David Ruelle

La théorie ergodique des systèmes dynamiques d'Anosov

Je voudrais présenter un transfert de technologie mathématique, entre la théorie ergodique basée sur la mécanique statistique de l'équilibre (que je vais décrire dans la première partie) et la théorie des systèmes dynamiques différentiables hyperboliques (que je décrirai dans la seconde partie).

On a commencé d'abord, dans les années 1970, à étudier rigoureusement une partie de la physique : la mécanique statistique de l'équilibre. En fait, il faut remarquer que, dans la conceptualisation de cette partie de la réalité, il y a des structures mathématiques sous-jacentes très intéressantes. De plus, on a développé une théorie ergodique qui est utile dans d'autres domaines. Et il se trouve que, je ne sais pas bien pourquoi, des personnes ont été impliquées à la fois en mécanique statistique de l'équilibre et en théorie des systèmes dynamiques hyperboliques. Il y a eu ainsi un transfert d'information entre les deux domaines.

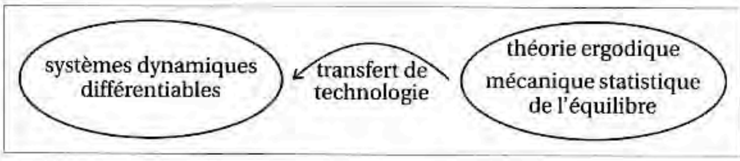

Fig. 1.

Peut-être Yakov Grigorievich Sinai est-il la personne que l'on peut citer en premier lieu. En se trouvant dans cet environnement russe où la physique théorique est moins éloignée des mathématiques que dans l'autre partie du globe, ce transfert de technologie était pour Sinai plus naturel. J'ai eu, par ailleurs, la chance de rencontrer Rufus Bowen [1], un mathématicien qui est malheureusement décédé très jeune, à l'âge de 31 ans. Bowen, grâce à son génie de la simplicité, a beaucoup contribué au développement de ce sujet. On reparlera de lui plus tard.

Cela dit, je vais commencer par décrire cette théorie ergodique, c'est-à-dire la mécanique statistique de l'équilibre. Je pense qu'il y a un certain nombre de mathématiciens qui, quand on parle de mécanique, et surtout de mécanique statistique, sont un peu mal à l'aise. Mais, bien que j'utiliserai des termes d'origine physique, le fait est qu'il s'agit d'une théorie mathématique. Donc, vous ferez plus ou moins d'efforts, comme il vous plaira, pour oublier (ou pas) la physique.

Systèmes de spins sur un réseau

Premières définitions

J'utilise au départ le langage physique en disant que j'ai un réseau dont les points s'appellent x. Toutefois, cela revient aussi bien à dire que l'on considère les points d'un réseau $\mathbb{Z}^\nu$, où $\mathbb{Z}$ est l'ensemble des nombres entiers et ν est un entier positif. Par la suite, Λ représentera un sous-ensemble fini de ce réseau. Par exemple, dans le cas où $\nu = 1$, il s'agit simplement de l'ensemble des entiers et Λ est un ensemble fini d'entiers.

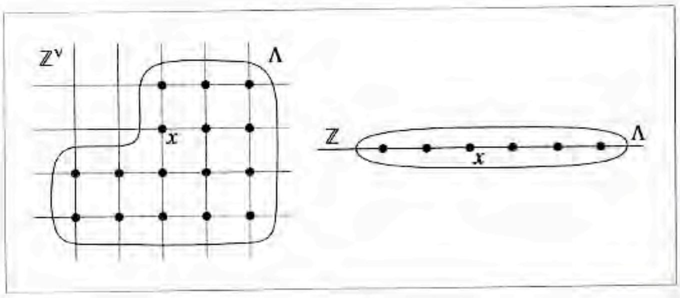

Fig. 2.

1. Robert Edward Bowen (1947-1978) était connu comme Rufus Bowen à cause de ses cheveux et sa barbe rouges. (N.d.R.)

L'idée est que l'on va prendre un spin σ_x en chaque point x du réseau ou en chaque point x de Λ. Qu'est-ce que c'est qu'un spin ? Inutile d'en donner le sens physique ! Pour nous, ce sera simplement un élément d'un ensemble fini. On pourrait lui donner n'importe quel nom, il ne faut pas être impressionné par le mot *spin*. Cela ne veut strictement rien dire dans ce cadre-ci. C'est simplement un mot ! Pour les sceptiques, j'ajoute que l'on pourrait faire de la géométrie en remplaçant droite et plan par table et chaise et que la systématisation euclidienne resterait. Ici j'ai un ensemble fini F dont les éléments s'appellent les valeurs de spin. Et une configuration de spins, $\underline{\sigma} = \{\sigma_x\}$, est simplement obtenue en donnant pour chaque point x une valeur de spin $\sigma_x \in \mathrm{F}$, où x peut varier soit sur tout le réseau $\mathbb{Z}^\nu$, soit sur un sous-ensemble fini Λ.

