p et parallèle à la tangente T_pC_2 de C_2 en p. Alors, chaque droite D de $\mathbb{C}^2$ qui passe par p coupe la conique en au plus un point q_D en dehors de p, et exactement un point dès qu'elle n'est parallèle ni aux directions asymptotiques de C_2, ni à la tangente T_pC_2. Par ailleurs, chacune de ces droites coupe L en un point r_D, à l'exception de T_pC_2 qui est parallèle à L. L'application de C_2 dans L qui associe r_D au point q_D est appelée « projection stéréographique ». Je laisse au lecteur le soin de vérifier que son application inverse fournit un paramétrage rationnel de C_2. Si C_3 est une cubique lisse par contre, une telle application n'est plus injective puisque la plupart de ses fibres sont formées de deux points distincts. Le raisonnement précédent ne permet donc pas d'obtenir un paramétrage rationnel de C_3. Une étude plus détaillée de la projection dans ce cas permet d'ailleurs d'établir que C_3 est homéomorphe à un tore privé d'un, deux ou trois points au maximum, voir la figure 3. On peut démontrer que cette topologie empêche C_3 d'être paramétrée par une fraction rationnelle dont la source est homéomorphe à une sphère privée de trois ou quatre points.

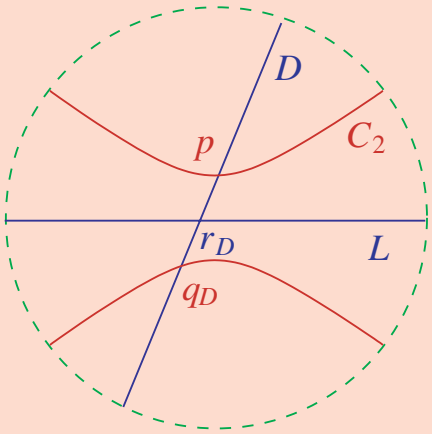


Figure 2 – *La projection stéréographique.*

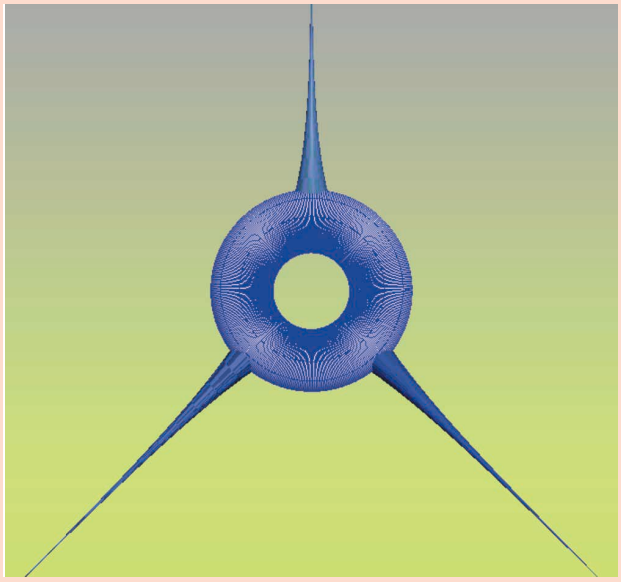


Figure 3 – *Une cubique lisse du plan complexe.*

Pour en savoir plus

[IKS1] ITENBERG (I.), KHARLAMOV (V.), SHUSTIN (E.), Welschinger invariant and enumeration of real rational curves, *Int. Math. Res. Not.*, (2003), no. 49, 2639–2653.

[IKS2] ITENBERG (I.), KHARLAMOV (V.), SHUSTIN (E.), Logarithmic equivalence of Welschinger and Gromov-Witten invariants, *Russian Math. Surveys*, (2004), no. 59, 1093-1116.

[KM] KONTSEVICH (M.), MANIN (Y.), Gromov-Witten classes, quantum cohomology, and enumerative geometry, *Comm. Math. Phys*, (1994), vol. 164, no. 3, 525–562.

[W1] WELSCHINGER (J.-Y.), Invariants of real symplectic 4-manifolds and lower bounds in real enumerative geometry, *Invent. Math*., (2005), vol. 162, no. 1, 195–234.

[W2] WELSCHINGER (J.-Y.), Towards relative invariants of real symplectic four-manifolds, *Geom. Funct. Anal.*, à paraître. (Voir la prépublication no. 332 de l'Ecole normale supérieure de Lyon, 2005.)

La Vérité et la Machine

Benjamin WERNER*

Longtemps reservée aux informaticiens et logiciens, la vérification formelle de démonstration commence à être utilisée par une fraction grandissante de la communauté mathématique.

En décembre 2004, Georges Gonthier a annoncé l'achèvement de la formalisation complète de la preuve du théorème des quatre couleurs avec le logiciel Coq [Go]. Cette nouvelle a depuis été assez largement relayée par la presse scientifique de grand public et même généraliste ; c'est d'autant plus remarquable que la nature exacte de ce résultat est finalement relativement difficile à expliquer. Par ailleurs, plusieurs résultats mathématiques célèbres ont été revérifiés par des systèmes de preuve au cours de l'année écoulée, en particulier le théorème des nombres premiers ou le théorème de Jordan. On peut donc s'interroger s'il s'agit là d'une coïncidence chronologique ou du début d'un mouvement plus vaste.

Preuves formelles : historique

La quête de la correction

On considère en général que la logique mathématique moderne est née à la fin du XIX^{e} siècle, avec les travaux de logiciens tels que Frege, Peano, Zermelo ou Russell, qui ont contribué à la définition précise de formalismes tels que l'arithmétique, la théorie des ensembles ou les premières formes de la théorie des types. A partir de ce moment-là, on peut considérer une preuve mathématique comme étant elle-même un objet mathématique, dont la correction repose sur des règles syntaxiques, bien comprises et non ambiguës. Dans la plupart des formalismes, les preuves possèdent un structure d'arbre. Par exemple, étant donnés une preuve σ_A d'une proposition A et une preuve σ_B d'une autre proposition B, on peut les combiner pour construire une preuve de la proposition A *et* B ; on écrit la règle correspondante ainsi :

$$\frac{\dfrac{\sigma_A}{\vdash A} \qquad \dfrac{\sigma_B}{\vdash B}}{\vdash A \wedge B}$$

Le point de vue logique est donc qu'une proposition peut être déclarée vraie dans un certain formalisme si elle admet une preuve formelle vérifiant les règles de ce formalisme. Un texte mathématique traditionnel peut alors être vu comme une description informelle de cette preuve formelle ; le but est de convaincre le lecteur mathématicien de l'existence de cette preuve. C'est ainsi qu'au cours du siècle dernier, le consensus s'est fait autour de l'idée que la vérité mathématique était une notion exacte et objective.

