

MATHÉMATIQUES

Compression des images fixes

Yves Meyer

Ce texte a été rédigé par Yves Meyer à la suite de son intervention à la journée de conférences « La face cachée des mathématiques ». Vous trouverez en addendum un compte rendu de cette journée par une participante, Geneviève Courtade-Coulomb. Nous espérons pouvoir publier par la suite d'autres interventions à cette journée.

L'analyse par ondelettes est née d'une étonnante découverte faite par un ingénieur, Jean Morlet, à la fin des années soixante-dix. Comme nous le verrons, cette découverte fut aussitôt comprise et interprétée par Alexandre Grossmann, un physicien déjà célèbre pour ses travaux en *mécanique quantique*. Grossmann sut relier les idées de Jean Morlet à la théorie des *états cohérents* utilisée en mécanique quantique. Vingt ans après, l'analyse par ondelettes débouchait sur le nouveau standard, nommé JPEG2000, de compression des images fixes. Mais aujourd'hui, JPEG2000 est déjà dépassé : ce standard vient d'être battu par une jeune équipe de brillants chercheurs réunis autour de Stéphane Mallat. Ce groupe arrive à comprimer les photos d'identités en n'utilisant pas plus de 500 octets ! Mais pour décrire cet exploit il faut revenir aux débuts de l'épopée des ondelettes. Nous parlerons ensuite des problèmes de compression de l'image et cela nous amènera tout naturellement aux travaux de l'équipe de Mallat.

Tout débuta avec Jean Morlet

Ancien élève de l'École polytechnique, Morlet était ingénieur de recherche chez Elf-Aquitaine. Quand il découvrit les ondelettes, Morlet travaillait depuis déjà une vingtaine d'années dans le secteur de la vibrosismique. Morlet créa l'analyse par ondelettes pour surmonter certaines difficultés rencontrées dans l'analyse des signaux acquis lors des campagnes pétrolières.

Autrefois, pour chercher du pétrole, on faisait exploser des charges et les échos recueillis permettaient d'estimer la position, la profondeur et la forme de la cavité contenant l'or noir. Les experts engagés par les compagnies pétrolières, les sourciers, étaient alors des physiciens. Analyser les bruits répercutés par le sous-sol, c'était imiter le savoir-faire du médecin qui, à l'aide du stéthoscope, ausculte le malade en écoutant sa respiration ou les battements de son cœur.

Pierre Goupillaud, collègue et ami de Jean Morlet, a contribué à l'amélioration de la recherche pétrolière. Pierre Goupillaud travailla pour la compagnie pétrolière Conoco, (aujourd'hui ConocoPhillips) dans le secteur de la géophysique. Goupillaud suggéra d'envoyer dans le sous-sol une vibration, courte et modulée en fréquence, au lieu de faire exploser des charges. L'énergie dépensée et les dégâts occasionnés sont alors réduits. Ce même principe est utilisé par le sonar de la chauve-souris. La vibrosismique était née. Mais les échos recueillis sont bien plus complexes à analyser que dans le cas des explosions de charges. Les physiciens durent céder la place à des spécialistes du traitement du signal. Ces derniers élaborèrent des logiciels informatiques qui, en un sens, imitent le fonctionnement du cerveau de la chauve-souris. Grâce à la vibrosismique, Elf-Aquitaine a pu mener une campagne de prospection pétrolière à Paris même. Les camions-vibrateurs ont sillonné les artères parisiennes pendant une quinzaine de jours, au milieu des nuits d'hiver de février 1986. L'exploitation des résultats a demandé une année entière. Cela donne une idée des difficultés rencontrées dans la vibrosismique.

Jean Morlet analysait donc les signaux provenant de la vibrosismique. Ces signaux sont des courbes graphiques assez irrégulières qui présentent, en outre, de fortes parties transitoires (ce terme sera défini plus tard). Jean Morlet étudiait ces courbes à l'aide d'une technique éprouvée, l'analyse de Fourier à fenêtre (en fait à l'aide des ondelettes de Gabor, nous y reviendrons). Mais un jour, Morlet, lassé des artefacts (erreurs systématiques) entraînés par cette technique, découvrit une nouvelle façon de représenter ce type de signaux. Le nouvel algorithme fournit un meilleur rendu des zones transitoires.

C'est ainsi que Morlet créa l'analyse par ondelettes.

Une définition mathématique de l'analyse par ondelettes est présentée dans l'appendice.

Rappelons qu'analyser signifie décomposer un tout en ses éléments constituants. La qualité et la précision de l'analyse dépendent du choix des éléments constituants ou *atomes*. Un choix qui convient à une collection de signaux peut être déplorable dans un autre cas. En 1807, Joseph Fourier crée l'analyse harmonique où les éléments constituants d'une fonction, d'un signal ou d'une image sont les ondes $\exp(i\omega \cdot x)$. On reconstruit le signal en combinant toutes ces ondes après les avoir multipliées par des coefficients $\gamma(\omega)$ qui sont, au facteur $(2\pi)^{-n}$ près, donnés par la corrélation $\int f(x) \exp(-i\omega \cdot x) dx$ entre la fonction $f(x)$ que l'on analyse et l'onde utilisée. Pour Dennis Gabor, les éléments constituants seront les *ondelettes*, c'est-à-dire des petits morceaux d'ondes. L'analyse et la synthèse s'opèrent alors comme dans l'analyse de Fourier. Les ondelettes de Gabor ont une durée (ou un diamètre) fixe, mais une fréquence arbitraire; leur expression analytique est $w(x - x_0) \exp(i\omega \cdot x)$ où la fenêtre w est le plus souvent une gaussienne. Dans le cas d'une image, x est un point du plan; alors ω est un vecteur et $\omega \cdot x$ désigne un produit scalaire. Dans le cas d'un signal, c'est-à-dire d'une fonction d'une variable réelle, la variable est le temps, noté t . On analyse un signal ou une image $f(x)$ à l'aide de ces ondelettes en calculant les corrélations $W(x, \omega) = \int f(y) w(y - x) \exp(-i\omega \cdot y) dy$. Cela revient à déployer le signal dans le *plan temps-fréquence*. Cela revient aussi à faire une *analyse de Fourier à fenêtre* : on segmente d'abord le signal ou l'image en les délimitant à l'aide de la

fenêtre glissante $w(y - x)$; on effectue ensuite une analyse de Fourier ordinaire sur chaque morceau $f(\cdot)w(\cdot - x)$ du signal ou de l'image. L'analyse de Fourier convient aux signaux stationnaires, tandis que les ondelettes de Gabor sont l'un des outils que l'on utilise pour analyser les signaux quasi-stationnaires (comme le signal de parole). C'est pourquoi, Dennis Gabor appelait *logons* ce que nous appelons les ondelettes de Gabor.

Les ondelettes de Morlet diffèrent des précédentes en ce que leur durée est arbitrairement petite, ce qui permet de mieux cerner les phénomènes transitoires. On pense à l'attaque (qui dure quelques millisecondes) d'une note à la clarinette. L'analyse de Fourier est adaptée à l'étude de la note, une fois formée, mais pas à celle de son attaque. Les ondelettes de Morlet sont aussi localisées en fréquence autour d'une fréquence moyenne qui est l'inverse de leur durée. Tout naturellement les fréquences utilisées dans les algorithmes dérivés des ondelettes de Gabor forment une progression arithmétique (comme c'est le cas dans les séries de Fourier) tandis que celles intervenant dans les algorithmes utilisant les ondelettes de Morlet sont en progression géométrique (comme dans l'analyse de Littlewood-Paley). Cette dernière joue, pour l'analyse mathématique, le même rôle que les ondelettes de Morlet en traitement du signal.

Parlons encore un peu de mathématique ! L'illustration la plus simple de la supériorité des ondelettes de Morlet sur l'analyse de Fourier est fournie par la théorie de l'approximation non-linéaire. Comme il est bien connu, si $f(x)$ est une fonction continue et 2π -périodique de la variable réelle x , et si α est un exposant positif non entier, alors les deux propriétés suivantes sont équivalentes :

- (a) pour une certaine constante C et tout entier N , on peut approcher $f(x)$ par une somme trigonométriques $P_N(x) = \sum_{-N}^N a_{k,N} \exp(ikx)$ avec une précision $O(N^{-\alpha})$ (la qualité de l'approximation est mesurée en norme uniforme),
- (b) $f(x)$ est uniformément hölderienne d'exposant α .

Mais supposons, en outre, que $f(x)$ soit indéfiniment dérivable en dehors d'un ensemble fini de singularités de type cusp; c'est-à-dire qu'au voisinage de x_j on ait $f(x) = c_j|x - x_j|^{\alpha_j} + g(x)$ où $1 \leq j \leq N$ et $\alpha_j \geq \alpha$ et où $g(x)$ est indéfiniment dérivable. Utilisons alors, pour approcher $f(x)$, des fractions rationnelles $R_N(z) = \frac{P_N(z)}{Q_N(z)}$ en la variable $z = \exp(ix)$, en exigeant que les pôles de R_N n'appartiennent pas au cercle $|z| = 1$ et que les degrés de P_N et de Q_N ne dépassent pas N . Alors on passe d'une vitesse d'approximation lente en $O(N^{-\alpha})$ à une vitesse rapide, en $O(N^{-q})$ pour tout entier q . Les fractions rationnelles en question $\frac{P_N}{Q_N}$ sont des combinaisons linéaires de N éléments simples qui, en un sens, ressemblent à des ondelettes (si les pôles sont suffisamment proches du cercle unité). Ces singularités isolées (cusps) de $f(x)$ sont un exemple de transitoires et cet exemple mathématique nous explique pourquoi l'analyse par ondelettes convient aux signaux présentant de fortes transitoires. Les détails se trouvent dans les travaux de Ronald De Vore cités en référence.

J'ai souvent discuté avec Jean Morlet. Il ressemble beaucoup à Benoît Mandelbrot. Tout comme Mandelbrot, Morlet a une extraordinaire intuition et une réelle

vision scientifique. Il a tout de suite compris la portée de sa découverte et a essayé d'alerter Elf-Aquitaine. Mais Elf-Aquitaine venait d'être la victime consentante d'une énorme escroquerie ; un malfaiteur belge qui aujourd'hui coule des jours heureux au Brésil, était arrivé à persuader les têtes pensantes de l'entreprise que l'on pouvait « flairer le pétrole » à l'aide des trop célèbres « avions renifleurs ». Pour bluffer les « têtes pensantes » et autres « décideurs » d'ELF, l'escroc présentait, dans un certain ordre, des objets dans une pièce. Ces objets étaient « reniflés » par un miraculeux « gadget », situé dans une autre pièce. Ce gadget reconstruisait, en temps réel, les images des objets sur un écran d'ordinateur. Les décideurs d'ELF étaient médusés. L'escroquerie fut révélée par un physicien, Jules Horowitz, qui était à l'époque directeur de la physique au CEA. Jules Horowitz fut invité comme expert pour assister à une démonstration. Quand le « grand patron » et sa suite sont arrivés, pendant les salutations et les politesses entre « inventeurs » et « grand patron », Jules Horowitz s'est glissé dans la chambre de mesure et a tordu complètement, en équerre, la règle que le gadget pouvait voir à distance. Puis, la démonstration a eu lieu. Une image de la règle, bien droite est apparue. Jules Horowitz a conduit le grand patron et les inventeurs à la chambre où était placée la règle. Il a montré qu'elle était pliée. Cela a suffi et tout ELF a été convaincu. Jules Horowitz n'a fait aucun commentaire et est revenu à son travail quotidien.

Passant de l'extrême crédulité à une extrême méfiance, Elf-Aquitaine répondit à la découverte de Jean Morlet en lui octroyant une retraite anticipée. Plus de dix ans après cette mise à la retraite, Morlet obtint le prix Reginald Fessenden de la Société Américaine de Géophysique. Lors de la cérémonie, Pierre Goupillaud présenta l'œuvre de Morlet et dit : « *A product of the renowned École polytechnique, Morlet performed the exceptional feat of discovering a novel mathematical tool which has made the Fourier transform obsolete after 200 years of uses and abuses, particularly in its fast version... Until now, his only reward for years of perseverance and creativity in producing this extraordinary tool was an early retirement from ELF.* »

Jean Morlet est, comme Goupillaud nous le dit, un ancien élève de l'École polytechnique. Roger Balian y enseignait la physique. Morlet lui demanda son avis, car il voulait que son travail sur les ondelettes soit expertisé. Balian l'orienta vers Alexandre Grossmann. Alex Grossmann, directeur de recherches au CNRS, travaillait à Marseille-Luminy, au centre de physique théorique. Grossmann fut patient, subtil et comprit ce que Morlet avait dans l'esprit. Grâce à la clairvoyance de Grossmann, les résultats empiriques de Morlet purent être démontrés et furent publiés en 1984. Écouter Morlet n'était certainement pas une tâche aisée, tant ses idées étaient originales, allusives, approximatives et souvent exagérément optimistes. J'en parle d'expérience. Morlet pensait, par exemple, que l'analyse par ondelettes allait tout de suite révolutionner la vibrosismique et la prospection pétrolière. Quelque chose d'autre s'est produit : les ondelettes ne servent qu'à *comprimer et transmettre* les données recueillies dans les campagnes pétrolières. Ces données sont des signaux 1-D (représentant les variations ou fluctuations d'une fonction du temps). L'analyse de l'information contenue dans ces données est une tout autre histoire. En outre les ondelettes servent à *comprimer et transmettre* les images qui sont bi-dimensionnelles (ou 2-D).