Un concept qui est quand même plus intéressant, et qui vient aussi de la physique, est celui de l'*énergie*. Intuitivement, l'énergie n'est qu'une grandeur dont il est bon d'imaginer qu'elle devient grande lorsque l'ensemble Λ considéré est grand. Comment peut-on alors définir une énergie ? Eh bien, on va prendre une somme sur les sous-ensembles X quelconques de l'ensemble Λ, d'une fonction $\Phi = \{\Phi(\cdot|\mathrm{X})\}_\mathrm{X}$ (appelée « interaction »), définie sur les configurations de spins restreintes à X, c'est-à-dire

$$\mathrm{H}_\Lambda(\underline{\sigma}) = \sum_{\mathrm{X}\subset\Lambda} \Phi(\underline{\sigma}|\mathrm{X}).$$

Il faut d'abord observer que toute fonction sur F^Λ peut s'écrire de cette manière-là. En effet, étant donnée une fonction $\mathrm{U} : \bigcup_{\mathrm{X}\subset\Lambda} \mathrm{F}^\mathrm{X} \to \mathbb{R}$ telle que $\mathrm{U} = 0$ sur $\mathrm{F}^\varnothing$, il existe une fonction unique

$$\Phi_\mathrm{U} : \bigcup_{\mathrm{X}\subset\Lambda} \mathrm{F}^\mathrm{X} \to \mathbb{R}$$

telle que

$$\mathrm{U}(\underline{\sigma}) = \sum_{\mathrm{Y}\subset\mathrm{X}} \Phi_\mathrm{U}(\underline{\sigma}|\mathrm{Y}) \qquad \text{pour} \qquad \mathrm{X} \subset \Lambda.$$

Cette fonction est obtenue par induction sur $|\mathrm{X}|$ (le cardinal de X), avec $\Phi_\mathrm{U} = 0$ sur $\mathrm{F}^\varnothing$.

La description précédente de la fonctionnelle d'énergie va nous permettre aisément, par la suite, de passer à la limite lorsque l'ensemble Λ devient grand. Il est important de remarquer que Φ est définie sur les configurations finies de spins et conserve la même valeur, peu importe si la taille de Λ tend vers l'infini.

Je voudrais encore vous présenter une quantité usuelle en physique statistique : la fonction de partition. Curieusement, ce n'est pas une fonction. Il faut prendre l'expression « fonction de partition » en bloc. C'est inspiré du mot allemand *zustandssumme* et cela pourrait être traduit différemment[2]. La fonction de partition est une somme de termes $e^{-\beta H_\Lambda(\underline{\sigma})}$ sur toutes les configurations de spins dans une région finie Λ :

$$Z_\Lambda = \sum_{\underline{\sigma}\in F^\Lambda} e^{-\beta H_\Lambda(\underline{\sigma})}.$$

(On remarque qu'il y a un nombre fini de configurations de spins et que la somme a donc un sens.) En physique, β représente l'inverse du produit de la constante de Boltzmann par la température absolue. En fait, ici, on n'aura jamais besoin ni de la température absolue ni de la constante de Boltzmann, mais simplement d'un paramètre β strictement positif. De plus, pour simplifier, on prendra souvent la constante β égale à 1, sans restreindre la généralité du modèle précédent. Le choix de l'exponentielle permet de passer d'une grandeur additive à une grandeur multiplicative. Il est d'usage de conserver le signe moins dans l'exponentielle (qui vient de la physique). Nous venons ainsi d'obtenir une fonction de partition Z_Λ à partir des quantités additives.

Pourquoi a-t-on introduit cette fonction de partition ? Pour construire une mesure de probabilité sur l'ensemble des configurations de spins dans la région Λ. En effet, chaque configuration $\underline{\sigma}$ apparaît avec la probabilité donnée par

$$\rho_\Lambda\{\underline{\sigma}\} = \frac{e^{-\beta H_\Lambda(\underline{\sigma})}}{Z_\Lambda}.$$

La fonction de partition est simplement un facteur de normalisation.

Résumons. Qu'est-ce que je vous ai présenté ? J'ai d'abord introduit un espace de configurations de spins sur un morceau de réseau, puis associé à chacune de ces configurations une probabilité faisant intervenir une quantité additive appelée énergie. On aurait pu, si l'on avait voulu commencer avec des quantités multiplicatives, définir l'énergie comme un logarithme.