Toutefois, dans la pratique mathématique, cet objet-preuve est longtemps resté virtuel. S'il est, *en principe*, possible d'écrire, ou « dessiner », une preuve entièrement formalisée, sa taille, c'est-à-dire le nombre d'étapes élémentaires de déduction, rend cette entreprise à peu près impossible *en pratique*, si la preuve n'est pas mathématiquement triviale. Qui plus est, il est vain de croire qu'aligner des symboles peu intuitifs sur la papier réduira de quelque façon

* *LIX, Ecole Polytechnique, 91 128 PALAISEAU cedex.*
Benjamin.Werner@inria.fr

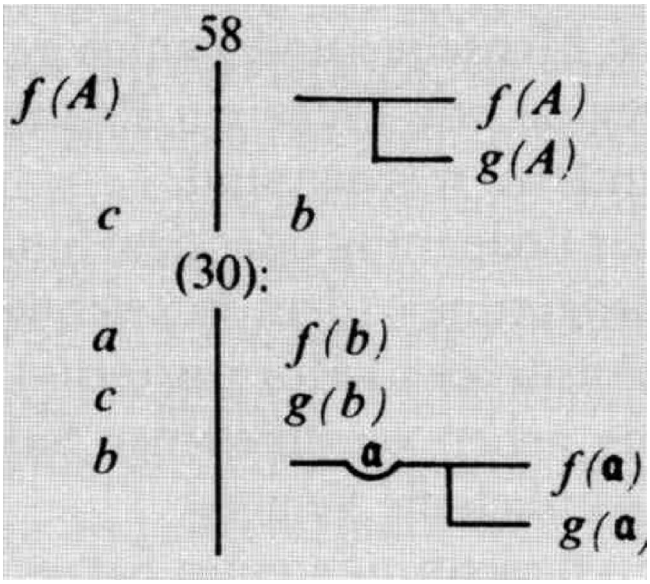


Figure 1 – *L'écriture mathématique de Frege (1872).*

que ce soit le risque d'erreur. En d'autres termes, la logique mathématique est, au départ, une discipline fondamentale et peu applicable destinée à être appliquée.

L'arrivée de l'ordinateur

Si un étudiant d'informatique d'aujourd'hui regarde les écrits de Frege, il ne peut être que frappé par ce qui lui apparaîtra comme la nature essentiellement informatique des notions développées. En effet, de part leur construction arborescentes, les propositions et les preuves sont typiquement des structures de données que les machines savent manipuler. L'ordinateur est le candidat idéal pour construire, stocker et surtout vérifier une preuve formelle.

Remarquons à ce titre, qu'il est important de distinguer les tâches de *construction* et de *vérification* de la preuve. En particulier, lors de la vérification, on ne cherche surtout pas à rendre l'ordinateur « intelligent ». Au contraire, c'est le manque d'imagination de la machine, sa précision mécanique voire bureaucratique, qui permettent d'accorder à une preuve vérifiée par ordinateur un grand degré de certitude. C'est évidemment le but recherché.

Chronologiquement, le premier système de traitement de preuve fut le logiciel Automath développé par l'équipe de N.G. de Bruijn dans les années 1960. On peut considérer ce pionnier comme l'ancêtre commun des systèmes de preuves actuels. La plupart sont développés en Europe ou aux Etats-Unis ; on peut citer parmi d'autres Coq, Isabelle et HOL (U.E.) ou PVS (USA).

Des preuves de programmes aux preuves mathématiques

Les systèmes de preuve, comme tous logiciels, vivent d'abord à travers l'utilisation qui en est faite. Si à leur débuts, les textes mathématiques, on pourrait presque dire les classiques, ont été au centre des tout premiers travaux de formalisation en général effectués par les concepteurs mêmes, les centres d'intérêt des formalisateurs se sont un moment éloignés des mathématiques pures, en même temps que l'on découvrait un domaine d'applications privilégiés des « méthodes formelles » : la preuve de propriétés de programmes informatiques.

En effet, à l'image d'un objet mathématique, un programme obéit à des règles formelles et précisément définissables. On utilise donc un raisonnement de type mathématique pour garantir que tel ou tel logiciel (comme une inversion de matrices) aura bien telle ou telle propriété (sera involutive). Malheureusement, « à mains nues » cela se révèle rapidement assez malcommode. En effet, une telle preuve de correction peut certe faire appel à des mathé-

matiques tout-à-fait subtiles et non-élémentaires ; mais la pratique montre que raisonner à propos d'un programme se révèle alors bien souvent plus bureaucratique et fastidieux que de démontrer un théorème. C'est en grande partie du au fait qu'un programme est lui-même un objet formel et le fait qu'il soit correct « dans les grandes lignes » n'empêche pas qu'une petite erreur suffise à faire échouer matériellement le programme.

La vérification de programmes est donc un domaine où :

1. La compréhension ou l'intuition du raisonnement est souvent moins utile que la pure rigueur ;
2. dont l'utilité des applications est évidente.

On comprend donc que la certification de programmes a fait l'objet de nombreux travaux théoriques et pratiques. On a donc intégré aux systèmes de preuves les techniques et les outils permettant de raisonner à propos de ces objets particuliers que sont les programmes.

Il est donc d'autant plus intéressant d'observer actuellement un certain retour de l'activité de formalisation vers les mathématiques. Une première raison est sans doute que les progrès de ces systèmes les rendent plus confortables et efficaces, et donc aussi plus intéressants pour les mathématiciens. Il semble toutefois qu'il y ait d'autres raisons plus profondes à ce mouvement, liées à la nature d'une partie des mathématiques contemporaines. C'est ce que nous allons essayer de décrire sommairement ici.

La question du calcul

Indépendamment de la question de la formalisation, le calcul informatique joue un rôle de plus en plus important dans un certain nombre de domaines mathématiques. Un exemple radical est la question des preuves de primalité : il ne viendrait à l'idée de personne de « prouver » une proposition comme « $2^{25964951} - 1$ est premier » sans recourir au calcul électronique. Si l'on veut toutefois réellement produire une preuve, la meilleure chose que l'on puisse espérer c'est de prouver que le programme utilisé établit bien la primalité. Autrement dit, si les mathématiques sont nécessaires pour prouver la correction de programmes, les programmes peuvent intervenir à leur tour dans des preuves.

Bien sûr, la propriété de primalité est directement liée au calcul ; mais il est aussi des théorèmes dont l'énoncé n'est pas essentiellement calculatoire et dont les seules preuves connues reposent pourtant sur des calculs importants, numériques ou symboliques. De ceux-là, le plus célèbre est sans doute le théorème des quatre couleurs. Ce dernier a été rejoint par la preuve de la conjecture de Kepler, également célèbre. De fait, à travers de tels résultats, l'ordinateur s'est invité à la table des mathématiciens et il importe de comprendre le statut des preuves qui l'utilisent.

La particularité des quatre couleurs

Le théorème des quatre couleurs dit, rappelons-le, qu'il est toujours possible de colorier une carte planaire, avec une couleur par pays, de telle manière que deux pays ayant une frontière commune ne soit pas de la même couleur. En général, on commence par ramener l'énoncé à la quatre-coloriabilité de graphes planaires (chaque pays correspondant alors à un sommet du graphe).