À l'École polytechnique, la photocopieuse que mathématiciens et physiciens partageaient était située dans une petite pièce. Jean Lascoux, directeur du centre de physique mathématique de l'École polytechnique, scientifique érudit et généreux, photocopiait tout ce qui pouvait l'intéresser ou être utile à l'un de ses amis. Quand j'avais aussi des photocopies à faire, j'attendais qu'il termine les siennes et j'en profitais pour discuter avec lui. C'est ainsi qu'il me montra, à l'automne 1984, un preprint de Grossmann et Morlet ouvrant la voie à ce qui allait devenir l'analyse par ondelettes. C'est ainsi que débutèrent mes recherches sur les ondelettes. Je pris alors le train pour Marseille et rejoignis ainsi le « club des ondelettistes ». Ce « club » se composait alors de Grossmann, de Morlet et de quelques jeunes qui deviendront célèbres. Parmi ces jeunes, citons Ingrid Daubechies qui travaillait avec Grossmann sur certains problèmes de la mécanique quantique et qui est aujourd'hui professeur à Princeton.

En arrivant à Marseille, je compris que Grossmann et Morlet avaient redécouvert un théorème qu'un mathématicien argentin, travaillant à l'université de Chicago, Alberto Calderón, avait établi vingt ans avant. La preuve qu'en donnait Grossmann utilisait la théorie des représentations des groupes, théorie qui fonde la mécanique quantique. Mais cette preuve utilisait l'irréductibilité de l'action du groupe affine sur l'espace de Hardy (composé des fonctions holomorphes dans le demi-plan supérieur et de carré intégrable sur la droite réelle). Cette approche limitait l'application de l'analyse de Morlet à la dimension 1, c'est-à-dire aux signaux en excluant les images. En outre le signal analysé devait être analytique, ce qui n'est pas un obstacle. Calderón ne savait pas et n'a jamais cru que son résultat puisse jouer un rôle important dans le traitement du signal. Ceci est d'autant plus surprenant que Calderón avait au départ une formation d'ingénieur. Morlet relia ainsi le théorème de Calderón (qu'il ne connaissait pas) au traitement du signal et la découverte de cette relation est révolutionnaire.

Mais les ondelettes devront muter avant d'envahir la science et la technologie contemporaines. Cette mutation peut être comparée à la découverte de la FFT, ou *fast Fourier transform*. La FFT est un algorithme exact. Il a été découvert, en 1965, aux États-Unis, par James W. Cooley et John W. Tukey. Sans la FFT le calcul d'une transformation de Fourier serait prohibitif. Il faudrait N^2 opérations pour un signal de longueur N . Avec la FFT ceci se réduit à $2N \log_2 N$. Plus concrètement cela revient à comparer un temps de calcul qui ne prend qu'une seconde à un temps de calcul qui prendrait plusieurs semaines. Sans la possibilité qu'offre la FFT de calculer en temps réel, l'imagerie médicale ou l'analyse des images de diffraction des molécules de la biologie n'auraient pas vu le jour.

La révolution numérique dont la FFT est un exemple repose sur l'utilisation d'algorithmes rapides et efficaces. Grossmann et Morlet avaient bien proposé un algorithme pour calculer la transformée en ondelettes. Cet algorithme consistait à remplacer les intégrales doubles (voir l'appendice) qui figurent dans le théorème de Calderón et dont le coût de calcul est élevé par des approximations. Mais ces approximations sont imprécises et engendrent des erreurs incontrôlables.

Un long chemin restait à parcourir avant que l'analyse par ondelettes bénéficie d'un algorithme ayant la précision et la rapidité de la FFT. C'est grâce aux travaux

d'Ingrid Daubechies et de Stéphane Mallat que la Fast Wavelet Transform a finalement pu se hisser au niveau atteint par la *FFT*. Le calcul de la *FWT* d'un signal de longueur N est exact et ne nécessite que CN opérations (C est une constante que nous retrouverons dans ce qui suit).

La mise en œuvre d'algorithmes exacts débuta par ma construction, pendant l'été 1985, d'une base orthonormée d'ondelettes appartenant à la classe de Schwartz. Les intégrales doubles de l'identité de Calderón devinrent alors de simples séries. La théorie des distributions de Laurent Schwartz put enfin s'écrire à l'aide d'algorithmes numériques performants. Plus précisément mes ondelettes sont une suite ψ_λ , $\lambda \in \Lambda$ de fonctions de la classe de Schwartz $S(\mathbf{R}^n)$ et les coefficients d'ondelettes d'une distribution tempérée f sont les produits scalaires $c_\lambda = \langle f, \psi_\lambda \rangle$. L'espace des distributions tempérées devient un espace de suites et l'appartenance de f aux espaces fonctionnels utilisés en analyse est déterminée par des conditions simples portant sur les modules des c_λ , $\lambda \in \Lambda$. En d'autres termes, les ondelettes constituent une base inconditionnelle universelle.

Du même coup, les calculs approchés de Grossmann et Morlet furent enfin remplacés par des formules exactes. Mais mon algorithme reposait sur des *filtres de longueur infinie*, ce qui n'est pas acceptable en traitement du signal.

C'est alors que Stéphane Mallat, alors qu'il était encore étudiant en thèse à Philadelphie, construisit, en 1986, la *FWT*, c'est-à-dire la *Fast Wavelet Transform*. Mallat travaillait sur l'analyse et la *compression* des images. Comprimer un signal ou une image, c'est « dire la même chose de façon plus concise ». Pour comprimer, il convient de comprendre. Nous y reviendrons.

La construction de la *FWT* bénéficiait de deux découvertes antérieures : les *algorithmes pyramidaux* et le *codage en sous-bandes*. Nous devons à Mallat d'avoir compris que les algorithmes pyramidaux, les ondelettes et le codage en sous-bandes sont apparentés. Les *algorithmes pyramidaux* furent découverts par P. Burt et E. H. Adelson (1983), dans le cadre du traitement de l'image. Un algorithme pyramidal est un algorithme rapide fournissant des vues de plus en plus synthétiques d'une image ; les mêmes que celles que l'on obtiendrait en s'éloignant d'une scène tout en continuant à l'examiner. Les échelles successives utilisées dans un algorithme pyramidal sont 1, 2, 4, 8, 16, 32, etc. Ces algorithmes pyramidaux sont rapides et importants. Ils sont rapides, car ils utilisent un schéma itératif qui fournit directement l'image simplifiée I_{n+1} à l'aide de I_n , sans que l'on ait besoin de revenir à l'image de départ. Ils sont importants, car ils permettent de mieux percevoir et comprendre une image. On rejoint ici la sagesse populaire qui enjoint de *prendre du recul* pour accéder à l'essentiel. L'analyse par ondelettes d'une image devient, dans les travaux de Mallat, un cas particulier des *algorithmes pyramidaux*. Chez Mallat, comme chez Burt et Adelson, analyser une image par ondelettes revient à étudier cette image en prenant du recul. Ce recul permet de comparer mentalement différentes vues à différentes échelles. Mais l'algorithme de Mallat va plus loin, car il fournit la différence entre les vues à deux échelles consécutives. Cette différence est fournie par l'analyse par ondelettes. La seconde découverte utilisée par Mallat dans la *FWT* est le *codage en sous-bandes* ou *subband coding*. Ce codage avait été inventé en 1977, au centre d'IBM, à La Gaude par D. Esteban et C. Galand.

La motivation était le téléphone digital. L'importance de la découverte du codage en sous-bande n'a pas été perçue par IBM. Ni Grossmann, ni Morlet, ni moi-même n'étions au courant de cette découverte importante qui précédait nos travaux. C'est Mallat qui a compris que les bases orthonormées d'ondelettes et le *subband coding* pouvaient, dans la plupart des cas, être identifiés. Les spécialistes du traitement du signal ou de l'image qui rejetaient les ondelettes de Grossmann-Morlet furent obligés de se rendre à l'évidence : sous un léger déguisement, les ondelettes existaient bel et bien en traitement du signal. Mallat a ainsi provoqué un appel d'air. Il a ouvert portes et fenêtres et a créé une discipline scientifique nouvelle où la physique théorique, les mathématiques, l'algorithmique, le traitement du signal et le traitement de l'image se sont rejoints.

La thèse de Mallat fut le point de départ du standard JPEG2000 de compression des images fixes.

Puis vint la construction, par Ingrid Daubechies (1987), des bases orthonormées d'ondelettes à support compact, de régularité r donnée. Cette régularité peut être aussi élevée que l'on veut, mais la base choisie dépend alors de r . Avant cette découverte, le seul cas connu était celui du système de Haar (1909) qui n'a pas de régularité. La longueur du support de l'ondelette est la constante C intervenant dans le nombre CN d'opérations. Les ondelettes à support compact conduisent à des algorithmes qui fonctionnent en temps réel, alors même que le signal défile. Le calcul se fait sur une *fenêtre mobile*, ne mettant en jeu qu'un nombre fixe, limité de valeurs du signal. Dans le cas du système de Haar, l'algorithme calcule la demi-somme et la demi-différence entre deux valeurs consécutives. Mais le manque de régularité ($r = 0$) du système de Haar excluait toute application à la compression des images fixes. L'année suivante, en collaboration avec Albert Cohen et Jean-Christophe Feauveau, I. Daubechies construisit les ondelettes bi-orthogonales. Ce sont elles qui seront utilisées dans JPEG2000.

Ondelettes, JPEG et JPEG2000

Aujourd'hui le « web », la toile, envahit notre vie et modifie même l'économie. Le téléchargement du réseau vers l'utilisateur d'une image est beaucoup plus lent que celui d'un texte. Transmettre les images à travers la toile, en utilisant les capacités limitées des câbles de transmission (parfois des lignes téléphoniques ordinaires) reste un défi. En effet numériser une image coûte un prix qui est souvent prohibitif. Rappelons qu'une seule image 1024×1024 en noir et blanc, dont le niveau de gris est codé sur huit bits, coûterait, si elle n'était pas comprimée, 1.048.576 octets, c'est-à-dire autant que l'intégralité des Essais de Montaigne. L'encombrement des autoroutes de l'information interdit de telles dépenses. Cet encombrement avait été prévu et calculé par Claude Shannon, il y a plus de cinquante ans. Shannon a évalué le nombre maximum de bits que l'on peut transmettre par seconde à l'aide d'un câble ayant une fréquence de coupure donnée. La fréquence de coupure est dictée par la technologie de transmission et l'usage des fibres optiques accroîtra la bande passante : l'autoroute de l'information acquiert alors une voie supplémentaire, ce qui accélère le débit.

Pour transmettre des images de la façon la plus efficace à travers un canal de transmission donné, il nous faut donc gérer au mieux le budget en bits qui nous est

alloué pour représenter une image. Comprimer un signal ou une image, c'est 'dire la même chose' (dans la compression sans perte ou *lossless*) ou 'dire quelque chose d'équivalent' (compression avec perte ou *lossy*) en moins de mots. La compression sans perte s'apparente à la sténographie, une technique pour prendre des notes que les secrétaires d'autrefois maîtrisaient. Nous nous concentrerons sur la compression avec perte. On rejoint ici le problème de la traduction d'une langue dans une autre. Les logiciels de traduction automatique fournissent encore des résultats décevants, car on ne peut traduire un texte sans l'avoir compris. Les logiciels n'ont pas accès au sens. De même, pour savoir ce que l'on peut perdre et sacrifier dans une image, il faut la comprendre, c'est-à-dire accéder à son contenu sémantique. C'est pourquoi les radiologues se méfient des algorithmes de compression, car ils les soupçonnent de faire disparaître des détails essentiels à leurs yeux. On ne peut compresser que si l'on a une certaine idée de ce que l'image contient ; en termes plus techniques, si l'on dispose d'un *modèle*. Il n'y a pas de modèle universel et il faudrait, en bonne logique, élaborer un modèle approprié à chaque collection d'images. Nous rejoignons alors un secteur très vivant des statistiques contemporaines : la théorie de l'apprentissage.

Les enjeux scientifiques et industriels de la compression des images fixes sont énormes. Comme Jacques Blamont le rappelait à l'Académie des sciences (séance du 25 septembre 2000), l'exploration de l'Espace passe par la compression des images, car ces images, grandes dévoreuses de bits, il faut les retransmettre à la Terre, tout en réduisant au maximum l'énergie nécessaire à cette transmission. Cette énergie est fournie par des équipements embarqués et il importe de réduire le plus possible la masse du satellite .

Il faut donc compresser les images pour les transmettre et c'est là que les ondelettes entrent dans la danse. En fait les ondelettes ne sont efficaces que si les images dont on s'occupe peuvent être représentées ou modélisées par ce que j'appellerais le croquis d'un dessinateur. Plus précisément, le modèle utilisé suppose que les images se composent d'un certain nombre de lignes délimitant des objets. Les lignes peuvent être assez irrégulières mais la longueur totale de l'ensemble de ces lignes est l'une des contraintes du modèle. L'autre est le fait que l'image est régulière à l'intérieur des domaines (ou objets) délimités par les lignes. Ces lignes définissent un croquis ou sketch noté u . Les croquis que nous venons de définir appartiennent à l'espace BV des fonctions à variation bornée. Le modèle est complété par une seconde composante de l'image, notée v , qui représente la texture et le bruit. L'image complète est alors la somme $f = u + v$. Les ondelettes ne sont pertinentes que si elles sont appliquées à ce type d'images.

Naturellement on n'avait pas attendu les ondelettes pour compresser les images et l'algorithme JPEG2000 doit être comparé à JPEG, algorithme utilisé jusqu'alors. JPEG ne repose pas sur un modèle préalable. C'est un algorithme robuste, 'tout-terrain' dont les réglages ont été fixés une fois pour toutes et ont été testés sur un stock standardisé d'images. L'algorithme JPEG peut être regardé comme une variante de l'analyse de Fourier à fenêtre. Il opère

- (1) un découpage de l'image en blocs 8×8 ,
- (2) une transformation de Fourier rapide (*FFT*) sur chaque bloc et

(3) une quantification adaptée des coefficients de Fourier ainsi calculés. La quantification est l'opération consistant à remplacer les nombres réels par des valeurs digitales approchées; elle peut être uniforme ou peut dépendre de la fréquence analysée. En fait on utilise la *DCT*, proche parente de la *FFT*.