---

2. Selon la traductrice personnelle du rédacteur, Katrin Gelfert, *zustand* signifie « état » et *summe* veut dire simplement « somme » ; ce qui donne « somme d'états » comme traduction littérale de *zustandssumme*. Merci, Katrin ! (N.d.R.)

Notions ergodiques

Jusqu'à présent, ce que j'ai fait est extrêmement trivial. La seule chose que l'on pourrait remarquer est que cette mesure de probabilité ρ_Λ jouit d'une propriété intéressante sur laquelle je voudrais insister. On peut définir pour une mesure de probabilité μ sur les $\underline{\sigma}$ une entropie qui est

$$h(\mu) = -\sum_{\underline{\sigma} \in F^\Lambda} \mu\{\underline{\sigma}\} \log \mu\{\underline{\sigma}\},$$

où l'on convient que $p \log p = 0$ si $p = 0$. On peut aussi définir la valeur moyenne d'une fonction $U : F^\Lambda \to \mathbb{R}$ comme étant simplement

$$\int U \, d\mu = \sum_{\underline{\sigma} \in F^\Lambda} \mu\{\underline{\sigma}\} U(\underline{\sigma}).$$

Alors, si l'on considère la différence

$$h(\mu) - \int H_\Lambda \, d\mu,$$

le maximum de cette quantité est obtenu précisément pour la mesure de probabilité ρ_Λ. Cette mesure de probabilité a donc comme propriété remarquable le fait de résoudre un certain problème variationnel : maximiser la différence entre l'entropie et la valeur moyenne de l'énergie. N'oublions pas que, jusqu'à présent, Λ est fini et qu'il n'y a donc aucune subtilité. Les subtilités vont arriver lorsqu'on essaiera de prendre la limite quand la taille de l'ensemble Λ tend vers l'infini.

Avant de prendre la limite et d'essayer d'examiner la propriété variationnelle analogue pour un ensemble infini, on va se simplifier un peu l'existence en supposant qu'il y a invariance par translation. On voit bien ce que signifie translater une configuration de spins par une translation $\theta : \mathbb{Z}^\nu \to \mathbb{Z}^\nu$ du réseau :

$$(\theta(\underline{\sigma}))_x = \sigma_{\theta(x)}.$$

On voit bien aussi ce que veut dire « invariance de l'énergie par translation », et plus précisément, « invariance de la fonction Φ par translation » :

$$\Phi(\theta(\underline{\sigma})) = \Phi(\underline{\sigma}).$$

Il est à noter que le fait de représenter une énergie H_Λ en fonction de Φ n'a en soi aucune importance. Cela dépend de ce que l'on va imposer à Φ. On ne demande pas de propriétés trop particulières, ni trop générales, qui conduiraient à une définition ou bien trop spécifique ou bien trop générale d'une fonction que l'on appellerait énergie. Si la fonction $\Phi = \{\Phi(\cdot|X)\}_X$ est définie sur les configurations de spins dans une région finie, on demande que $\Phi(\cdot|X)$ décroisse exponentiellement quand le diamètre de X tend vers l'infini. Cela donne une structure un peu locale au problème : la décroissance exponentielle à grande distance permet d'avoir une idée de ce qui est proche et de ce qui est loin.

Maintenant je veux introduire une fonction A_Φ définie sur les configurations infinies sur le réseau $\mathbb{Z}^\nu$ et qui va jouer le rôle d'une énergie. On sait définir une énergie pour des configurations de spins finies, mais naturellement, pour une configuration infinie, l'énergie va diverger. On peut par contre espérer définir une contribution d'un point à l'énergie du système infini et il y a un certain arbitraire dans la définition. On prend le point 0. Ainsi, $\Phi(\cdot|X)$ étant une contribution à l'énergie d'un ensemble fini $X \ni 0$, on va la diviser par le nombre de points de cet ensemble, simplement pour ne compter que la contribution du point 0. On pose donc :

$$A_\Phi(\underline{\sigma}) = -\beta \sum_{X \ni 0} \frac{1}{|X|} \Phi(\underline{\sigma}|X).$$

(Le signe moins sert à se débarrasser du signe moins arbitraire qui a été introduit auparavant.) C'est une quantité qui a le sens d'une énergie par site (on peut l'interpréter comme la contribution du point 0 à l'énergie infinie du système) et qui nous permettra d'oublier largement la fonction Φ que nous avions introduite auparavant. À nouveau le β va disparaître : on peut prendre $\beta = 1$ si l'on veut.