La suite de l'histoire est assez connue. Conjecturée en 1852, cette proposition simple et remarquablement « concrète » a défié pendant plus d'un siècle les efforts des nombreux mathématiciens, fameux ou anonymes, qui ont cherché à la démontrer. L'aura quelque peu mystérieuse de ce théorème auprès du grand public se renforçant encore lorsque la première démonstration fut annoncée en 1976, car cette dernière « faisait appel à l'ordinateur ».

Sans chercher, bien sûr, à exposer les détails de la preuve, il est intéressant de comprendre quel genre de tâche doit être dévolue à l'ordinateur. La preuve commence par restreindre le problème à une classe de graphes planaires triangulés appelés quasi 6-connexes. Ensuite, on se donne une liste de petits graphes particuliers appelés configurations. Dans la preuve originelle, on en comptait 1476, nombre ramené à 633 dans une preuve de 1995. On peut alors montrer que dans tout graphe triangulé et quasi 6-connexe apparaît au moins un de ces configurations ; c'est la propriété dite d'inévitabilité. Il faut noter que même si cette étape de la preuve est, pour des raisons évidentes particulièrement fastidieuse, elle reste à la portée d'une équipe de mathématiciens. De fait elle fut démontrée « à la main » pour la preuve de 1976.

On raisonne ensuite par récurrence sur le nombre de sommets du graphe. En se donnant un graphe triangulé et quasi 6-connexe, on sait donc qu'il contient une configuration. On peut alors « ôter » cette configuration (en fait la remplacer par un sous-graphe plus petit) et 4-colorier le graphe obtenu par hypothèse de récurrence. Il resterait alors à étendre ce coloriage du reste de la carte à la configuration pour conclure. Las, il n'est en général pas possible de trouver un coloriage de la configuration qui corresponde au coloriage du reste de la carte. C'est là qu'intervient l'explosion combinatoire de la preuve : en analysant, pour chaque configuration l'ensemble de tous les coloriages possibles, on arrive à montrer qu'il est toujours possible de *réorganiser le coloriage* du reste de la carte pour obtenir un coloriage qui s'étende à la configuration.

Il faut pour cela considérer l'ensemble des coloriages de chaque configuration (jusqu'à 20 000 pour une configuration), mais aussi l'ensemble des appareillements par composantes bicolores (jusqu'à 1 500 000). Même pour un ordinateur moderne utilisant un algorithme efficace, cela reste un calcul non-trivial. Cette partie est évidemment hors de portée d'un humain ou d'un groupe d'humains. On ne peut que faire confiance à la machine.

Un langage commun

La formalisation de la preuve des quatre couleurs en Coq signifie donc deux choses. D'une part on a prouvé formellement la correction des programmes utilisés pour vérifier la réductibilité des configurations. Mais surtout, on a pu, dans le même langage, celui de Coq, effectuer l'ensemble de la preuve, c'est-à-dire :

1. construire la théorie des graphes planaires et prouver, formellement, une série de lemmes et théorèmes,
2. écrire les programmes nécessaires à démontrer, formellement, le lien entre ces programmes et la propriété de réductibilité,
3. vérifier la preuve des quatre couleurs, processus qui fait intervenir les propriétés (1) et l'exécution des programmes (2).

On peut donc dire que c'est la première fois qu'une preuve du théorème des quatre couleurs est entièrement écrite : les textes existants jusqu'à maintenant ressemblaient à des textes mathématiques, jusqu'à la preuve de reductibilité. Il fallait ensuite passer des mathématiques à l'informatique en changeant de langage et donner un programme informatique ; on laissait le lecteur s'assurer qu'un résultat positif de l'exécution du programme signifiait bien l'existence d'une preuve.

Le langage de Coq évite cette rupture, puisqu'il inclut à la fois déduction et programmes. L'encadré 2 donne quelques détails sur la manière dont calcul et raisonnement peuvent s'articuler dans un formalisme comme celui de Coq.

Pour résumer la situation des preuves calculatoires, on peut donc dire d'abord que l'on a affaire à des preuves dont la vérification est hors de la portée de l'esprit humain sans assistance extérieure, et ce pour des raisons quasiment « mécaniques ». Le premier problème que cela pose est celui du langage dans lequel décrire ces preuves, puisque ce dernier doit inclure à la fois le discours mathématique traditionnel, mais aussi la possibilité de décrire des algorithmes informatiques. Les systèmes de preuve calculatoires proposent donc de tels langages hybrides, mais ils répondent aussi à la seconde interrogation à propos de l'interface entre ces deux modes : comment être sûr que le programme donné à bien les propriétés attendues, c'est-à-dire que tel ou tel résultat d'un calcul implique bien telle ou telle proposition mathématique (*cf* encadré 1).

D'un coté, on peut espérer que l'alliance entre raisonnement et calcul ouvre une nouvelle voie aux mathématiques : elle met à notre portée des résultats qui pour de simples raisons « mécaniques » (longueur de la preuve) étaient inaccessibles jusqu'à maintenant. On peut même imaginer l'appropriation par les mathématiciens d'un nouveau mode de recherche plus expérimental. Mais l'introduction de calculs informatiques rend plus difficile la vérification de ces résultats en même temps qu'est perdue une certaine forme d'intuition. La formalisation et la vérification informatique apparaissent comme le meilleur moyen de garantir la correction de ces nouvelles constructions mathématiques. Un atout étant que, comme évoqué plus haut, la vérification de logiciels est historiquement l'un des premiers débouchés pour la formalisation.

D'un autre coté, formaliser ces raisonnements d'un type nouveau nécessite une complexification du langage mathématique, puisqu'il faut lui incorporer un langage de programmation. Or, même si ce langage peut être plus épuré que celui qui sert à programmer un jeu vidéo, il se pose néanmoins la question du standard. On se trouve déjà dans une situation de concurrence entre plusieurs formalismes et plusieurs systèmes de preuve et la question est

ouverte de savoir lesquels pourront s'établir comme des standards pour la communauté mathématique et par quels processus. On peut toutefois penser que le défi pour les développeurs des systèmes de preuve est de les rendre suffisamment pratiques pour les faire adopter au moins par les mathématiciens dont le domaine de recherche gagnerait à être formalisé. Pour les mathématiciens, il s'agira d'apprivoiser les systèmes de preuves, techniquement mais aussi socialement, en les intégrant aux critères de publication existants.

Conjecture de Kepler ou théorème de Hales ?

La question de la crédibilité des preuves mêlant raisonnement et calcul s'est reposée de manière brûlante voici peu de temps. Pendant vingt ans, la preuve du théorème des quatre couleurs est restée unique en son genre, et l'emploi important qu'elle faisait du calcul a surtout contribué à lui donner une aura particulière. Mais l'Histoire s'est répétée une première fois en 1998, lorsque Thomas Hales a proposé une preuve de la célèbre *conjecture de Kepler*. Là encore il s'agit d'une conjecture ancienne, puisqu' énoncée en 1611 et qui peut être expliquée dans un langage non-mathématique : il n'y a pas de façon de ranger des oranges qui soit meilleure que celle des étals des marchés.