Comme le disait avec force Martin Vetterli, il fallait être totalement inconscient pour essayer de vaincre JPEG qui reposait sur la *FFT* et avait ensuite bénéficié de 25 ans de perfectionnements. Mais la raison était de notre côté. Voici pourquoi.

Une image fournit des informations intéressantes à toutes les échelles. La notion d'échelle semble donc plus pertinente que celle de fréquence pour décrire les images. L'idéal serait de relier échelle et fréquence. C'est exactement ce que permet de faire l'analyse par ondelettes qui fournit l'équation $\omega = 1/a$ où a désigne l'échelle et ω la fréquence. Un cas particulier de cette équation nous dit que la fréquence est l'inverse de la longueur d'onde. Retournant à l'analyse par ondelettes, elle permet de décomposer une image en utilisant des canaux fréquentiels indépendants couvrant à peu près une octave : les fréquences (spatiales) y vont de 2^j à 2^{j+1} . À l'intérieur de chacun de ces canaux, on peut, à l'aide du théorème de Shannon, échantillonner en variable d'espace avec un pas d'échantillonnage égal à 2^{-j} ; la grille d'échantillonnage est alors $2^{-j}Z^2$. Les ondelettes fournissent cette double localisation : la première en fréquence et la seconde en variable d'espace.

Un second atout en faveur des ondelettes est le suivant : l'utilisateur de l'image peut zoomer sur une partie de l'image qui l'intéresse et dépenser son allocation en bits en optimisant ce zoom, un peu comme un peintre qui, faisant un portrait, accorderait plus d'importance aux yeux qu'à des détails de l'habillement. Cette capacité de zoomer provient de la construction même des ondelettes et est totalement absente dans l'analyse de Fourier ou dans l'ancien JPEG. Les zooms sont déjà apparus dans notre discussion lorsque nous avons évoqué les algorithmes pyramidaux, lointains ancêtres des ondelettes.

Par ailleurs certaines cellules du cortex visuel primaire fournissent une information qui ressemble à celle que l'on retire des coefficients d'ondelettes de l'image observée. Ces *coefficients d'ondelettes* ne sont pas ceux que nous avons utilisés précédemment, mais doivent être pris en un sens qui annonce les travaux de David Donoho et de la start-up *Let It Wave*. Citons d'abord David Hubel, prix Nobel de médecine : « *Cells in the primary visual cortex, to which the optic nerve projects (with one intermediate nucleus interposed) are far more exacting in their stimulus requirements. The commonest type of cells fire most vigorously not to a circular spot, but to a short line segment, to a dark line, a bright line, or to an edge boundary between dark and light. Furthermore each cell is influenced in its firing only by a restricted range of line orientations : a line more than about 15 to 30 degrees from the optimum generally evokes no response. Different cells prefer different orientations, and no one orientation, vertical, horizontal or oblique, is represented more than any other...Evidently cells in this part of the cortex are determining whether there are contours (light-dark or color) in the visual scene, and collectively registering their orientations.* »

Les travaux, en collaboration avec Torsten Wiesel et Margaret Livingstone décrits ici par David Hubel portent sur les chats et les singes. Cependant ils s'appliquent

au système visuel humain. On dispose aujourd'hui de preuves indirectes de ce fait. La première leçon que l'on tire de ces découvertes est que le système visuel humain *ne calcule pas une transformée de Fourier*. David Marr était arrivé à la même conclusion. La seconde est que l'image n'est pas perçue comme un tout par le système visuel, mais bien décomposée en entités élémentaires qui sont acquises de façon indépendante. La dernière leçon est le rôle essentiel joué par la géométrie.

Émuler le système visuel humain pourrait alors être la stratégie idéale pour résoudre le problème de la compression, car les erreurs impliquées par cette compression seraient alors les mêmes que celles que l'œil aurait commises, elles seraient « masquées » et perceptuellement tolérables.

C'est pour toutes ces raisons que l'analyse par ondelettes a gagné. Il a cependant fallu transformer une argumentation scientifique en une réalité technologique. Ce ne fut pas une mince affaire et l'on se doit d'admirer le remarquable travail effectué par les chercheurs qui ont œuvré au programme JPEG2000. Mais certains de ces arguments militent encore plus fortement en faveur de travaux plus récents (Donoho et Mallat).

500 octets suffisent, car les ondelettes ne sont pas optimales

Pendant les années où JPEG2000 a été élaboré, la « quête du graal » a continué. Il s'agissait de faire mieux que les ondelettes dans la course à la compression. Plusieurs groupes ont ainsi cherché à mieux tirer parti des particularités géométriques des images. Un des modèles utilisés consiste à décrire les images en imitant le peintre Ingres. Une image est alors un ensemble de lignes assez régulières délimitant des zones où l'éclairement varie peu. Ces lignes seront appelées des bords. Si l'on utilise l'analyse par ondelettes pour décrire une telle image et si l'on désire localiser les bords d'un objet avec une précision de l'ordre de $1/\sqrt{N}$, il faudra utiliser au moins N ondelettes distinctes, ce qui revient à coder les positions d'au moins N points appartenant à un voisinage du bord. L'erreur est ici mesurée en valeur quadratique moyenne. Il est clair que ce codage conduit à un gaspillage et que la même précision peut être obtenue en utilisant seulement \sqrt{N} données. Il suffit pour le voir de fournir les positions de \sqrt{N} points des bords et des \sqrt{N} tangentes à ces points. En utilisant l'approximation des bords par les cercles osculateurs, on fait encore mieux et tout dépend ici de la régularité supposée des bords. Les ondelettes, quant à elles, sont bien sensibles à la présence de bords, mais n'arrivent pas à suivre la direction de ce bord, car elles sont 'isotropes'. Elles ne comportent pas un paramètre indiquant une direction. Seules la position et l'échelle sont prises en compte.

La première percée dans la direction de l'anisotropie a été réalisée par Emmanuel Candès et David Donoho dans un travail retentissant. Ces deux chercheurs ont créé les *ridgelets* qui sont des *ondelettes de seconde génération*. Les *ridgelets* et leurs 'cousines' les *curvelets* forment une base de l'espace de référence (en fait, un *frame* dans les cas des *curvelets*). En outre, cette base est adaptative, ce qui semble une contradiction dans les termes. Lorsqu'on analyse une image géométrique dans cette nouvelle base, les *curvelets* semblent s'orienter et s'allonger en épousant automatiquement la géométrie d'un bord éventuel. La construction des *ridgelets* repose sur celle des ondelettes.

Stimulé par les travaux de Candès et Donoho, Stéphane Mallat a fourni une solution plus réaliste en inventant les *bandelettes*. Elles aussi épousent le plus longtemps possible le tracé des bords des images. La nouvelle technologie proposée par Mallat est développée par la start-up *Let It Wave* et s'appelle *Let It Wave Codec*. *Let It Wave* a obtenu le premier prix de l'innovation technologique 2002, décerné par le Ministre de la Recherche et de la Technologie.

Quelques mots sur ce dernier algorithme. Le point de départ théorique est la thèse d'Erwann Le Pennec. Dans ce travail, une image est modélisée comme il vient d'être indiqué. Rappelons que les objets contenus dans l'image sont délimités par des bords nets, que l'on peut dessiner comme le ferait Ingres. L'éclairage de l'objet est supposé très régulier. À cette image brute, assez schématique est ajouté un bruit aléatoire. Le problème posé est la description la plus concise, la plus économique, de ce type d'images. L'algorithme utilisé par Le Pennec est hybride, car il mêle les *level sets* de Jean-Michel Morel et Stanley Osher aux ondelettes. Le point de départ est la recherche des bords (supposés doux et réguliers) des objets. En fait, on cherche seulement une information directionnelle décrivant les lignes parallèles aux bords. Ces lignes parallèles ressemblent alors aux lignes de niveau d'une carte d'état major. C'est ici qu'apparaissent les 'level sets'. On tire ensuite parti de la régularité de l'éclairage pour l'analyser à l'aide d'ondelettes anisotropes dont les paramètres sont réglés à l'aide de l'information directionnelle déjà obtenue. Les *bandelettes* de Mallat et Le Pennec ne sont pas fixées une fois pour toutes, comme le sont les ridgelets, mais leur construction s'adapte (automatiquement) à l'image analysée. Le passage d'un travail de recherche à l'application industrielle développée par *Let It Wave* n'était pas une mince affaire et le résultat final obtenu par l'équipe de Stéphane Mallat est une prouesse.

Appendice : mais que sont les ondelettes ?

Les ondelettes sont l'un des outils utilisés en traitement du signal. Les signaux en question proviennent de mesures ou d'enregistrements. Voici des exemples : électrocardiogrammes, fluctuations des valeurs boursières ou signaux audio (parole, musique). Le traitement du signal a plusieurs buts : la compression, la transmission, l'archivage, l'analyse et le diagnostic. Un exemple est décrit dans *Le premier cercle* de Soljénitsyne. Dans ce magnifique roman, l'auteur nous raconte que la police secrète veut identifier une personne qui a téléphoné d'une cabine publique en déterminant ses 'empreintes vocales'. Cet exemple nous amène à nous pencher sur le signal fourni par la parole. Rappelons que l'enregistrement de la parole est donné par les variations, les fluctuations de la pression de l'air au cours du temps.

Les recherches sur les ondelettes (ici ce mot a un sens très général) ont d'abord porté sur le signal de parole. C'est pourquoi certaines ondelettes étaient appelées des *logons*. Ces recherches répondaient aux besoins de l'industrie téléphonique. Nous y reviendrons. Les linguistes nous enseignent que le signal de parole peut être fragmenté en une suite de phonèmes. Un phonème est 'insécable'. C'est un atome. Il ne peut être fragmenté en unités acoustiques plus brèves. Les phonèmes sont-ils des ondelettes ? Peut-on généraliser à d'autres signaux cette décomposition du signal de parole en 'atomes' ou 'briques élémentaires' ? Ce problème fondamental a reçu des réponses différentes au cours des siècles. Fourier suggérait en 1807 que les 'atomes'

soient les fonctions $f_{(\omega, \varphi)}(t) = \cos(\omega t + \varphi)$, c'est-à-dire des ondes pures. Il énonçait le résultat paradoxal suivant : toute fonction, tout signal, aussi compliqués soient-ils, n'est pas autre chose qu'une combinaison judicieusement pondérée d'ondes $f_{(\omega, \varphi)}$. Mais une telle représentation soulève de nombreux problèmes quand elle est appliquée aux signaux transitoires (la définition de ce mot sera précisée dans quelques lignes). Voici l'un d'eux : alors même que le signal que l'on veut représenter a cessé d'exister, (par exemple, la source a cessé d'émettre) les ondes qui servent à le composer continueront à vibrer. Leur combinaison est alors exactement nulle, par interférence destructive. C'est comme un orchestre qui continuerait à jouer, mais qui produirait le silence. Comme l'écrivait Jean Ville, c'est mathématiquement exact, mais cela n'a aucun rapport avec la réalité. Si le son n'est pas entendu, c'est qu'il n'est pas émis.

L'explication de ces paradoxes vient de ce que l'analyse de Fourier est adaptée aux *signaux stationnaires*. Un signal est stationnaire si rien d'imprévu ne peut advenir. En particulier un signal stationnaire ne peut s'arrêter, s'éteindre. Il est immortel, comme le sont les ondes de Fourier. Un signal stationnaire laisse une part au hasard, mais ces aléas sont aussi prévisibles. Le signal de parole n'est évidemment pas stationnaire. On ne sait pas à l'avance ce que je vais dire. Le signal de parole reste stationnaire pendant de très courts intervalles de temps (une voyelle vibre un peu comme une onde et le chant en tire profit). On dit que le signal de parole est quasi-stationnaire. Ceci ne concerne que les parties tonales et exclut les 'plosives', comme le 't'. C'est pourquoi une analyse plus réaliste que celle de Fourier a été introduite par Dennis Gabor (physicien anglais, né à Budapest, comme von Neumann). Gabor se proposait de trouver une représentation efficace des signaux quasi-stationnaires. Pour cela, il découpa les « ondes » en petits morceaux (comme on découpe un saucisson). Ces morceaux sont donc des « ondelettes de Gabor ». Les ondelettes de Gabor sont les fonctions $\exp(-(t - t_0)^2) \cos(\omega t + \varphi)$ et Gabor découpait les signaux à un certain rythme fourni par :

$$t_0 = 0, \pm 1, \pm 2, \dots, \quad \omega = 0, \pm 2\pi, \pm 4\pi, \dots$$

L'analyse de Gabor telle qu'elle vient d'être définie est incorrecte : les ondelettes de Gabor ne constituent pas un 'frame'. Une suite de vecteurs $e_j \in H$ est un frame de l'espace de Hilbert H si l'application $J : l^2 \mapsto H$ définie par $\alpha_j \mapsto \sum \alpha_j e_j$ est continue et surjective. L'analyse proposée par Gabor a été perfectionnée par Kenneth Wilson, Henrique Malvar, Ronald Coifman et moi-même et donne de très bons résultats dans le traitement des signaux quasi-stationnaires, comme la composante tonale de certains signaux audios. Cette analyse est utilisée dans le son numérique Dolby.