Rappelons maintenant quelques définitions de la théorie ergodique. On va appeler I l'ensemble des mesures de probabilité sur les configurations du système infini qui sont invariantes par translation. Rappelons que $\mathbb{Z}^\nu$ est le réseau, F est un ensemble fini et $F^{\mathbb{Z}^\nu}$ est l'ensemble des configurations sur le réseau infini (en chaque point $x \in \mathbb{Z}^\nu$, on se donne un élément de F). $F^{\mathbb{Z}^\nu}$ est un produit d'ensembles finis et donc compacts : $F^{\mathbb{Z}^\nu}$ est ainsi un ensemble compact, un ensemble de Cantor pour une topologie naturelle, la topologie produit. En somme, on s'intéresse à des mesures sur cet ensemble compact invariantes par l'effet des translations de $\mathbb{Z}^\nu$. Or, si l'on suit les idées de Bourbaki, une mesure sur un ensemble compact est un élément du dual de l'espace des fonctions continues. Et on sait que, pour la topologie faible, la sphère unité du dual de l'espace des fonctions continues est un ensemble compact : donc, I est un ensemble compact.

Une notion intéressante que l'on peut définir dans ce cadre-ci est la notion d'entropie moyenne. C'est simplement une fonction définie sur des mesures appartenant à I et prenant des valeurs positives ou nulles. On l'appelle aussi « invariant de Kolmogorov-Sinai » dans le cas où $\nu = 1$. Cependant, les définitions explicites sont toutes compliquées à détailler, car elles nécessitent un passage à la limite sur plusieurs indices. Je pense que décrire des concepts trop explicitement n'éclaire pas forcément la situation.

Grossièrement, on se donne une mesure invariante μ sur les configurations infinies. Cela définit une mesure sur chaque configuration d'une région finie Λ par projection $p_\Lambda : F^{\mathbb{Z}^\nu} \to F^\Lambda$ en faisant

$$\mu_\Lambda = \mu \circ p_\Lambda^{-1}.$$

Alors, comme précédemment, cette mesure-ci a une entropie qui est

$$h(\mu_\Lambda) = -\sum_{\underline{\sigma} \in F^\Lambda} \mu_\Lambda\{\underline{\sigma}\} \log \mu_\Lambda\{\underline{\sigma}\}.$$

Si l'on prend cette expression divisée par le nombre de points dans la région finie Λ et si l'on fait tendre la taille de la région Λ vers l'infini, on obtient une limite pourvu que l'on choisisse Λ habilement. On ne va pas, par exemple, prendre un domaine de forme trop biscornue, mais plutôt un domaine cubique ou sphérique. En prenant une suite adéquate de domaines, on a la limite exacte

$$h(\mu) = \lim_{|\Lambda| \to \infty} \frac{h(\mu_\Lambda)}{|\Lambda|}.$$

Cette limite s'appelle l'entropie moyenne de μ : c'est une entropie par unité de volume ou par site sur un réseau infini. La définition que je viens de donner ne montre pas cependant que c'est une quantité invariante. Mais, enfin, pour la suite de l'exposé, il suffit d'avoir en tête cette notion-ci. Finalement, c'est une quantité ergodique bien définie que nous allons essayer d'utiliser.

Mesure d'équilibre et mesure de Gibbs

Je vous ai raconté tout à l'heure que la mesure ρ_Λ maximise la différence entre l'entropie et la valeur moyenne de l'énergie. Comme nous l'avons vu, si une énergie H_Λ est définie sur un ensemble fini de configurations de spins, on peut introduire cette mesure ρ_Λ, appelée ensemble de Gibbs. Cette terminologie peut surprendre les mathématiciens. C'est toutefois une tradition de la physique. Ce que l'on appelle « ensembles » en mécanique statistique sont des mesures de probabilité. Enfin, on a une mesure de probabilité à laquelle est associé le nom de Gibbs (ou de Boltzmann si l'on préfère). Et comme nous l'avons vu précédemment, elle satisfait à un principe variationnel :

$$h(\rho_\Lambda) - \int H_\Lambda \, d\rho_\Lambda = \max\left\{h(\mu) - \int H_\Lambda \, d\mu : \mu \text{ probabilité sur } F^\Lambda\right\}.$$

La question que l'on peut se poser alors est de savoir si le principe variationnel va survivre quand on prend la limite sur une suite de volumes tendant vers l'infini. Cette question fait partie d'une collection de problèmes qui ont été clarifiés dans les années 1960.