A l'image de celle du théorème des quatre couleurs, la preuve de Hales repose sur des calculs informatiques importants. Elle est toutefois nettement plus complexe, d'une part parce que ces calculs interviennent à plusieurs niveaux dans la preuve, et surtout parce que la partie « conventionnelle », c'est-à-dire non calculatoire de la preuve est elle aussi plus longue et utilise plus de résultats pré-existants que celle des quatre couleurs.

Thomas Hales a soumis une série d'articles décrivant sa preuve en 1999, mais ces derniers ne paraîtront qu'en 2005 ; les relecteurs ne sachant pas comment traiter les parties calculatoires. Qui plus est, ces articles seront accompagné d'une mise en garde de l'éditeur précisant que ces parties informatiques de la preuve n'ont pas pu être vérifiées. On peut voir là le reflet d'une certaine réticence vis-à-vis de l'utilisation d'outils informatiques, mais il est indéniable que ce type de preuves est difficile à valider avec les procédures habituelles. D'une part parce que leur lecture complète nécessitent des compétences en programmation. Mais aussi parce que même pour un informaticien professionnel, la correction d'un programme est difficile à établir, pour les raisons déjà mentionnées.

Depuis cette relative mésaventure, Thomas Hales est devenu un ardent défenseur des mathématiques formelles. Qui plus est, ayant eu vent de l'effort de formalisation de la preuve des quatre couleurs, il a lancé un projet de formalisation de sa propre preuve qu'il a intitulé Flyspeck. Toutefois, il faudra sans doute attendre longtemps l'achèvement de ce projet. D'une part à cause de la complexité de la preuve elle-même, mais aussi parce que la littérature mathématique sur laquelle elle repose n'est pas encore formalisée.

Encadré 1

On peut prendre un exemple simple pour illustrer l'articulation preuve/calcul dans un formalisme moderne. Il est facile d'écrire un programme `test` *qui prend un nombre entier n en entrée et va essayer de le diviser par tous les nombres entiers compris entre 2 et* $n-1$. *Ce programme rendra* `true` *s'il ne trouve pas de diviseur et* `false` *sinon. Dans le formalisme, ce programme est une fonction des entiers vers les booléens. On peut alors facilement prouver le « théorème » suivant :*

$$\forall n.\texttt{test}(n) = \texttt{true} \Longrightarrow \texttt{premier}(n)$$

c'est-à-dire que ce programme est effectivement un test correct pour la primalité.
Si on applique ce théorème à, par exemple, 1789, on obtient une preuve de

$$\texttt{test}(1789) = \texttt{true} \Longrightarrow \texttt{premier}(1789).$$

Pour déduire que 1789 est premier, il nous suffit donc d'exhiber une preuve de $\texttt{test}(1789) = \texttt{true}$. *Or comme le programme* $\texttt{test}(1789)$ *rend la valeur* `true` *(on dit aussi qu'il s'évalue vers* `true`*), les objets mathématiques* $\texttt{test}(1789)$ *et* `true` *sont identifiés. Par congruence, cela veut dire que la proposition* $\texttt{test}(1789) = \texttt{true}$ *est elle identifiée à* $\texttt{true} = \texttt{true}$ *qui est une conséquence immédiate de la réflexivité de l'égalité.*
Pour prouver la primalité de nombres plus grands par la même méthode, il suffit alors d'utiliser des programmes moins naïfs et plus efficaces que `test`.

Encadré 2

La programmation fonctionnelle

La programmation fonctionnelle, à l'image de sa cousine la programmation orientée-objet, est un style de programmation commun à un certain nombre de langages de programmation. Les plus connu en France sont sans doute les différents dialectes de Caml, qui est aujourd'hui enseigné en classes préparatoires, mais on peut également citer Scheme, Standard ML ou Haskell. L'appellation de ces langages vient de ce qu'ils manipulent indifféremment des structures informatiques usuelles comme des entiers, chaînes de caractères ou tableaux, que des fonctions opérant sur ces structures. On peut ainsi définir une fonction prenant d'autres fonctions comme arguments.
*Ces langages permettent souvent une programmation claire et concise. Mais aussi, ils ont une nature plus mathématiques que les langages de programmation plus traditionnel. En effet, pour ces derniers, les programmes sont composés d'*instructions *dont l'exécution se traduit par une modification de la mémoire de l'ordinateur. C'est pourquoi on parle alors de langage* impératifs.
En programmation fonctionnelle, l'unité de base est la fonction et on laisse l'environnement d'exécution de l'ordinateur gérer l'espace mémoire pour calculer les valeurs de ces fonctions. De l'extérieur, un tel programme apparaît donc comme entièrement décrit par la manière dont il associe des valeurs aux arguments, exactement à l'image de la notion mathématique de fonction.
L'avantage pour les logiciens est qu'il devient possible de construire un formalisme logique dont les objets sont ces fonctions définies par un programme. De plus, ces programmes étant exécutables, l'expression 2+2 se simplifiera automatiquement en 4.

En sus de cet ambitieux projet, Hales a récemment annoncé et publié une preuve du théorème de Jordan, entièrement formalisée dans le système HOL-light.

Le monstre et la machine

Si l'ordinateur semble l'outil adapté pour dompter les monstres engendrés par les mathématiques contemporaines, ces derniers ne doivent pas tous leur caractère particulier au seul calcul. L'un des résultats les plus singuliers de l'époque récente est évidemment la classification des groupes finis. Il s'agit là d'un résultat mathématique que l'on peut dire « conventionnel » de part son énoncé comme de sa preuve, si ce n'était leurs tailles : la preuve a été estimée à 15 000 pages, dispersées dans la littérature mathématique. Il paraît donc là aussi difficile de parler d'intuition ou d'évidence cartésienne à propos d'un tel ensemble de preuves et une formalisation de cet important corpus serait sans doute un apport à ce délicat édifice[1].

On peut espérer voir naître de tels efforts de formalisation dans un futur très proche. Il sera particulièrement intéressant d'en observer les progrès, car ce défi recouvre deux points importants et qui restent aujourd'hui encore particulièrement délicats à maîtriser. Le premier est qu'un tel travail ne pourrait être que collectif, nécessiterait donc de coordonner un important travail d'équipe autour d'un développement. Le second concerne la question des bibliothèques.

Une question de style

Si la communauté mathématique témoigne d'un intérêt grandissant pour la vérification informatique du raisonnement on doit aussi se poser la question des freins à cette évolution. Un premier obstacle à l'utilisation d'un système de preuve est la nécessité d'apprendre le langage avec lequel les preuves sont communiquées à l'ordinateur, mais si cela demande un certain investissement initial, c'est en général un obstacle surmontable particulièrement pour des mathématiciens rompus au langage algébrique. Un facteur plus critique actuellement reste la difficulté à construire des bibliothèques de développements mathématiques formels réellement pratiques et réutilisables. De fait, on observe qu'au début d'un développement important, les utilisateurs préfèrent encore souvent repartir de zéro plutôt que de commencer à construire sur des définitions ou des lemmes repris sur un autre travail. Ce phénomène est identifié depuis longtemps en programmation, où l'on parle souvent de la question de la « réutilisabilité du code ».