Pour traiter des phénomènes imprévisibles, violents, transitoires, (par exemple les 'plosives' du signal de parole) Morlet eut l'idée de raccourcir arbitrairement la fenêtre gaussienne $\exp(-(t - t_0)^2)$ utilisée par Gabor. Il partit de la fonction $\psi(t) = \exp(-t^2) \cos(5t)$, puis la déplaça (par des décalages temporels) et enfin changea d'échelle. Il obtint alors les fonctions $\psi(\frac{t-t_0}{a})$ où a est arbitrairement petit (l'agrandissement sera $1/a$). L'analyse par ondelettes s'apparente donc à des flashes instantanés, comme ceux produits par le laser femtoseconde.

Morlet comprit que cette analyse convenait aux signaux non-stationnaires et établit l'existence d'un aller-retour 'analyse-synthèse'. L'analyse s'opère en calculant les *coefficients d'ondelette*

$$W(a, b) = \int_{-\infty}^{\infty} f(x) \psi_{(a,b)}(x) dx$$

où $\psi_{(a,b)}(x) = a^{-1/2} \psi((x-b)/a)$. Ici $a > 0$ définit l'échelle ($1/a$ est le grossissement). Pour reconstruire $f(x)$ il suffit de combiner les ondelettes $\psi_{(a,b)}$ en utilisant la pondération donnée par les coefficients $W(a, b)$. En d'autres termes,

$$f(x) = \int_0^{\infty} \int_{\mathbf{R}} W(a, b) \psi_{(a,b)}(x) db da / a^2$$

ce qui est précisément l'identité découverte par Alberto Calderón. On doit ici supposer que ψ est une fonction localisée ($\psi \in L^2$ et $x\psi(x) \in L^2$ suffisent), à valeurs réelles et d'intégrale nulle. En outre, on doit normaliser ψ de façon appropriée. Comme nous l'avons signalé, le calcul de cette intégrale double est numériquement prohibitif et c'est pour cela que la FWT s'est substituée à l'identité de Calderón.

La découverte des bases orthonormées d'ondelettes permet de remplacer ces intégrales par des séries en restant exact. Il vient, pour toute fonction f ,

$$f(x) = \sum_j \sum_k c(j, k) 2^{j/2} \psi(2^j x - k), \quad \text{où } c(j, k) = \int f(x) \psi_{(j,k)}(x) dx.$$

L'ondelette 'mère' ψ peut être choisie dans la classe de Schwartz et l'identité précédente fournit une analyse multiéchelle de n'importe quelle distribution tempérée. En outre tout ceci se généralise en dimension arbitraire et permet de résoudre de façon très efficace les principaux problèmes posés par la théorie de l'approximation non-linéaire (maillages adaptatifs, etc.). On consultera à ce propos les travaux de Ron DeVore (le site est indiqué ci-dessous).

Références

Voici des sites d'où vous pouvez télécharger des informations complémentaires.

En ce qui concerne le standard JPEG2000, le meilleur site est : <http://jj2000.epfl.ch> (l'absence de www n'est pas une erreur) mais on pourra aussi consulter <http://www.jpg.com> (qui est le site de la firme Pegasus) ou <http://www.jpeg.org> (qui est le site officiel du comité JPEG2000)

Les travaux de David Donoho se trouvent à l'adresse suivante : <http://www-stat.stanford.edu/~donoho>

On consultera aussi le site d'Emmanuel Candès : <http://www.acm.caltech.edu/~emmanuel>

On apprend tout sur les ondelettes en consultant le site : <http://www.cmapx.polytechnique.fr/~mallat>

Si l'on s'intéresse aux liens entre la théorie des ondelettes et l'approximation non-linéaire, on consultera :

<http://www.math.sc.edu/~devore>

Les travaux de Mallat sur la compression des photos d'identité se trouvent sur le site : <http://www.letitwave.fr/>

Enfin les mordus d'images astronomiques se régaleront en consultant le site :

<http://oposite.stsci.edu>

qui fournit, en direct, les images du télescope spatial Hubble, ainsi qu'une plongée vertigineuse vers les confins de l'Univers ! Là encore, pas de www.

Bon web !!

Addendum

Compte rendu de la journée « La face cachée des mathématiques »

La manifestation « La face cachée des mathématiques » se déroula le 18 mars 2004, entre 11h30 et 19h, au Centre Georges Pompidou à Paris. Elle était organisée par la Société Mathématique de France, la Société de Mathématiques Appliquées, l'Institut des Hautes Études Scientifiques et la revue *Pour La Science*.

Le programme, qui a été respecté, comportait huit exposés de trente minutes. Chaque exposé était suivi de quinze minutes de débat avec le public.

Les thèmes abordés

Cinq exposés eurent pour point de départ des problèmes actuels :

1. *La compression d'images* : c'était le titre de la conférence d'Y. Meyer.
2. *Le contrôle* : « Contrôle d'oscillateurs classiques et quantiques » par P. Rouchon.
3. *Chaos et contrôle* : « l'exemple du pendule forcé » par J. Hubbard (ceux qui trouvent le terme « pendule forcé » trop technique penseront aux bras des robots).
3. *Cryptage et décryptage* : « Comment concevoir un algorithme de chiffrement rapide et solide ? » par A. Canteaut.
4. *Recherche d'une stratégie optimale dans le cadre de la concurrence* : « Dilemme du prisonnier, théorie des jeux et négociations » par E. Jouini. Il est à noter que la « concurrence » envisagée dans l'exposé pouvait être aussi bien économique (concurrence entre producteurs ou distributeurs ; vente aux enchères) que stratégique (guerre) ou politique (élections).

Deux exposés traitaient de la vie interne des mathématiques et de l'histoire des mathématiques.

- *Théorie des nœuds* par A. Quéguiner.
- *Autour du théorème du point fixe* par J. Mawhin.

Un exposé présentait un exemple d'interaction entre les mathématiques et l'art : *Escher et les vaches qui rient*¹ par P. Stevenhagen

¹ M.C. Escher a écrit, en introduction à son œuvre graphique : « *En exposant mes sens aux énigmes de l'univers, en réfléchissant à ces sensations et en les analysant, je m'approche du domaine des mathématiques. Bien que manquant totalement de connaissances et de formation dans le domaine des sciences exactes, je me sens plus proche des mathématiciens que de mes collègues artistes* ».

Les conférences

Par manque de temps, les exposés de « La face cachée des mathématiques » n'ont pas comporté de démonstration. Tous les conférenciers ont tenté de faire découvrir les idées principales, et de donner un coup de projecteur sur des théories mathématiques sous-jacentes : l'analyse de Fourier pour la compression d'images ; les équations différentielles et systèmes différentiellement plats pour le contrôle ; la théorie des nombres et les courbes elliptiques pour le cryptage. C'était une occasion de découvrir ou de redécouvrir la grande variété de théories mathématiques, et la multiplicité de leurs interactions : lorsque l'on a, par exemple, demandé à E. Jouini quelles mathématiques se trouvaient derrière la théorie des jeux, il a répondu : « *le théorème du point fixe, les probabilités, la géométrie différentielle, etc.* »

À la suite de cette journée, on comprenait mieux la difficulté de planifier les recherches mathématiques, et la nécessaire liberté de choix des sujets de recherche pour les mathématiciens. C'est, par exemple, dans les années 1870 que Tait a entrepris le travail de classification des nœuds. Beaucoup de progrès ont été faits depuis, qui ont donné naissance à la théorie des nœuds. Cette théorie a trouvé récemment des applications en biologie (étude de l'activité biologique de la molécule d'ADN) mais on ne dispose toujours pas d'une table de classification de nœuds complète et sans répétition.

Les conférenciers étaient très différents, tant par l'âge, que par le tempérament qui pouvait se deviner à travers leur exposé et leurs réponses aux questions posées. Ils donnaient, à mon avis, une bonne idée de la diversité des mathématiciens. Un esprit chagrin ajouterait qu'ils n'étaient peut-être pas représentatifs de l'ensemble de la communauté puisqu'il manquait le mathématicien dans la lune qu'on aime à caricaturer dans le grand public.

Le public et les débats

Le grand amphi de Beaubourg était au trois quart plein (environ deux cent cinquante participants), la moitié des auditeurs semblant être des jeunes, accompagnés par un professeur.

Après les exposés, les questions posées furent nombreuses et variées : précisions d'ordre mathématique ; questions sur les applications aujourd'hui de telle ou telle théorie ; interrogations sur le métier de chercheur.

Pour donner une idée de la qualité des débats, je citerai la réponse d'Y. Meyer à la question :

« *Quelle a été votre réaction lorsque l'algorithme de compression d'image JPEG 2000 (issu de la théorie des ondelettes dont il était un des fondateurs et de ses propres travaux sur la compression d'images) a été supplanté par l'algorithme des bandelettes ?* »

Il répondit que la recherche mathématique s'apparente au combat de Jacob avec l'ange : « *On sait que l'on sera battu, mais on aime se battre*² ».

² Ce combat est décrit dans la *Bible*, au livre de *La Genèse*, ch. 31, versets 23 à 32.

Géométrisation des variétés de dimension 3 via le flot de Ricci¹

Michael T. Anderson²

Introduction

La classification des surfaces fermées a été une étape importante du développement de la topologie, tellement importante qu'elle est maintenant enseignée au cours des premières années d'université en guise d'introduction à la topologie. Depuis la résolution du problème de l'uniformisation des surfaces par Poincaré et Koebe, cette classification topologique est mieux comprise en termes de géométrisation des surfaces : toute surface fermée Σ admet une métrique à courbure de Gauss constante qui vaut soit $+1$, soit 0 , soit -1 , et ainsi elle est « uniformisée » par l'une des surfaces modèle \mathbb{S}^2 , \mathbb{R}^2 ou \mathbb{H}^2 . Ainsi toute surface Σ est le quotient soit de \mathbb{S}^2 , soit de \mathbb{R}^2 , soit de \mathbb{H}^2 par un groupe discret Γ agissant librement et isométriquement.

La classification des variétés de dimension plus grande est, bien sûr, beaucoup plus difficile. En effet, à cause de la complexité du groupe fondamental, une classification complète comme dans le cas des surfaces n'est pas possible en dimension supérieure ou égale à 4. En dimension 3 cet argument ne s'applique pas et la classification complète des variétés de dimension 3 est depuis longtemps le rêve des topologues. Un cas particulier est d'ailleurs la conjecture de Poincaré.

Dans cet article nous évoquerons le remarquable travail récent de Grisha Perelman [15]-[17], qui pourrait bien avoir résolu le problème de la classification des variétés de dimension 3 (dans un sens très naturel). Le travail de Perelman est en ce moment l'objet d'intenses investigations et vérifications par plusieurs groupes de chercheurs à travers le monde. À l'heure qu'il est, l'essentiel de ce travail a été validé par les experts de ce domaine. Même s'il est trop tôt pour annoncer une solution définitive à ce problème, on peut affirmer que les idées de Perelman sont hautement originales et très profondes. De plus, ses résultats sont déjà utilisés par d'autres chercheurs dans d'autres domaines. Tout ceci justifie, même à ce stade, l'écriture d'un article, ce qui pourrait paraître prématuré autrement.

Le travail de Perelman est bâti sur des travaux antérieurs de Thurston et Hamilton. Dans les deux sections suivantes nous présenterons les points de vue de Thurston sur les variétés de dimension 3 et du flot de Ricci introduit et étudié par Hamilton. Pour des informations complémentaires, en particulier sur la conjecture de Poincaré on renvoie à l'article de Milnor [14] (et qui a été traduit par la Gazette des Mathématiciens dans le numéro 99) ainsi qu'aux références contenues dans cet

¹ Ce texte a été publié par les *Notices* de l'AMS volume 51, numéro 2, sous le titre "Geometrization of 3-manifolds via the Ricci flow". Il a été traduit pour la *Gazette* par Zidine Djadli (Université de Cergy-Pontoise).

² E-mail : anderson@math.sunysb.edu

article. Pour plus d'informations et de détails sur le travail de Perelman on renvoie à [13].

La conjecture de géométrisation

Alors que la conjecture de Poincaré date d'environ un siècle, les remarquables idées de Thurston à la fin des années 70 ont amené à une possible et réaliste classification des variétés fermées de dimension 3 de manière analogue à la classification des surfaces via le théorème d'uniformisation.

Pour expliquer cela, nous avons besoin de préciser ce que sont les « géométries » correspondantes en dimension 3. En terme de géométrie riemannienne, une structure « géométrique » sur une variété M est la donnée d'une métrique riemannienne g complète et localement homogène. Ainsi M peut être décrite comme un quotient $\Gamma \backslash G/H$, où G est le groupe d'isométries du revêtement universel (\tilde{M}, g) et Γ et H sont des sous-groupes du groupe de Lie G , Γ étant discret et H étant compact. Thurston a montré qu'il n'y a que 8 quotients de ce type (G/H) qui soient simplement connexes et qui admettent des quotients compacts³.

Comme en dimension 2, les géométries les plus importantes sont celles qui sont à courbure constante : géométrie hyperbolique \mathbb{H}^3 de courbure -1 , géométrie euclidienne \mathbb{R}^3 de courbure 0 et géométrie sphérique \mathbb{S}^3 de courbure $+1$. Les cinq autres géométries modèles sont des produits ou des produits tordus de surfaces. Les fibrés \mathbb{S}^1 triviaux sur une surface de genre $g > 1$ ont une géométrie $\mathbb{H}^2 \times \mathbb{R}$, tandis que les fibrés non triviaux ont une géométrie $\tilde{S}L(2, \mathbb{R})$; les fibrés \mathbb{S}^1 non triviaux sur \mathbb{T}^2 ont une *Nil*-géométrie, tandis que les fibrés \mathbb{T}^2 sur \mathbb{S}^1 ont une *Sol*-géométrie (ou une *Nil*-géométrie ou une géométrie de type \mathbb{R}^3); finalement, les fibrés \mathbb{S}^1 sur \mathbb{S}^2 ont une géométrie du type $\mathbb{S}^2 \times \mathbb{R}$ (ou \mathbb{S}^3). Par exemple, toute variété de dimension 3 Seifert fibrée (une variété qui admet une action de \mathbb{S}^1 localement libre), a une structure géométrique.