1. A une échelle plus humaine, il faut mentionner les travaux en géométrie algébrique comme ceux de Carlos Simpson. La complexité croissante de certains de ses énoncés l'ont conduit à formaliser ses travaux en Coq.

Si cette réutilisabilité ne va pas forcément de soi en mathématiques formelles, c'est que celle-ci sont particulièrement sensibles aux détails des définitions et des énoncés. Dans un texte traditionnel, on utilisera la notion de polynôme indifféremment pour désigner une classe particulière de fonctions réelles, ou une classe d'expressions algébrique (à savoir les sommes de monômes). De plus, on regardera le même polynôme parfois comme une fonction et parfois comme une expression algébrique. Si l'on travaille formellement, la fonction et l'expression sont deux objets distincts ; on peut certes définir les traductions de l'une vers l'autre, mais chaque invocation de ces traductions alourdira énoncés et preuves.

Or la lourdeur est l'écueil principal de la preuve formelle. Dans les faits, lorsqu'un développement ne peut être mené à son terme, c'est en généralement parce qu'à partir d'un moment il risque de crouler sous le nombre d'hypothèses locales, le nombre de « petits lemmes » ou le nombre d'étapes « triviales ».

Pour éviter cela, la meilleure arme reste un choix judicieux des définitions et des énoncés, qui permettra la preuve la moins bureaucratique possible, c'est-à-dire la plus élégante. En d'autre termes, et c'est sans doute une bonne nouvelle, tout en étant, comme on le lui demande, un vérificateur sans âme, l'ordinateur devient également un garant du beau style mathématique.

Pour en savoir plus

[Coq] Le Système Coq, Diffusé par l'INRIA, http ://coq.inria.fr

[Go] Gonthier (G.), A computer-checked proof of the Four Color Theorem. Disponible sur http ://research.microsoft.com/ gonthier/

[Hales] Hales (T.), divers articles et documents sur http ://www.math.pitt.edu/ thales/

[GnThWe] Grégoire (B), Théry (L.) et Werner (B.), a Computational Approach to Pockington Certificates in Type Theory. FLOPS 2006, LNCS 3945, Springer, Venlag 2006.

La Vérité et la Machine

Benjamin WERNER*

Longtemps reservée aux informaticiens et logiciens, la vérification formelle de démonstration commence à être utilisée par une fraction grandissante de la communauté mathématique.

En décembre 2004, Georges Gonthier a annoncé l'achèvement de la formalisation complète de la preuve du théorème des quatre couleurs avec le logiciel Coq [Go]. Cette nouvelle a depuis été assez largement relayée par la presse scientifique de grand public et même généraliste ; c'est d'autant plus remarquable que la nature exacte de ce résultat est finalement relativement difficile à expliquer. Par ailleurs, plusieurs résultats mathématiques célèbres ont été revérifiés par des systèmes de preuve au cours de l'année écoulée, en particulier le théorème des nombres premiers ou le théorème de Jordan. On peut donc s'interroger s'il s'agit là d'une coïncidence chronologique ou du début d'un mouvement plus vaste.

Preuves formelles : historique

La quête de la correction

On considère en général que la logique mathématique moderne est née à la fin du XIX^{e} siècle, avec les travaux de logiciens tels que Frege, Peano, Zermelo ou Russell, qui ont contribué à la définition précise de formalismes tels que l'arithmétique, la théorie des ensembles ou les premières formes de la théorie des types. A partir de ce moment-là, on peut considérer une preuve mathématique comme étant elle-même un objet mathématique, dont la correction repose sur des règles syntaxiques, bien comprises et non ambiguës. Dans la plupart des formalismes, les preuves possèdent un structure d'arbre. Par exemple, étant donnés une preuve σ_A d'une proposition A et une preuve σ_B d'une autre proposition B, on peut les combiner pour construire une preuve de la proposition A *et* B ; on écrit la règle correspondante ainsi :

$$\frac{\dfrac{\sigma_A}{\vdash A} \qquad \dfrac{\sigma_B}{\vdash B}}{\vdash A \wedge B}$$

Le point de vue logique est donc qu'une proposition peut être déclarée vraie dans un certain formalisme si elle admet une preuve formelle vérifiant les règles de ce formalisme. Un texte mathématique traditionnel peut alors être vu comme une description informelle de cette preuve formelle ; le but est de convaincre le lecteur mathématicien de l'existence de cette preuve. C'est ainsi qu'au cours du siècle dernier, le consensus s'est fait autour de l'idée que la vérité mathématique était une notion exacte et objective.

Toutefois, dans la pratique mathématique, cet objet-preuve est longtemps resté virtuel. S'il est, *en principe*, possible d'écrire, ou « dessiner », une preuve entièrement formalisée, sa taille, c'est-à-dire le nombre d'étapes élémentaires de déduction, rend cette entreprise à peu près impossible *en pratique*, si la preuve n'est pas mathématiquement triviale. Qui plus est, il est vain de croire qu'aligner des symboles peu intuitifs sur la papier réduira de quelque façon

* *LIX, Ecole Polytechnique, 91 128 PALAISEAU cedex.*
Benjamin.Werner@inria.fr

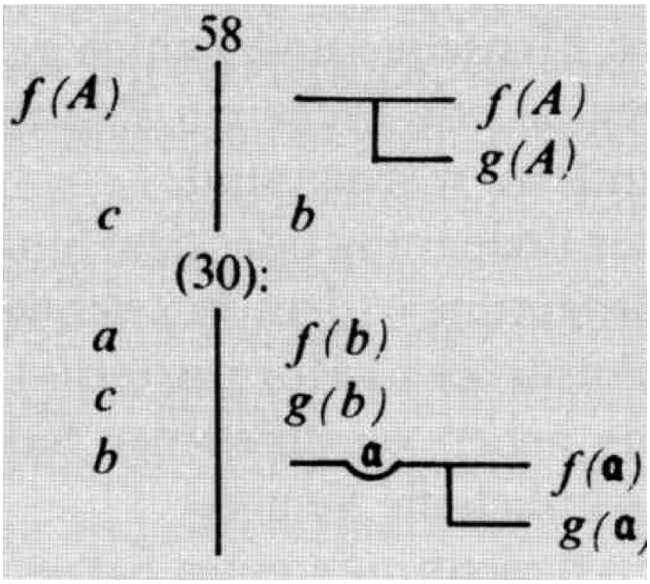


Figure 1 – *L'écriture mathématique de Frege (1872).*

que ce soit le risque d'erreur. En d'autres termes, la logique mathématique est, au départ, une discipline fondamentale et peu applicable destinée à être appliquée.

L'arrivée de l'ordinateur

Si un étudiant d'informatique d'aujourd'hui regarde les écrits de Frege, il ne peut être que frappé par ce qui lui apparaîtra comme la nature essentiellement informatique des notions développées. En effet, de part leur construction arborescentes, les propositions et les preuves sont typiquement des structures de données que les machines savent manipuler. L'ordinateur est le candidat idéal pour construire, stocker et surtout vérifier une preuve formelle.

Remarquons à ce titre, qu'il est important de distinguer les tâches de *construction* et de *vérification* de la preuve. En particulier, lors de la vérification, on ne cherche surtout pas à rendre l'ordinateur « intelligent ». Au contraire, c'est le manque d'imagination de la machine, sa précision mécanique voire bureaucratique, qui permettent d'accorder à une preuve vérifiée par ordinateur un grand degré de certitude. C'est évidemment le but recherché.

Chronologiquement, le premier système de traitement de preuve fut le logiciel Automath développé par l'équipe de N.G. de Bruijn dans les années 1960. On peut considérer ce pionnier comme l'ancêtre commun des systèmes de preuves actuels. La plupart sont développés en Europe ou aux Etats-Unis ; on peut citer parmi d'autres Coq, Isabelle et HOL (U.E.) ou PVS (USA).

Des preuves de programmes aux preuves mathématiques

Les systèmes de preuve, comme tous logiciels, vivent d'abord à travers l'utilisation qui en est faite. Si à leur débuts, les textes mathématiques, on pourrait presque dire les classiques, ont été au centre des tout premiers travaux de formalisation en général effectués par les concepteurs mêmes, les centres d'intérêt des formalisateurs se sont un moment éloignés des mathématiques pures, en même temps que l'on découvrait un domaine d'applications privilégiés des « méthodes formelles » : la preuve de propriétés de programmes informatiques.

En effet, à l'image d'un objet mathématique, un programme obéit à des règles formelles et précisément définissables. On utilise donc un raisonnement de type mathématique pour garantir que tel ou tel logiciel (comme une inversion de matrices) aura bien telle ou telle propriété (sera involutive). Malheureusement, « à mains nues » cela se révèle rapidement assez malcommode. En effet, une telle preuve de correction peut certe faire appel à des mathé-

matiques tout-à-fait subtiles et non-élémentaires ; mais la pratique montre que raisonner à propos d'un programme se révèle alors bien souvent plus bureaucratique et fastidieux que de démontrer un théorème. C'est en grande partie du au fait qu'un programme est lui-même un objet formel et le fait qu'il soit correct « dans les grandes lignes » n'empêche pas qu'une petite erreur suffise à faire échouer matériellement le programme.

La vérification de programmes est donc un domaine où :

1. La compréhension ou l'intuition du raisonnement est souvent moins utile que la pure rigueur ;
2. dont l'utilité des applications est évidente.

On comprend donc que la certification de programmes a fait l'objet de nombreux travaux théoriques et pratiques. On a donc intégré aux systèmes de preuves les techniques et les outils permettant de raisonner à propos de ces objets particuliers que sont les programmes.

Il est donc d'autant plus intéressant d'observer actuellement un certain retour de l'activité de formalisation vers les mathématiques. Une première raison est sans doute que les progrès de ces systèmes les rendent plus confortables et efficaces, et donc aussi plus intéressants pour les mathématiciens. Il semble toutefois qu'il y ait d'autres raisons plus profondes à ce mouvement, liées à la nature d'une partie des mathématiques contemporaines. C'est ce que nous allons essayer de décrire sommairement ici.

La question du calcul

Indépendamment de la question de la formalisation, le calcul informatique joue un rôle de plus en plus important dans un certain nombre de domaines mathématiques. Un exemple radical est la question des preuves de primalité : il ne viendrait à l'idée de personne de « prouver » une proposition comme « $2^{25964951} - 1$ est premier » sans recourir au calcul électronique. Si l'on veut toutefois réellement produire une preuve, la meilleure chose que l'on puisse espérer c'est de prouver que le programme utilisé établit bien la primalité. Autrement dit, si les mathématiques sont nécessaires pour prouver la correction de programmes, les programmes peuvent intervenir à leur tour dans des preuves.

Bien sûr, la propriété de primalité est directement liée au calcul ; mais il est aussi des théorèmes dont l'énoncé n'est pas essentiellement calculatoire et dont les seules preuves connues reposent pourtant sur des calculs importants, numériques ou symboliques. De ceux-là, le plus célèbre est sans doute le théorème des quatre couleurs. Ce dernier a été rejoint par la preuve de la conjecture de Kepler, également célèbre. De fait, à travers de tels résultats, l'ordinateur s'est invité à la table des mathématiciens et il importe de comprendre le statut des preuves qui l'utilisent.

La particularité des quatre couleurs

Le théorème des quatre couleurs dit, rappelons-le, qu'il est toujours possible de colorier une carte planaire, avec une couleur par pays, de telle manière que deux pays ayant une frontière commune ne soit pas de la même couleur. En général, on commence par ramener l'énoncé à la quatre-coloriabilité de graphes planaires (chaque pays correspondant alors à un sommet du graphe).

La suite de l'histoire est assez connue. Conjecturée en 1852, cette proposition simple et remarquablement « concrète » a défié pendant plus d'un siècle les efforts des nombreux mathématiciens, fameux ou anonymes, qui ont cherché à la démontrer. L'aura quelque peu mystérieuse de ce théorème auprès du grand public se renforçant encore lorsque la première démonstration fut annoncée en 1976, car cette dernière « faisait appel à l'ordinateur ».

Sans chercher, bien sûr, à exposer les détails de la preuve, il est intéressant de comprendre quel genre de tâche doit être dévolue à l'ordinateur. La preuve commence par restreindre le problème à une classe de graphes planaires triangulés appelés quasi 6-connexes. Ensuite, on se donne une liste de petits graphes particuliers appelés configurations. Dans la preuve originelle, on en comptait 1476, nombre ramené à 633 dans une preuve de 1995. On peut alors montrer que dans tout graphe triangulé et quasi 6-connexe apparaît au moins un de ces configurations ; c'est la propriété dite d'inévitabilité. Il faut noter que même si cette étape de la preuve est, pour des raisons évidentes particulièrement fastidieuse, elle reste à la portée d'une équipe de mathématiciens. De fait elle fut démontrée « à la main » pour la preuve de 1976.

On raisonne ensuite par récurrence sur le nombre de sommets du graphe. En se donnant un graphe triangulé et quasi 6-connexe, on sait donc qu'il contient une configuration. On peut alors « ôter » cette configuration (en fait la remplacer par un sous-graphe plus petit) et 4-colorier le graphe obtenu par hypothèse de récurrence. Il resterait alors à étendre ce coloriage du reste de la carte à la configuration pour conclure. Las, il n'est en général pas possible de trouver un coloriage de la configuration qui corresponde au coloriage du reste de la carte. C'est là qu'intervient l'explosion combinatoire de la preuve : en analysant, pour chaque configuration l'ensemble de tous les coloriages possibles, on arrive à montrer qu'il est toujours possible de *réorganiser le coloriage* du reste de la carte pour obtenir un coloriage qui s'étende à la configuration.