Les variétés « géométriques » de dimension 3, c'est-à-dire les variétés de dimension 3 admettant une structure géométrique, sont les briques élémentaires pour construire des variétés de dimension 3 plus compliquées. Pour simplifier on supposera dans cet article que toutes les variétés considérées sont orientables. Les briques élémentaires sont donc « assemblées » le long de sphères de dimension 2, en utilisant l'opération de somme connexe, et le long de tores de dimension 2. Comme exemple d'un tel assemblage, soit $\{M_i\}$ une famille finie de variétés Seifert fibrées de dimension 3 sur des surfaces Σ_i à bord non vide, de tel façon que ∂M_i soit un tore pour tout i . Ces tores peuvent être « collés » les uns aux autres par paire en utilisant des difféomorphismes pour obtenir une variété fermée de dimension 3 ou une variété de dimension 3 à bord torique. Une variété de dimension 3 assemblée de cette façon est appelée une variété graphe (on associe un sommet à chaque espace Seifert fibré et une arête à chaque tore qui relie deux espaces Seifert fibrés). Un fibré en tore au-dessus de \mathbb{S}^1 est une variété graphe puisqu'union de deux espaces Seifert fibrés au-dessus de $\mathbb{S}^1 \times I$. Les variétés graphes ont été introduites, et complètement analysées par Waldhausen.

³ La classification de Thurston est essentiellement un cas particulier de la classification plus ancienne de Bianchi des métriques homogènes apparaissant en relativité générale; voir [3] pour des remarques supplémentaires sur la correspondance entre ces classifications

Réciproquement soit M une variété fermée de dimension 3 quelconque (mais orientable comme on l'a dit précédemment). On peut alors décomposer ou éclater la variété en pièces élémentaires, à savoir des sphères ou des tores. Topologiquement ceci peut se faire grâce au résultat suivant de topologie des variétés de dimension 3.

Décomposition en sphère (ou prime decomposition) (d'après Kneser, Milnor)

Soit M une variété fermée de dimension 3. Alors M peut être décomposée comme somme connexe

$$(1) \quad M = (K_1 \# \dots \# K_p) \# (L_1 \# \dots \# L_q) \# (\#_1^r \mathbb{S}^2 \times \mathbb{S}^1).$$

Les facteurs K et L sont des variétés fermées de dimension 3 et qui sont irréductibles, c'est-à-dire que toute sphère de dimension 2 qui y est plongée borde une boule de dimension 3. Les facteurs K ont un groupe fondamental infini et sont asphériques, tandis que les facteurs L ont un groupe fondamental fini et leur revêtement universel est une sphère d'homotopie. Puisque $M \# \mathbb{S}^3 = M$ on suppose qu'aucun facteur L n'est une sphère sauf si M est elle-même une sphère. Les facteurs dans cette écriture sont ainsi uniques à une permutation près et sont obtenus à partir de M par chirurgie sur une famille de sphères de dimension 2 essentielles (i.e. topologiquement non triviales) qui sont incluses dans M (en remplaçant les régions du type $\mathbb{S}^2 \times I$ par deux copies de B^3). Voir la figure 1 pour une représentation schématique de cela.

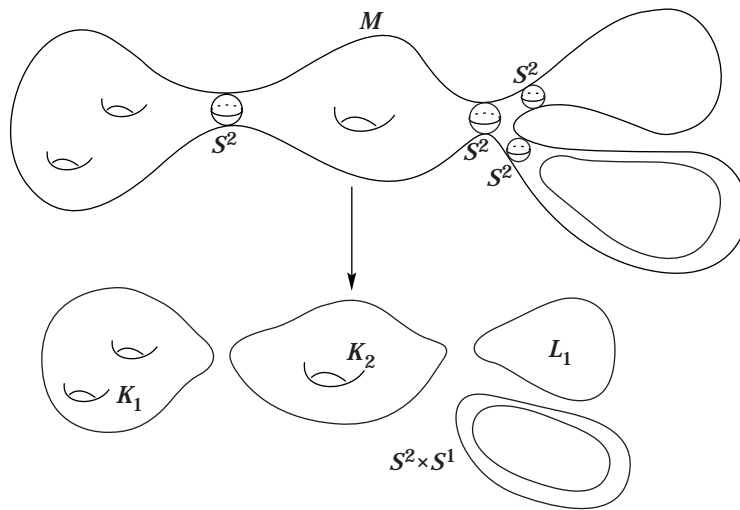


FIG. 1. Décomposition en sphères

Les facteurs K peuvent aussi contenir des tores topologiquement essentiels. Un tore \mathbb{T}^2 plongé dans M est dit incompressible si l'inclusion canonique induit une injection sur Π_1 . Une variété de dimension 3 N est dite toriquement irréductible si tout tore plongé incompressible peut être déformé en un tore dans ∂N . Ainsi si $\partial N = \emptyset$ alors N ne possède aucun tore incompressible.

Décomposition en tore (Jaco-Shalen, Johannsen)

Soit M une variété de dimension 3, fermée, irréductible. Alors il existe une famille finie, éventuellement vide, de tores disjoints incompressibles dans M qui séparent M en une collection finie de variétés de dimension 3 compactes (à bord torique), et qui sont chacun soit toriquement irréductible soit Seifert fibré.

Une décomposition plus complexe, mais essentiellement équivalente, est donnée par des tores séparant M en des composantes toriquement irréductible ou des variétés graphes. Voir la figure 2.

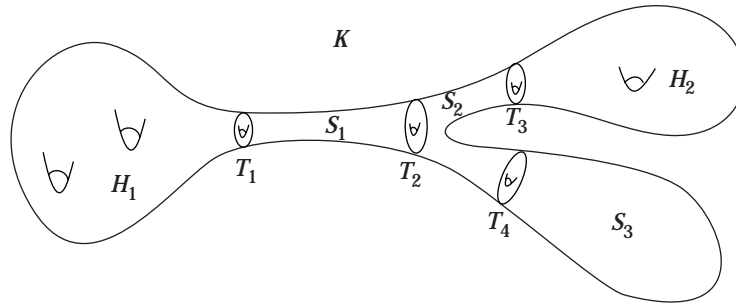


FIG. 2. Décomposition en tores

À l'exception de $\mathbb{S}^2 \times \mathbb{S}^1$ et son quotient $\mathbb{R}P^3 \# \mathbb{R}P^3$, les sphères de dimension 2 essentielles sont des obstructions à l'existence de structure géométriques sur une variété de dimension 3. La même chose est vraie pour les tores essentiels, sauf si M est Seifert fibrée ou une Sol variété de dimension 3.

Donc les décompositions en sphères et tores divisent M topologiquement en composantes où ces obstructions connues sont éliminées.

Conjecture d'hyperbolisation (Thurston). — Soit M une variété de dimension 3, fermée et orientée. Alors chaque composante de la décomposition en sphères et tores admet une structure géométrique.

La conjecture de géométrisation donne une classification complète et effective de toutes les variétés fermées, ressemblant de très près sur plusieurs aspects à la classification des surfaces. Plus précisément, elle réduit la classification à celle des variétés géométriques de dimension 3. La classification des variétés géométriques de dimension 3 est assez simple et complètement comprise, sauf dans le cas des variétés hyperboliques, qui reste un champ de recherche actif.

Pour illustrer la puissance de la conjecture de Thurston, voyons maintenant pourquoi elle implique la conjecture de Poincaré. Soit M une variété de dimension 3, simplement connexe; alors la décomposition (1) implique que M doit être de type L . La conjecture de géométrisation dit alors que L est géométrique, et donc $L = \mathbb{S}^3/\Gamma$. Et donc $M = L = \mathbb{S}^3$.

La formulation de la conjecture de géométrisation et le travail de Thurston sur celle-ci ont véritablement révolutionné la topologie des variétés de dimension 3. Voir à ce propos [18] et [19] et les références contenues dans ces articles. Il a ainsi mis le doigt sur le fait que dans la classe de toutes les variétés de dimension 3

irréductibles, les variétés hyperboliques sont celles qui prévalent le plus, comme dans le cas des surfaces, et il a développé un vaste champ d'idées nouvelles et de méthodes neuves pour comprendre la géométrie des variétés de dimension 3. Thurston et de nombreux autres auteurs ont prouvé la conjecture de géométrisation dans des cas particuliers importants, le plus important d'entre eux étant le cas des variétés Haken : si M est une variété de dimension 3 Haken et irréductible (i.e. M contient une surface incompressible de genre supérieur ou égal à 1), alors la conjecture de géométrisation est vraie pour M .

Un ingrédient important dans l'approche de Thurston est la déformation et la dégénérescence des structures hyperboliques sur les variétés non compactes (ou la déformation des variétés hyperboliques singulières sur les variétés compactes). Les huit structures géométriques sont rigides au sens où il n'y a pas de géométrie qui les « interpole » continûment. Ainsi, sur un assemblage de telles structures, chaque composante doit dégénérer en passant d'une composante à l'autre ; il n'y a pas de structure unique ou de métrique donnant la géométrie totale de la variété M . Par exemple, sur la figure 2 les composantes H peuvent être des variétés hyperboliques séparées par des tores, des variétés Seifert fibrées S . Même si cet éclatement est bien défini, les géométries ne s'accordent pas bien dans la région du recollement, et métriquement il n'y a pas de région naturelle où faire ce recollement.

Indépendamment et à peu près simultanément, Gromov [6], [7] a aussi étudié la déformation et la dégénérescence de métriques riemanniennes plus générales ayant une courbure bornée plutôt que constante. L'idée est que l'on peut contrôler le comportement d'une métrique, ou d'une famille de métriques en se donnant une borne uniforme sur le tenseur riemannien $Riem$ de la métrique⁴. Cela conduit à l'important théorème de compacité de Gromov, le théorème de structure des variétés presque plates et la théorie du « collapsing » qui fut développée en détails par Cheeger et Fukaya.

Une version de ces résultats est particulièrement utile pour notre propos. Soit (M, g) une variété riemannienne fermée, avec volume normalisé et supposons que

$$(2) \quad |Riem| \leq \Lambda,$$

pour un certain Λ fini. La métrique g fournit alors une décomposition naturelle de M en parties fine et épaisse, $M = M^\nu \cup M_\nu$, où

$$(3) \quad \begin{aligned} M^\nu &= \{x \in M, Vol B_x(1) \geq \nu\}, \\ M_\nu &= \{x \in M, Vol B_x(1) \leq \nu\}; \end{aligned}$$

ici $B_x(1)$ est la boule géodésique de centre x et de rayon 1 et $\nu > 0$ est un réel arbitraire mais que l'on fixe petit. Maintenant considérons la classe de toutes les variétés de dimension n de volume 1 et qui vérifient (2), et considérons la décomposition correspondantes (3). Alors la géométrie et la topologie de M sont contrôlées a priori. Pour tout $\nu > 0$ donné, il existe un nombre fini (nombre ne dépendant que de λ et de ν) de type topologique possible pour M^ν . De plus, l'espace des métriques sur M^ν est compact dans un sens très naturel ; toute suite

⁴ Le tenseur de courbure est un tenseur de type $(3, 1)$ compliqué qui s'exprime à l'aide des dérivées secondes de la métrique ; dans un système de coordonnées locales en un point il est donné par $R_{ijk}^l = -\frac{1}{2}(\partial_i \partial_k g_{jl} + \partial_j \partial_l g_{ik} - \partial_j \partial_k g_{il} - \partial_i \partial_l g_{jk})$.

de métriques sur M^ν possède une sous-suite qui converge au sens $C^{1,\alpha}$, $\alpha < 1$ (modulo les difféomorphismes). Pour ν assez petit, la partie fine M_ν admet une F -structure au sens de Cheeger et Gromov ; en dimension 3 cela signifie simplement que M_ν est une variété graphe à bord torique ou vide. En particulier la topologie de M_ν est fortement restreinte. Une métrique sur M_ν s'effondre très fortement au sens où les cercles dans les composantes Seifert fibrées et les tores recollant ces composantes ont un diamètre très petit, dépendant de ν ; voir la figure 3 pour une illustration schématique. De surcroît, pour tout $\nu > 0$ fixé, la distance entre M^ν et la partie fine arbitraire $M_{\nu'}$ devient arbitrairement grande lorsque $\nu'/\nu \rightarrow 0$.

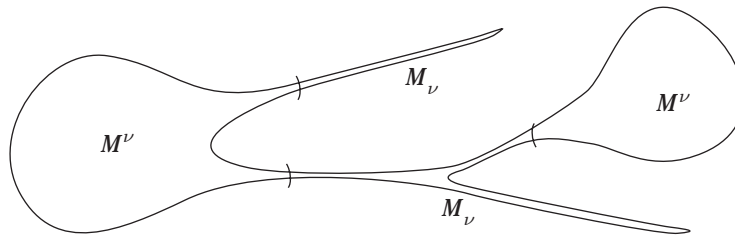


FIG. 3. Décomposition en partie fine et partie épaisse

Nous insistons sur le fait que des résultats similaires sont vrais localement et pour des variétés non compactes ; il en découle que l'hypothèse de volume normalisé faite plus haut n'est pas essentielle.

L'approche de Thurston du problème de géométrisation a fait faire d'énormes progrès sur la partie « hyperbolique » de la conjecture. En comparaison avec cela, très peu de progrès ont été faits sur la partie « à courbure positive » de la conjecture, par exemple la conjecture de Poincaré. Il est important de noter que parmi les huit géométries possibles, \mathbb{H}^3 et \mathbb{S}^3 sont les plus importantes à comprendre (dans le sens de la caractérisation des variétés géométriques). Les autres géométries (mixtes) sont en comparaison plus simples.

Du point de vue de la géométrie riemannienne, la conjecture de Thurston affirme l'existence d'une « meilleure » métrique sur une variété fermée de dimension 3 arbitraire. Dans le cas où M n'est pas elle-même géométrique, on peut autoriser cette métrique optimale à avoir des régions de dégénérescence. La discussion que nous avons eue précédemment ainsi que les figures qui l'illustrent suggèrent que ces dégénérescences doivent se produire via le pincement de sphères de dimension 2 (décomposition en sphère) et l'effondrement de variétés graphes le long de cercles ou de tores (décomposition en tore).

Le flot de Ricci

Une méthode pour trouver une métrique optimale sur une variété consiste à trouver une équation d'évolution naturelle, décrite par un champ de vecteurs dans l'espace des métriques, et essayer de prouver que le flot admet des solutions en tout temps et converge vers une limite (géométrique). Si le flot ne converge pas,

les métriques correspondantes dégènèrent et on doit relier la dégénérescence à la topologie de M .

Il y a essentiellement un seul champ de vecteurs qui soit simple et naturel (ou plus précisément une famille de champs de vecteurs) dans l'espace des métriques. Il est donné par

$$(4) \quad \frac{d}{dt}g(t) = -2Ric_{g(t)} + \lambda(t)g(t).$$

Ici Ric est le tenseur de Ricci, donné en coordonnées locales par $R_{ij} = \sum_k R_{ijk}^k$, de telle façon que Ric est la trace du tenseur riemannien. La constante 2 est seulement placée là pour des raisons de confort et pourrait être changée en faisant un changement homotétique sur le temps; $\lambda(t)$ est une constante dépendant du temps. Le flot de Ricci, introduit par Hamilton [11], est obtenu en prenant $\lambda = 0$, c'est-à-dire

$$(5) \quad \frac{d}{dt}g(t) = -2Ric_{g(t)}.$$

La raison pour laquelle (4) est le seul flot naturel est essentiellement la même que celle qui conduit aux équations du champ d'Einstein en relativité générale. Le tenseur de Ricci est une forme bilinéaire symétrique, tout comme la métrique. À part les multiples de la métrique, c'est la seule forme bilinéaire symétrique qui dépende au plus des dérivées secondes de la métrique et invariante par changement de coordonnées. En faisant un changement homotétique sur la métrique et sur le temps, on peut transformer (5) en (4). Par exemple, en faisant un changement homothétique sur la métrique de telle façon que le volume de $(M, g(t))$ soit constant on récupère (4) avec $\lambda(t) = \frac{2}{Vol(M, g(t))} \int_M R_{g(t)}$ où R est la courbure scalaire.

Dans un système de coordonnées adéquat, l'équation (5) a une forme très naturelle. En effet, au temps t , choisissons des coordonnées harmoniques locales de telle façon que les fonctions coordonnées soient localement des fonctions harmoniques pour $g(t)$. (5) prend alors la forme

$$(6) \quad \frac{d}{dt}g_{ij} = \Delta g_{ij} + Q_{ij}(g, \partial g),$$

où Δ est l'opérateur de Laplace-Beltrami sur les fonctions pour la métrique $g = g(t)$, et Q est un terme quadratique en g et les dérivées d'ordre inférieur de g . C'est une équation de la chaleur non linéaire pour g_{ij} . À partir de l'analyse de cette équation aux dérivées partielles, on obtient l'existence et l'unicité (pour une donnée initiale arbitraire) d'une solution de cette équation sur un intervalle de temps. C'est la raison de la présence du signe $-$ dans (5); un signe $+$ donne une équation de la chaleur « rétroactive » (c'est-à-dire dans les temps antérieurs), qui n'a pas de solution en général.

Donnons quelques exemples de solutions explicites pour le flot de Ricci. Si la métrique initiale est à courbure de Ricci constante, $Ric = a.g$, alors $g(t) = (1 - 2at)g(0)$. Notons que si $a > 0$ le flot contracte la métrique, tandis que si $a < 0$ le flot dilate la métrique, et cela uniformément dans toutes les directions. Ainsi si on fait un changement homothétique pour avoir un volume constant la solution est constante. Les points stationnaires du flot de Ricci avec volume

normalisé sont exactement les métriques d'Einstein, c'est-à-dire les métriques à courbure de Ricci constante. En dimension 3 les métriques d'Einstein sont à courbure constante et donnent donc les modèles \mathbb{H}^3 , \mathbb{S}^3 et \mathbb{R}^3 .

Plus généralement, si $Ric(x, t) > 0$ alors le flot contracte la métrique $g(t)$ au voisinage de x , dans le futur, tandis que si $Ric(x, t) < 0$, alors le flot dilate $g(t)$ au voisinage de x . D'une façon générale, il y aura des directions à courbure de Ricci positives et des directions à courbure de Ricci négatives le long desquelles la métrique se contracte ou se dilate localement.

Supposons que $g(0)$ est une métrique produit sur $\mathbb{S}^1 \times \Sigma$, où Σ est une surface à courbure constante. Alors $g(t)$ reste une métrique produit, la longueur restant constante sur le facteur \mathbb{S}^1 et se contractant ou se dilatant sur la surface selon le signe de la courbure de Σ .

Finalement, le flot de Ricci commute avec l'action du groupe des difféomorphismes et ainsi préserve les isométries de la métrique initiale. Ainsi, les variétés géométriques de dimension 3 restent géométriques. Pour les géométries mixtes « négatives » ($\mathbb{H}^2 \times \mathbb{R}$, $\tilde{S}L(2, \mathbb{R})$, *Nil* et *Sol*), le flot de Ricci normalisé contracte les fibres \mathbb{S}^1 ou \mathbb{T}^2 et dilate le facteur constitué par la surface de base; pour les géométries mixtes « positives » $\mathbb{S}^2 \times \mathbb{R}$, le flot de Ricci normalisé contracte le facteur \mathbb{S}^2 et dilate le facteur \mathbb{R} .

Maintenant considérons le flot de Ricci (5) en général. D'après la forme de l'équation il est clair que le flot $g(t)$ existera si et seulement si la courbure de Ricci reste bornée. Cela suggère le fait que l'on peut considérer le flot induit sur la courbure par le flot sur la métrique. Le plus simple étant le flot induit par le flot sur la métrique sur la courbure scalaire :

$$(7) \quad \frac{d}{dt}R = \Delta R + 2|Ric|^2.$$

En évaluant cette relation en un point qui réalise le minimum R_{min} de R sur M on obtient le fait important que R_{min} est monotone croissante le long du flot. En particulier le flot de Ricci préserve la positivité de la courbure scalaire (en toute dimension). De plus si $R_{min}(0) > 0$, par le même argument on a $\frac{d}{dt}R_{min} \geq \frac{2}{n}R_{min}^2$ (notons qu'on utilise l'inégalité de Cauchy-Schwarz $|Ric|^2 \geq \frac{1}{n}R^2$). Une simple intégration par parties donne donc

$$(8) \quad t \leq \frac{n}{2R_{min}(0)}.$$

Ainsi le flot de Ricci existe pour un temps fini T et $T \leq \frac{n}{2R_{min}(0)}$. Par opposition, dans les régions où la courbure de Ricci est définie négative le flot existe en tout temps.

L'évolution de la courbure de Ricci a la même forme que l'équation (7) :

$$(9) \quad \frac{d}{dt}R_{ij} = \Delta R_{ij} + \tilde{Q}_{ij}.$$

L'expression de \tilde{Q} est beaucoup plus compliquée que le second terme du membre de droite de (7) mais ne fait intervenir que des termes quadratiques en la courbure. Malgré tout \tilde{Q} fait intervenir le tenseur riemannien et pas seulement la courbure de Ricci (tout comme (7) fait intervenir la courbure de Ricci et pas seulement la

courbure scalaire). Un fait important en dimension 3 est que le tenseur riemannien est entièrement déterminé par le tenseur de Ricci. Cela implique qu'en général le flot de Ricci a de meilleures chances de « marcher » en dimension 3. Par exemple, l'analyse du terme \tilde{Q} montre que le flot de Ricci conserve la positivité du tenseur de Ricci en dimension 3 : si $Ric_{g(0)} > 0$ alors $Ric_{g(t)} > 0$ pour tout $t > 0$. Ceci n'est pas le cas en dimension plus grande. D'un autre côté, en dimension supérieure ou égale à 3, le flot de Ricci ne préserve pas la négativité du tenseur de Ricci pas plus qu'il ne préserve une borne inférieure $Ric \geq -\lambda$ pour $\lambda > 0$. Dans tout le reste de cet article nous supposons que la dimension de la variété est 3.

Dans le théorème de compacité de Gromov et la décomposition fine/épaisse (3), l'hypothèse $|Riem|$ bornée peut aussi être remplacée par une borne sur $|Ric|$ (car nous sommes ici en dimension 3). Il en découle que sur les intervalles de temps où $|Ric|$ reste bornée les métriques $g(t)$ sont toutes quasi-isométriques : à savoir qu'il existe c et C dépendant du temps t tels que $cg(0) \leq g(t) \leq Cg(0)$. De ce fait, la région fine M_ν pour $\nu \ll 1$ ne peut apparaître, sous l'hypothèse $|Ric|$ bornée, que pour des temps très grands.

De tout cela il vient que le flot de Ricci est très naturel et a de nombreuses propriétés intéressantes. On peut en effet voir des liens avec la vision de Thurston des variétés de dimension 3. Pourtant, la première indication que le flot de Ricci est un outil important pour traiter des problèmes géométriques est le résultat suivant de Hamilton :

Space form theorem. — *Si $g(0)$ est une métrique à courbure de Ricci strictement positive sur une variété de dimension 3, alors le flot normalisé a une solution en tout temps et converge vers la métrique standard de \mathbb{S}^3/Γ , où Γ est un sous-groupe fini de $SO(4)$ agissant librement sur \mathbb{S}^3 .*

Ainsi le flot de Ricci « géométrise » les variétés de dimension 3 à courbure de Ricci strictement positive. Depuis ce résultat important, la question de savoir si ce résultat peut être étendu aux variétés à courbure scalaire strictement positive reste complètement ouverte.

Même si l'évolution de la courbure le long du flot est très compliquée pour des métriques initiales quelconques, une analyse détaillée de (9) donne le résultat suivant :

Estimation sur le pincement de la courbure : [10], [12]. — *Soit $g(t)$ une solution du flot de Ricci sur une variété fermée M de dimension 3. Alors il existe une fonction $\varphi :]-\infty, +\infty[\rightarrow \mathbb{R}$, qui tend vers 0 en l'infini, et une constante C ne dépendant que de la métrique initiale $g(0)$ telles que*

$$(10) \quad Riem(x, t) \geq -C - \varphi(R(x, t)) \cdot |R(x, t)|.$$

Cet énoncé signifie que toutes les courbures sectionnelles R_{ijji} de $g(t)$, où e_i est une base orthonormée en (x, t) , sont minorées par le membre de droite de (10).

Cette estimation n'implique pas une borne inférieure sur $Riem(x, t)$ uniforme en temps. Pourtant lorsqu'on la combine avec le fait que la courbure scalaire est minorée uniformément en temps, elle implique que lorsque $R(x, t) \gg 1$ alors $|Riem(x, t)| \gg 1$. Par conséquent pour contrôler la taille de $|Riem|$, il est suffisant de contrôler la courbure scalaire par le haut. Ceci est tout à fait remarquable puisque

la courbure scalaire est un invariant de la métrique beaucoup moins rigide que le tenseur riemannien. De surcroît, aux points où la courbure est assez grande, (10) montre que $\frac{Riem(x,t)}{R(x,t)} \geq -\delta$ avec δ petit. Il est donc clair que si on normalise la métrique de façon que $R(x,t) = 1$ alors $Riem(x,t) \geq -\delta$. Pour une telle normalisation la métrique a une courbure presque positive au voisinage de (x,t) .

Inégalité de Harnack : [9]. — *Soit $(N, g(t))$ une solution du flot de Ricci telle que $Riem \geq 0$ et $|Riem|$ bornée, et supposons que $g(t)$ est une métrique complète sur N . Alors pour tout $0 < t_1 \leq t_2$,*

$$(11) \quad R(x_2, t_2) \geq \frac{t_1}{t_2} \exp\left(-\frac{d_{t_1}(x_1, x_2)}{2(t_2 - t_1)}\right) R(x_1, t_1),$$

où d_{t_1} est la distance sur (M, g_{t_1}) .

Cette estimation nous permet de relier ou de contrôler la géométrie des solutions en différents points de l'espace-temps.

L'obtention d'une estimée analogue à celle de (11) en général (*i.e.* sans l'hypothèse $Riem \geq 0$) a été l'un des obstacles majeurs pour des progrès plus importants sur le flot de Ricci.

L'analyse faite plus haut montre que le flot de Ricci semble favoriser la courbure strictement positive. Le flot tend à évoluer de manière à rendre la courbure plus positive, et les résultats les plus forts ont été prouvés dans le cas de la courbure strictement positive, et cela contraste d'une certaine façon avec l'approche de Thurston.