Il faut pour cela considérer l'ensemble des coloriages de chaque configuration (jusqu'à 20 000 pour une configuration), mais aussi l'ensemble des appareillements par composantes bicolores (jusqu'à 1 500 000). Même pour un ordinateur moderne utilisant un algorithme efficace, cela reste un calcul non-trivial. Cette partie est évidemment hors de portée d'un humain ou d'un groupe d'humains. On ne peut que faire confiance à la machine.

Un langage commun

La formalisation de la preuve des quatre couleurs en Coq signifie donc deux choses. D'une part on a prouvé formellement la correction des programmes utilisés pour vérifier la réductibilité des configurations. Mais surtout, on a pu, dans le même langage, celui de Coq, effectuer l'ensemble de la preuve, c'est-à-dire :

1. construire la théorie des graphes planaires et prouver, formellement, une série de lemmes et théorèmes,
2. écrire les programmes nécessaires à démontrer, formellement, le lien entre ces programmes et la propriété de réductibilité,
3. vérifier la preuve des quatre couleurs, processus qui fait intervenir les propriétés (1) et l'exécution des programmes (2).

On peut donc dire que c'est la première fois qu'une preuve du théorème des quatre couleurs est entièrement écrite : les textes existants jusqu'à maintenant ressemblaient à des textes mathématiques, jusqu'à la preuve de reductibilité. Il fallait ensuite passer des mathématiques à l'informatique en changeant de langage et donner un programme informatique ; on laissait le lecteur s'assurer qu'un résultat positif de l'exécution du programme signifiait bien l'existence d'une preuve.

Le langage de Coq évite cette rupture, puisqu'il inclut à la fois déduction et programmes. L'encadré 2 donne quelques détails sur la manière dont calcul et raisonnement peuvent s'articuler dans un formalisme comme celui de Coq.

Pour résumer la situation des preuves calculatoires, on peut donc dire d'abord que l'on a affaire à des preuves dont la vérification est hors de la portée de l'esprit humain sans assistance extérieure, et ce pour des raisons quasiment « mécaniques ». Le premier problème que cela pose est celui du langage dans lequel décrire ces preuves, puisque ce dernier doit inclure à la fois le discours mathématique traditionnel, mais aussi la possibilité de décrire des algorithmes informatiques. Les systèmes de preuve calculatoires proposent donc de tels langages hybrides, mais ils répondent aussi à la seconde interrogation à propos de l'interface entre ces deux modes : comment être sûr que le programme donné à bien les propriétés attendues, c'est-à-dire que tel ou tel résultat d'un calcul implique bien telle ou telle proposition mathématique (*cf* encadré 1).

D'un coté, on peut espérer que l'alliance entre raisonnement et calcul ouvre une nouvelle voie aux mathématiques : elle met à notre portée des résultats qui pour de simples raisons « mécaniques » (longueur de la preuve) étaient inaccessibles jusqu'à maintenant. On peut même imaginer l'appropriation par les mathématiciens d'un nouveau mode de recherche plus expérimental. Mais l'introduction de calculs informatiques rend plus difficile la vérification de ces résultats en même temps qu'est perdue une certaine forme d'intuition. La formalisation et la vérification informatique apparaissent comme le meilleur moyen de garantir la correction de ces nouvelles constructions mathématiques. Un atout étant que, comme évoqué plus haut, la vérification de logiciels est historiquement l'un des premiers débouchés pour la formalisation.

D'un autre coté, formaliser ces raisonnements d'un type nouveau nécessite une complexification du langage mathématique, puisqu'il faut lui incorporer un langage de programmation. Or, même si ce langage peut être plus épuré que celui qui sert à programmer un jeu vidéo, il se pose néanmoins la question du standard. On se trouve déjà dans une situation de concurrence entre plusieurs formalismes et plusieurs systèmes de preuve et la question est

ouverte de savoir lesquels pourront s'établir comme des standards pour la communauté mathématique et par quels processus. On peut toutefois penser que le défi pour les développeurs des systèmes de preuve est de les rendre suffisamment pratiques pour les faire adopter au moins par les mathématiciens dont le domaine de recherche gagnerait à être formalisé. Pour les mathématiciens, il s'agira d'apprivoiser les systèmes de preuves, techniquement mais aussi socialement, en les intégrant aux critères de publication existants.

Conjecture de Kepler ou théorème de Hales ?

La question de la crédibilité des preuves mêlant raisonnement et calcul s'est reposée de manière brûlante voici peu de temps. Pendant vingt ans, la preuve du théorème des quatre couleurs est restée unique en son genre, et l'emploi important qu'elle faisait du calcul a surtout contribué à lui donner une aura particulière. Mais l'Histoire s'est répétée une première fois en 1998, lorsque Thomas Hales a proposé une preuve de la célèbre *conjecture de Kepler*. Là encore il s'agit d'une conjecture ancienne, puisqu' énoncée en 1611 et qui peut être expliquée dans un langage non-mathématique : il n'y a pas de façon de ranger des oranges qui soit meilleure que celle des étals des marchés.

A l'image de celle du théorème des quatre couleurs, la preuve de Hales repose sur des calculs informatiques importants. Elle est toutefois nettement plus complexe, d'une part parce que ces calculs interviennent à plusieurs niveaux dans la preuve, et surtout parce que la partie « conventionnelle », c'est-à-dire non calculatoire de la preuve est elle aussi plus longue et utilise plus de résultats pré-existants que celle des quatre couleurs.

Thomas Hales a soumis une série d'articles décrivant sa preuve en 1999, mais ces derniers ne paraîtront qu'en 2005 ; les relecteurs ne sachant pas comment traiter les parties calculatoires. Qui plus est, ces articles seront accompagné d'une mise en garde de l'éditeur précisant que ces parties informatiques de la preuve n'ont pas pu être vérifiées. On peut voir là le reflet d'une certaine réticence vis-à-vis de l'utilisation d'outils informatiques, mais il est indéniable que ce type de preuves est difficile à valider avec les procédures habituelles. D'une part parce que leur lecture complète nécessitent des compétences en programmation. Mais aussi parce que même pour un informaticien professionnel, la correction d'un programme est difficile à établir, pour les raisons déjà mentionnées.

Depuis cette relative mésaventure, Thomas Hales est devenu un ardent défenseur des mathématiques formelles. Qui plus est, ayant eu vent de l'effort de formalisation de la preuve des quatre couleurs, il a lancé un projet de formalisation de sa propre preuve qu'il a intitulé Flyspeck. Toutefois, il faudra sans doute attendre longtemps l'achèvement de ce projet. D'une part à cause de la complexité de la preuve elle-même, mais aussi parce que la littérature mathématique sur laquelle elle repose n'est pas encore formalisée.