La formation des singularités

L'analyse la plus profonde du flot de Ricci concerne les singularités pouvant apparaître le long du flot en temps fini. Comme (8) le montre déjà, le flot de Ricci n'existera pas en général en tout temps. Dans le cas d'une métrique initiale à courbure de Ricci strictement positive, ce problème est écarté grâce à une renormalisation du volume. Le Space Form Theorem d'Hamilton montre que le flot normalisé existe en tout temps et converge C^∞ vers la métrique ronde. Pourtant la situation est nécessairement plus compliquée en dehors de la classe des variétés à courbure de Ricci strictement positive. Considérons par exemple des métriques initiales à courbure scalaire strictement positive. Toute variété qui est une somme connexe de facteurs pouvant être \mathbb{S}^3/Γ ou $\mathbb{S}^2 \times \mathbb{S}^1$ possède une métrique à courbure scalaire strictement positive (comparer cela avec la décomposition en sphère (1)). Ainsi pour des raisons topologiques évidentes, le flot de Ricci normalisé ne peut pas converger vers une métrique ronde ; même le flot normalisé peut développer des singularités.

Des singularités apparaissent fréquemment dans de nombreux types d'EDP non linéaires et celles-ci ont été intensivement étudiées depuis plusieurs décennies. Tout spécialement dans un contexte géométrique, le problème classique pour comprendre la structure des singularités est de renormaliser ou de faire un changement d'échelle sur une suite qui converge vers la singularité de manière à rendre la solution bornée et essayer de passer à la limite sur la solution renormalisée. Une telle solution limite sert de modèle pour cette singularité, et l'on peut espérer (ou attendre) que

la singularité possèdera certaines caractéristiques qui la rendra plus simple qu'une solution arbitraire de l'équation.

Un singularité ne peut se former le long du flot de Ricci que là où la courbure devient non bornée. Supposons $\lambda_i^2 = |\text{Riem}|(x_i, t_i) \rightarrow \infty$ pour une suite de points $x_i \in M$ et une suite de temps $t_i \rightarrow T < \infty$. Il est alors naturel de considérer les métriques renormalisées et le temps renormalisé

$$(12) \quad \tilde{g}_i(\tilde{t}_i) = \lambda_i^2 g(t), \quad \tilde{t}_i = \lambda_i^2(t - t_i).$$

Les métriques \tilde{g}_i sont aussi solutions du flot de Ricci et sont à courbure bornée en $(x_i, 0)$. Pour un choix adéquat de x_i et t_i , la courbure sera bornée au voisinage de x_i , et pour des temps voisins négatifs $\tilde{t}_i \leq 0$; par exemple, on pourrait prendre les points où la courbure est maximale sur $(M, g(t))$, $0 \leq t \leq t_i$.

La renormalisation (12) dilate les distances d'un facteur λ_i et le temps d'un facteur λ_i^2 . Donc on en vient à étudier de très petites régions, de tailles spatiales d'ordre $r_i = \lambda_i^{-1}$, et en utilisant un « microscope » pour examiner les phénomènes à petite échelle dans cette région sur une échelle de taille environ 1. Bien sûr, un changement de coordonnées est implicitement fait dans cette analyse, i.e. l'utilisation de difféomorphismes locaux en accompagnement de la renormalisation de la métrique.

Une version locale du théorème de compacité de Gromov autorise donc à passer à la limite sur les solutions du flot de Ricci, au moins localement en temps et espace, à condition que les volumes locaux de la renormalisation soient minorés par une quantité strictement positive; plus précisément on a besoin que $x_i \in M^\nu(\tilde{g}_i(\tilde{t}_i))$, pour un certain $\nu > 0$; voir (3). En terme de flot original (non normalisé), cela signifie que la métrique $g(t)$ ne doit pas s'effondrer localement, à l'échelle de sa courbure, i.e.

$$\text{Vol}B_{x_i}(r_i, t_i) \geq \nu r_i^3.$$

Une limite maximal connexe $(N, \bar{g}(\bar{t}), x)$ contenant le point base $x = \lim x_i$ est alors appelée un modèle singulier. Observons que la topologie de la limite N peut très bien être différente de celle de la variété originale M , beaucoup de celle-ci pouvant avoir été envoyée à l'infini par la renormalisation.

Pour décrire l'utilité potentielle de ce procédé, supposons que nous avons un non effondrement local sur les métriques normalisées et que nous avons choisi des points en temps et en espace de courbure maximale. On a alors, au moins pour une sous-suite, une solution limite du flot de Ricci $(N, \bar{g}(\bar{t}), x)$, avec point base x , définie au moins sur $]-\infty, 0]$; de plus $\bar{g}(\bar{t})$ est une métrique riemannienne complète sur N . Elles sont appelées « ancient solutions » du flot de Ricci dans la terminologie d'Hamilton. Les estimations dans (10) et (11) peuvent maintenant être utilisées pour montrer que les modèles singuliers ont en fait d'importantes caractéristiques qui les rendent plus simples que les solutions générales du flot de Ricci. Comme nous le disons à la suite de (10), l'estimation de pincement implique que la limite est à courbure positive ou nulle. De plus, la topologie des variétés complètes N à courbure positive ou nulle est complètement comprise en dimension 3. Si N est non compacte alors N est difféomorphe à \mathbb{R}^3 , $\mathbb{S}^3 \times \mathbb{R}$ ou un quotient de ces variétés. Si N est compacte, alors une forme légèrement plus forte du théorème d'Hamilton

ci-dessus implique que N est difféomorphe à \mathbb{S}^3/Γ , $\mathbb{S}^2 \times \mathbb{S}^1$ ou $\mathbb{S}^2 \times_{\mathbb{Z}_2} \mathbb{S}^1$. De plus l'inégalité de Harnack (11) est vraie à la limite.

Ces propriétés générales des modèles singuliers sont certainement encourageants. Néanmoins il reste de nombreux problèmes à résoudre pour tirer un vrai bénéfice de ceci.

- I On doit prouver le non-effondrement à l'échelle de la courbure pour obtenir un modèle singulier.
- II En général, la courbure peut exploser à différentes échelles ou vitesses, et il n'est pas entièrement suffisant de comprendre seulement les modèles singuliers aux points de courbure maximale. Ce type de phénomène (appelé généralement « bubbling ») se produit dans beaucoup d'autres problèmes variationnels issus de la géométrie, par exemple les applications harmoniques, les champs de Yang-Mills, les métriques d'Einstein, et d'autres. Dans de tels problèmes elliptiques ces problèmes d'échelles multiples ont effectivement été résolus.
- III Même en sachant résoudre les deux premiers problèmes listés ci-dessus, on doit tout de même relier la structure des singularités avec la topologie de la variété sous-jacente.

L'étude de la formation des singularités le long du flot de Ricci a été initiée par Hamilton [10]; voir aussi [4] pour un survey récent. Même s'il y a eu de nombreux progrès techniques en la matière pendant la dernière décennie, les problèmes essentiels sur l'existence et la structure des modèles singuliers et de leurs relations avec la topologie restaient ouverts jusqu'à l'apparition des travaux de Perelman en 2002 et 2003.

Le travail de Perelman

Le travail récent de Perelman [15]-[17] (ainsi qu'un article moins crucial et qui reste à paraître) implique une solution complète de la conjecture de géométrisation. Ceci se fait grâce à l'introduction de nombreuses idées hautement originales en géométrie et de nombreuses techniques nouvelles pour comprendre le flot de Ricci. En particulier, le travail de Perelman résout complètement les questions I-III ci-dessus. Commentons maintenant, nécessairement brièvement, certaines des avancées dues à Perelman.

I Non effondrement

Considérons la fonctionnelle d'Einstein-Hilbert

$$(13) \quad \mathcal{R}(g) = \int_M R(g) dv(g)$$

définie sur l'espace des métriques riemanniennes \mathbb{M} sur une variété M . Les points critiques de \mathcal{R} sont les métriques Ricci-plates ($Ric = 0$). Cette fonctionnelle peut être modifiée, par exemple en ajoutant une constante cosmologique -2Λ , pour donner une fonctionnelle dont les points critiques seraient des métriques d'Einstein à courbure de Ricci constante⁵. Il est naturel d'essayer de relier le flot de Ricci

⁵ L'action (13) mène à l'équation de champ d'Einstein dans le vide en relativité générale pour les métriques de Lorentz sur une variété de dimension 4. Le terme $\lambda(t)$ dans (4) est l'analogue de la constante cosmologique.

à \mathcal{R} ; par exemple la flot de Ricci est-il le flot du gradient de \mathcal{R} (par rapport à la métrique naturelle L^2 sur \mathbb{M})? Même s'il semblait que ce devait être presque vrai, il fut reconnu que tel n'est pas le cas. En fait le flot du gradient de \mathcal{R} n'existe même pas puisqu'il implique une équation rétroactive de la chaleur pour la courbure scalaire R (similaire à (7) mais avec un signe moins devant le laplacien).

Considérons maintenant la fonctionnelle suivante

$$(14) \quad \mathcal{F}(g, f) = \int_M (R + |\nabla f|^2) e^{-f} dv(g).$$

C'est une fonctionnelle sur l'espace plus grand $\mathbb{M} \times C^\infty(M, \mathbb{R})$, ou de façon équivalente une famille de fonctionnelles sur \mathbb{M} paramétrée par $C^\infty(M, \mathbb{R})^6$. Prenons une mesure lisse dm sur M et considérons le couplage de Perelman demandant que (g, f) satisfasse

$$(15) \quad e^{-f} dg(g) = dm.$$

La fonctionnelle ainsi obtenue

$$(16) \quad \mathcal{F}^m(g, f) = \int_M (R + |\nabla f|^2) dm$$

devient une fonctionnelle sur \mathbb{M} . Au premier abord il pourrait sembler que ceci est beaucoup plus compliqué que (13); pourtant pour tout $g \in \mathbb{M}$, il existe une large classe de fonctions f (ou de mesures dm) telles que le flot du gradient L^2 de \mathcal{F}^m existe en g , et il est donné par

$$(17) \quad \frac{d\tilde{g}}{dt} = -2(\text{Ric}_{\tilde{g}} + D^2f),$$

où D^2f est la hessienne de f pour la métrique \tilde{g} . L'équation d'évolution (17) pour \tilde{g} est simplement le flot de Ricci (5) modifié par une difféomorphisme infinitésimal : $D^2f = \frac{d}{dt}(\varphi_t^* \tilde{g})$, où $\frac{d}{dt}\varphi_t = \nabla f$. Il s'ensuit que le flot du gradient de \mathcal{F}^m est le flot de Ricci aux difféomorphismes près. Différents choix de dm correspondent à des choix différents de difféomorphismes. En particulier la fonctionnelle \mathcal{F}^m croît le long du flot de Ricci.

Que peut-on faire avec cette fonctionnelle plus compliquée? Il s'avère qu'étant donnée une métrique initiale $g(0)$ et $t > 0$, la fonction f (et ainsi la mesure dm) peut être librement déterminée en $g(t)$ ($g(t)$ étant l'évolution de $g(0)$ le long du flot de Ricci (5)). Perelman utilise alors cette « liberté » pour vérifier la géométrie de $g(t)$ avec des choix adéquats de f .

Par exemple, on peut montrer par une analyse très simple de \mathcal{F}^m que l'effondrement ou le non effondrement de la métrique $g(t)$ au voisinage d'un point $x \in M$ peut être détecté à partir de la taille de $\mathcal{F}^m(g(t))$ en choisissant e^{-f} de telle façon qu'elle soit une approximation d'une masse de Dirac en x . Plus l'effondrement de $g(t)$ en x est grand plus la valeur de $\mathcal{F}^m(g(t))$, sera négative. L'effondrement est donc écarté à toutes les échelles en temps fini en associant ceci avec le fait que la fonctionnelle \mathcal{F}^m croît avec le temps le long du flot de Ricci. En fait, cet argument

⁶ La fonctionnelle (14) apparaît en théorie des cordes comme énergie de moindre action [5]; la fonction ou le champ scalaire f est appelée la dilatation. Il est intéressant de noter dans ce contexte que le champ gravitationnel et le champ de dilatation proviennent simultanément de la quantification du « string world sheet » (σ -model) [5]

est mis en œuvre en utilisant une fonctionnelle plus compliquée que \mathcal{F}^m , et invariante par changement d'échelle; motivé par certaines analogies avec la physique statistique Perelman l'appelle la fonctionnelle entropie.

Modèles singuliers. — Une seconde avancée de [15] est essentiellement une classification de tous les modèles singuliers complets $(N, g(t))$ pouvant apparaître en temps fini. Ici complet signifie que la métrique $g(0)$ est une métrique complète sur N . Si N est compacte et lisse, il s'ensuit grâce au space-form theorem d'Hamilton que N est difféomorphe à \mathbb{S}^3/Γ , $\mathbb{S}^2 \times \mathbb{S}^2$ ou $\mathbb{S}^2 \times_{\mathbb{Z}_2} \mathbb{S}^2$. Dans le cas plus difficile où N est complète non compacte, Perelman prouve que la géométrie de N à l'infini est la plus simple et la plus naturelle possible. Au temps $t = 0$ et en un point x avec $r(x) = \text{dist}(x, x_0) \gg 1$, pour un point base fixé x_0 , un voisinage « de grande taille » de x à une échelle où $R(x) = 1$ est ε -proche d'un voisinage « de grande taille » sur $\mathbb{S}^2 \times \mathbb{R}$ muni de sa métrique standard. Ici ε peut être rendu arbitrairement petit en choisissant $r(x)$ grand. Une telle région est appelée ε -neck. Ainsi la géométrie à l'infini de N est celle d'une réunion de ε -necks, où le rayon de \mathbb{S}^2 variant lentement peut être soit uniformément borné soit diverger vers l'infini, mais seulement à une vitesse inférieure à celle de $r(x)$. De surcroît, cette structure est aussi valable sur un intervalle long pour des temps négatifs, de telle façon que sur ces régions la solution est proche de celle du flot de Ricci rétroactive sur $\mathbb{S}^2 \times \mathbb{R}$. Topologiquement N est difféomorphe à \mathbb{R}^3 ou (N, g) est isométrique à $\mathbb{S}^2 \times \mathbb{R}$.

Perelman montre que ce résultat de structure pour les modèles singuliers est aussi vrai pour la solution $g(t)$ au voisinage d'un temps singulier T . Ainsi, pour un point de base (x, t) où la courbure est assez grande, le changement d'échelle fait en (12) sur l'espace-temps par la courbure est suffisamment proche, sur des domaines compacts de grande taille, de domaines de grande taille sur un modèle singulier complet. Le modèle singulier complet « idéal » décrit en réalité la géométrie et la topologie près d'une singularité. Par conséquent, on a une compréhension détaillée de la géométrie à petite échelle et de la topologie partout sur $(M, g(t))$ pour t proche de T . En particulier cela montre une version générale de l'inégalité de Harnack (11).

Ces résultats sont bien sûr assez techniques et les preuves ne sont pas simples. Pourtant elles ne sont pas exceptionnellement difficiles et elles reposent sur des idées novatrices et des outils nouveaux pour comprendre le flot de Ricci. Une idée clé est l'utilisation du résultat de non effondrement ci-dessus à toutes les échelles.

III Lien avec la topologie

Le point important maintenant est l'apparition de sphères de dimension 2 au voisinage des singularités. Rappelons, d'après (1), que l'on doit d'abord faire une décomposition en sphère de M avant de pouvoir la géométrer. Il n'y a pas de géométrie correspondant à la décomposition en sphère⁷. Alors que la décomposition en sphère est l'opération la plus simple topologiquement, géométriquement et analytiquement elle est de loin la plus difficile à comprendre. Comment peut-on repérer des sphères de dimension 2 dans M sur lesquelles faire de la chirurgie à partir de la géométrie donnée par une métrique? Nous allons voir maintenant que de telles

⁷ On pourrait penser que la géométrie de $\mathbb{S}^2 \times \mathbb{R}$ correspond à la décomposition en sphère mais ceci n'est pas tout à fait exact; au mieux ceci pourrait avoir un sens dans un contexte limité.

sphères, plongées dans les ε -necks dont on a parlé ci-dessus, apparaissent naturellement au voisinage des singularités du flot de Ricci.

L'idée est alors de faire de la chirurgie sur M le long de sphères de dimension 2 juste avant le temps T d'apparition de la première singularité. La figure 4 donne une description schématique d'une métrique partiellement singulière $g(T)$ sur M . La métrique $g(T)$ est lisse sur un domaine maximal Ω , où la courbure est localement bornée mais singulière, *i.e.* pas complètement définie, sur le complémentaire de l'ensemble où la courbure explose lorsque $t \rightarrow T$.

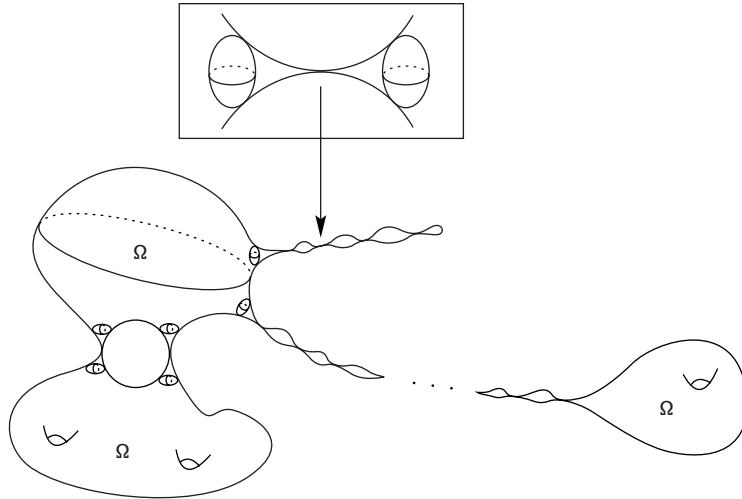


FIG. 4. Cornes sur la limite singulière

Supposons d'abord que Ω est vide, de telle manière que la courbure de $g(t)$ explose partout sur M quand $t \rightarrow T$. On dit que la solution du flot de Ricci s'éteint au temps T . Notons que $R(x, t) \gg 1$ pour tous $x \in M$ et t au voisinage de T (d'après l'estimation de pincement (10)). Etant entendu que nous avons une compréhension complète des modèles singuliers, il n'est pas difficile de voir que M est alors difféomorphe à \mathbb{S}^3/Γ , $\mathbb{S}^2 \times \mathbb{S}^1$ ou $\mathbb{S}^2 \times_{\mathbb{Z}_2} \mathbb{S}^1$. Dans cette situation on a fini et M est géométrique.

Si Ω n'est pas vide le point important est que les petits voisinages de $\partial\Omega$ sont composés de cornes. Une corne est une métrique sur $\mathbb{S}^2 \times [0, \delta]$ où le facteur \mathbb{S}^2 est presque rond de rayon $\rho(r)$ avec $\rho(r)$ petit et $\frac{\rho(r)}{r} \rightarrow 0$ lorsque r tend vers 0. Alors une corne est une union de ε -necks assemblés à des échelles de plus en plus petites. La figure 4 représente une métrique partiellement singulière sur une variété lisse $\mathbb{S}^2 \times I$, constituée de paires de cornes jointes par une métrique dégénérée. Au temps T il peut y avoir une infinité de composantes de Γ , de taille arbitrairement petite, contenant de telles cornes. Pourtant seul un nombre fini de telles composantes sont des cornes doubles, qui sont toutes topologiquement de la forme $\mathbb{S}^2 \times I$. En termes quantitatifs, il existe une constante petite $\rho_0 > 0$ telle que si Ω ne contient pas de cornes contenant une sphère $\mathbb{S}^2 \times \{\delta\}$ de rayon supérieur ou égal à ρ_0 , alors, comme ci-dessus dans le cas Ω vide, M est difféomorphe à \mathbb{S}^3/Γ , $\mathbb{S}^2 \times \mathbb{S}^1$ ou $\mathbb{S}^2 \times_{\mathbb{Z}_2} \mathbb{S}^1$ et on

a fini. S'il existe des cornes contenant une sphère $\mathbb{S}^2 \times \delta$ d'une taille fixée ρ_0 dans Ω , alors on peut faire de la chirurgie sur chacune de ces cornes en la tronquant le long de \mathbb{S}^2 de rayon ρ_0 et en recollant une boule de dimension 3, donnant ainsi une famille disjointe de variétés de dimension 3.

Maintenant que l'on a déconnecté M par chirurgie sur des sphères de dimension 2 en un nombre fini de composantes, on continue ainsi avec le flot de Ricci sur chaque composante. Un argument conceptuellement simple mais techniquement difficile basé sur la décroissance du volume associé à chaque chirurgie, montre que le nombre de ces chirurgies est localement fini : sur tout segment de temps il n'y a qu'un nombre fini de temps où des singularités peuvent se former.

Comme exemple concret supposons que la métrique de départ soit à courbure scalaire strictement positive. Alors l'estimée (8) montre que le flot de Ricci est complètement terminé, i.e. il s'éteint en temps fini. Ainsi on ne fait qu'un nombre fini de chirurgies sur M le long du flot de Ricci et il s'ensuit que M est difféomorphe à une somme connexe d'un nombre fini de facteurs de la forme \mathbb{S}^3/Γ ou $\mathbb{S}^2 \times \mathbb{S}^1$.

Le point important de cette procédure est que si successivement on ignore ces composantes qui s'éteignent en temps fini (et qui ont déjà été identifiées topologiquement), le flot de Ricci avec chirurgie existera sur $[0, \infty[$. À quoi ressemble alors la géométrie des composantes $\{\hat{M}_i\}$ pour un temps très grand T_0 ? Ici la décomposition fine et épaisse de Gromov-Thurston apparaît. Fixons $\hat{M} \in \{\hat{M}_i\}$ et regardons la métrique normalisée $\hat{g}(t) = t^{-1}g(t)$, pour $t = T_0$. Il est facile de voir en utilisant l'équation du flot de Ricci que le volume de $(\hat{M}, g(t))$ est uniformément borné. Pour ν assez petit Perelman montre qu'il y a un contrôle suffisant sur la partie ν -épaisse \hat{M}^ν pour montrer que \hat{M}^ν est difféomorphe à une variété hyperbolique complète H de dimension 3 (ayant un nombre fini de composantes connexes). Le flot de Ricci existe sur \hat{M}^ν en tout temps, et les métriques normalisées $t^{-1}g(t)$ convergent vers la métrique hyperbolique de courbure $-\frac{1}{4}$ lorsque $t \rightarrow \infty$. Alors qu'il y a moins de contrôle sur la partie ν -fine, il y en a assez pour conclure que \hat{M}_ν est difféomorphe à une variété graphe G (ayant un nombre fini de composantes). Même s'il y peut y avoir une infinité de chirurgies cette fois pour pouvoir continuer le flot, elles ont toutes lieu dans G . Ainsi la variété originale M a été décomposée topologiquement (pour des temps très grands)

$$M = (K_1 \# \dots \# K_p) \# (\#_1^q \mathbb{S}^3/\Gamma_i) \# (\#_1^r \mathbb{S}^2 \times \mathbb{S}^1).$$

Perelman a montré récemment [17] que les facteurs \mathbb{S}^3/Γ et $\mathbb{S}^2 \times \mathbb{S}^1$ s'éteignent nécessairement en temps fini (avec une borne sur ce temps d'extinction qui dépend de la métrique initiale), et donc seuls les facteurs K subsistent après un temps assez long (ce résultat n'est de toute façon pas nécessaire pour la conjecture de géométrisation).

De plus chaque $K = K_i$ se décompose en partie fine et épaisse comme union

$$K = H \cup G,$$

où H est une variété hyperbolique complète de volume fini (qui peut être non connexe) et G est une variété graphe (qui peut être non connexe). L'union de H et G se fait le long d'une collection de tores plongés. Perelman utilise les preuves de [11] ou [1] et [2] pour conclure que chacun de ces tores est incompressible dans K .

Ce procédé donne donc à la fois la décomposition en sphère et en tore de la variété M . Même s'il n'est pas avéré que le flot de Ricci détecte les décompositions de G en composantes Seifert fibrées, ceci devrait être comparativement plutôt élémentaire d'un point de vue topologique. Les composantes toriquement irréductibles de K sont en fait des variétés hyperboliques.

Ceci termine notre survol de la conjecture de géométrisation. Le travail de Perelman a provoqué beaucoup d'émulation au sein de la communauté mathématique, ainsi que parmi les gens qui s'intéressent à la science en général. Alors même que des vérifications détaillées du travail de Perelman sont en cours, la beauté et la profondeur de ses contributions sont évidentes.

Je suis très reconnaissant à Bruce Kleiner, John Lott et Jack Milnor pour leurs nombreuses suggestions et commentaires, qui ont contribué à améliorer de manière significative le contenu de cet article.

Références

- [1] M. Anderson, Scalar curvature and geometrization conjectures for 3-manifolds, in *Comparison Geometry*, (Berkeley 1993-94), MSRI Publications, **30**, (1997), 49-82.
- [2] M. Anderson, Scalar curvature and the existence of geometric structures on 3-manifolds, I, *Jour. Reine Angew. Math.* **553**, (2002), 125-182.
- [3] L. Andersson, The global existence problem in general relativity, (preprint, 1999), gr-qc/9911032, to appear in "50 Years of the Cauchy Problem in General Relativity", eds. P.T. Chruściel and H. Friedrich.
- [4] H.-D. Cao and B. Chow, Recent developments on the Ricci flow, *Bull. Amer. Math. Soc.* **36**, (1999), 59-74.
- [5] E. D'Hoker, String theory, in *Quantum Fields and Strings : A Course for Mathematicians*, 2, Amer. Math. Soc. (1999).
- [6] M. Gromov, *Structures Métriques pour les Variétés Riemanniennes*, Cedic/Fernand Nathan, Paris, (1981).
- [7] M. Gromov, Volume and bounded cohomology, *Publ. Math. I.H.E.S.* **50**, (1982), 5-100.
- [8] R. Hamilton, Three manifolds of positive Ricci curvature, *Jour. Diff. Geom.* **17**, (1982), 255-306.
- [9] R. Hamilton, The Harnack estimate for the Ricci flow, *Jour. Diff. Geom.* **37**, (1993), 225-243.
- [10] R. Hamilton, Formation of singularities in the Ricci flow, in *Surveys in Differential Geometry*, Vol. 2, International Press, (1995), 7-136.
- [11] R. Hamilton, Non-singular solutions of the Ricci flow on three-manifolds, *Comm. Anal. Geom.* **7**, (1999), 695-729.
- [12] T. Ivey, Ricci solitons on compact three-manifolds, *Diff. Geom. and Appl.* **3**, (1993), 301-307.
- [13] B. Kleiner and J. Lott, Ricci flow website : <http://www.math.lsa.umich.edu/research/ricciflow/perelman.html>
- [14] J. Milnor, Towards the Poincaré Conjecture and the classification of 3-manifolds, (preprint, 2003), to appear in *Notices Amer. Math. Soc.*
- [15] G. Perelman, The entropy formula for the Ricci flow and its geometric applications, (preprint, 2002), math.DG/0211159.
- [16] G. Perelman, Ricci flow with surgery on three-manifolds, (preprint, 2003), math.DG/0303109.
- [17] G. Perelman, Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, (preprint, 2003), math.DG/0307245.
- [18] W. Thurston, *The Geometry and Topology of Three-Manifolds*, (preprint, 1978), Princeton University; available online at <http://www.msri.org/publications/books/gt3m>, and *Three-Dimensional Geometry and Topology*, Vol. 1, Princeton University Press, (1997).
- [19] W. Thurston, Three dimensional manifolds, Kleinian groups and hyperbolic geometry, *Bull. Amer. Math. Soc.* **6**, (1982), 357-381.