Encadré 1

On peut prendre un exemple simple pour illustrer l'articulation preuve/calcul dans un formalisme moderne. Il est facile d'écrire un programme `test` *qui prend un nombre entier n en entrée et va essayer de le diviser par tous les nombres entiers compris entre 2 et* $n-1$. *Ce programme rendra* `true` *s'il ne trouve pas de diviseur et* `false` *sinon. Dans le formalisme, ce programme est une fonction des entiers vers les booléens. On peut alors facilement prouver le « théorème » suivant :*

$$\forall n.\texttt{test}(n) = \texttt{true} \Longrightarrow \texttt{premier}(n)$$

c'est-à-dire que ce programme est effectivement un test correct pour la primalité.
Si on applique ce théorème à, par exemple, 1789, on obtient une preuve de

$$\texttt{test}(1789) = \texttt{true} \Longrightarrow \texttt{premier}(1789).$$

Pour déduire que 1789 est premier, il nous suffit donc d'exhiber une preuve de $\texttt{test}(1789) = \texttt{true}$. *Or comme le programme* $\texttt{test}(1789)$ *rend la valeur* `true` *(on dit aussi qu'il s'évalue vers* `true`*), les objets mathématiques* $\texttt{test}(1789)$ *et* `true` *sont identifiés. Par congruence, cela veut dire que la proposition* $\texttt{test}(1789) = \texttt{true}$ *est elle identifiée à* $\texttt{true} = \texttt{true}$ *qui est une conséquence immédiate de la réflexivité de l'égalité.*
Pour prouver la primalité de nombres plus grands par la même méthode, il suffit alors d'utiliser des programmes moins naïfs et plus efficaces que `test`.

Encadré 2

La programmation fonctionnelle

La programmation fonctionnelle, à l'image de sa cousine la programmation orientée-objet, est un style de programmation commun à un certain nombre de langages de programmation. Les plus connu en France sont sans doute les différents dialectes de Caml, qui est aujourd'hui enseigné en classes préparatoires, mais on peut également citer Scheme, Standard ML ou Haskell. L'appellation de ces langages vient de ce qu'ils manipulent indifféremment des structures informatiques usuelles comme des entiers, chaînes de caractères ou tableaux, que des fonctions opérant sur ces structures. On peut ainsi définir une fonction prenant d'autres fonctions comme arguments.
*Ces langages permettent souvent une programmation claire et concise. Mais aussi, ils ont une nature plus mathématiques que les langages de programmation plus traditionnel. En effet, pour ces derniers, les programmes sont composés d'*instructions *dont l'exécution se traduit par une modification de la mémoire de l'ordinateur. C'est pourquoi on parle alors de langage* impératifs.
En programmation fonctionnelle, l'unité de base est la fonction et on laisse l'environnement d'exécution de l'ordinateur gérer l'espace mémoire pour calculer les valeurs de ces fonctions. De l'extérieur, un tel programme apparaît donc comme entièrement décrit par la manière dont il associe des valeurs aux arguments, exactement à l'image de la notion mathématique de fonction.
L'avantage pour les logiciens est qu'il devient possible de construire un formalisme logique dont les objets sont ces fonctions définies par un programme. De plus, ces programmes étant exécutables, l'expression 2+2 se simplifiera automatiquement en 4.

En sus de cet ambitieux projet, Hales a récemment annoncé et publié une preuve du théorème de Jordan, entièrement formalisée dans le système HOL-light.

Le monstre et la machine

Si l'ordinateur semble l'outil adapté pour dompter les monstres engendrés par les mathématiques contemporaines, ces derniers ne doivent pas tous leur caractère particulier au seul calcul. L'un des résultats les plus singuliers de l'époque récente est évidemment la classification des groupes finis. Il s'agit là d'un résultat mathématique que l'on peut dire « conventionnel » de part son énoncé comme de sa preuve, si ce n'était leurs tailles : la preuve a été estimée à 15 000 pages, dispersées dans la littérature mathématique. Il paraît donc là aussi difficile de parler d'intuition ou d'évidence cartésienne à propos d'un tel ensemble de preuves et une formalisation de cet important corpus serait sans doute un apport à ce délicat édifice[1].

On peut espérer voir naître de tels efforts de formalisation dans un futur très proche. Il sera particulièrement intéressant d'en observer les progrès, car ce défi recouvre deux points importants et qui restent aujourd'hui encore particulièrement délicats à maîtriser. Le premier est qu'un tel travail ne pourrait être que collectif, nécessiterait donc de coordonner un important travail d'équipe autour d'un développement. Le second concerne la question des bibliothèques.

Une question de style

Si la communauté mathématique témoigne d'un intérêt grandissant pour la vérification informatique du raisonnement on doit aussi se poser la question des freins à cette évolution. Un premier obstacle à l'utilisation d'un système de preuve est la nécessité d'apprendre le langage avec lequel les preuves sont communiquées à l'ordinateur, mais si cela demande un certain investissement initial, c'est en général un obstacle surmontable particulièrement pour des mathématiciens rompus au langage algébrique. Un facteur plus critique actuellement reste la difficulté à construire des bibliothèques de développements mathématiques formels réellement pratiques et réutilisables. De fait, on observe qu'au début d'un développement important, les utilisateurs préfèrent encore souvent repartir de zéro plutôt que de commencer à construire sur des définitions ou des lemmes repris sur un autre travail. Ce phénomène est identifié depuis longtemps en programmation, où l'on parle souvent de la question de la « réutilisabilité du code ».

1. A une échelle plus humaine, il faut mentionner les travaux en géométrie algébrique comme ceux de Carlos Simpson. La complexité croissante de certains de ses énoncés l'ont conduit à formaliser ses travaux en Coq.

Si cette réutilisabilité ne va pas forcément de soi en mathématiques formelles, c'est que celle-ci sont particulièrement sensibles aux détails des définitions et des énoncés. Dans un texte traditionnel, on utilisera la notion de polynôme indifféremment pour désigner une classe particulière de fonctions réelles, ou une classe d'expressions algébrique (à savoir les sommes de monômes). De plus, on regardera le même polynôme parfois comme une fonction et parfois comme une expression algébrique. Si l'on travaille formellement, la fonction et l'expression sont deux objets distincts ; on peut certes définir les traductions de l'une vers l'autre, mais chaque invocation de ces traductions alourdira énoncés et preuves.

Or la lourdeur est l'écueil principal de la preuve formelle. Dans les faits, lorsqu'un développement ne peut être mené à son terme, c'est en généralement parce qu'à partir d'un moment il risque de crouler sous le nombre d'hypothèses locales, le nombre de « petits lemmes » ou le nombre d'étapes « triviales ».

Pour éviter cela, la meilleure arme reste un choix judicieux des définitions et des énoncés, qui permettra la preuve la moins bureaucratique possible, c'est-à-dire la plus élégante. En d'autre termes, et c'est sans doute une bonne nouvelle, tout en étant, comme on le lui demande, un vérificateur sans âme, l'ordinateur devient également un garant du beau style mathématique.

Pour en savoir plus

[Coq] Le Système Coq, Diffusé par l'INRIA, http ://coq.inria.fr

[Go] Gonthier (G.), A computer-checked proof of the Four Color Theorem. Disponible sur http ://research.microsoft.com/ gonthier/

[Hales] Hales (T.), divers articles et documents sur http ://www.math.pitt.edu/ thales/

[GnThWe] Grégoire (B), Théry (L.) et Werner (B.), a Computational Approach to Pockington Certificates in Type Theory. FLOPS 2006, LNCS 3945, Springer, Venlag 2006.