Pour répondre à ce problème, nous allons introduire une première notion : la notion de mesure d'équilibre associée à une interaction Φ. On va exiger que cette mesure maximise une quantité similaire au cas des ensembles finis. On dira ainsi qu'une mesure de probabilité $\rho \in I$ est une mesure d'équilibre pour Φ si

$$h(\rho) + \int A_\Phi \, d\rho = \max_{\mu \in I}\left(h(\mu) + \int A_\Phi \, d\mu\right).$$

Cette formule tient compte de la somme de deux termes : l'entropie par site, $h(\rho)$, et la valeur moyenne de la fonction A_Φ, qui représente, au signe près, la contribution d'un site à l'énergie. On aurait pu prendre le signe opposé devant A_Φ et minimiser ; on aurait appelé cette quantité énergie libre. Le maximum de la quantité initiale est appelé pression et on le désigne habituellement par $P(A_\Phi)$. Par contre, une fois encore ne faites pas trop attention au mot pression. Cela correspond à une pression dans le cas d'un gaz sur un réseau. Cependant, dans le cas où ce n'est pas un gaz, le mot pression a un avantage sur énergie libre : c'est un seul mot au lieu de deux !

Jusqu'à présent, nous n'avons fait qu'introduire une définition. Ce qu'on appelle « mesure d'équilibre » est n'importe quelle mesure invariante par translation qui satisfait à un principe variationnel. C'est évidemment une tentative pour trouver un équivalent à un ensemble de Gibbs dans le cas d'un volume ou d'un réseau infinis.

Néanmoins, une autre notion apparaît tout aussi naturelle. C'est la notion de mesure de Gibbs. Cette fois-ci, on ne va plus supposer l'invariance par translation. On se donne donc à nouveau une interaction Φ et on va demander que la mesure ρ soit de mesure conditionnelle fixée, prescrite dans le cas où on indique quelle est la configuration en dehors d'une région finie donnée. On demande alors que ces mesures conditionnelles soient obtenues de la manière suivante :

$$\rho(\underline{\sigma}|\mathbb{Z}^{\nu} - \Lambda) = \exp\Big(-\beta \sum_{X\cap\Lambda\neq\emptyset} \Phi(\underline{\sigma}|X)\Big) \Big/ \text{normalisation.}$$

On prend l'exponentielle de manière à passer d'une notation additive à une notation multiplicative. Dans la formule précédente, les contributions à l'énergie sont obtenues en sommant sur tous les sous-ensembles X du réseau qui ont une intersection non vide avec la région Λ en question. Interprétée autrement, cette somme représente l'énergie de ce qui se trouve à l'intérieur de la région Λ, ayant fixé la configuration à l'extérieur de Λ. On normalise enfin pour obtenir une probabilité. On appelle mesure de Gibbs toute mesure vérifiant la formule précédente pour chaque région Λ.

Comment justifier cette définition ? Si on reprend la définition donnée pour un ensemble fini

$$\rho_{\Lambda}\{\underline{\sigma}\} = \exp\Big(-\beta \sum_{X\subset\Lambda} \Phi(\underline{\sigma}|X)\Big) \Big/ \text{normalisation,}$$

si on ne s'intéresse qu'à des configurations localisées à un sous-ensemble fini et si on fait tendre Λ vers l'infini, on peut s'attendre à ce que la suite de ces ρ_{Λ} converge vers une mesure de Gibbs.

Théorème DLR

On voit ainsi que les deux définitions que je viens de donner (mesure d'équilibre et mesure de Gibbs) sont des tentatives pour trouver des équivalents à la notion d'ensemble de Gibbs pour un réseau infini. Que cherche-t-on ? On cherche à montrer que la mesure naturelle ρ_{Λ} tend vers une mesure de Gibbs lorsque Λ tend vers l'infini. Si de plus la mesure limite est invariante par translations, nous allons voir qu'on obtient une mesure d'équilibre.

On ne peut pas s'attendre à l'existence de cette limite sans supposer certaines propriétés de décroissance vers l'infini de la fonction Φ. J'ai déjà mentionné la décroissance exponentielle avec le diamètre ; cette propriété pourrait être affaiblie, mais cela nécessiterait trop de détails. Et le théorème dit effectivement que la limite existe et qu'en plus une mesure d'équilibre est équivalente à une mesure de Gibbs invariante par translation :

Théorème (Dobrushin, Lanford-Ruelle). *Une mesure d'équilibre équivaut à une mesure de Gibbs invariante par translation.*

Les deux notions sont donc des notions naturelles. Si l'on prend une mesure de Gibbs et si l'on suppose en plus son invariance par translation, elle devient une mesure d'équilibre. Si l'on prend une mesure d'équilibre, on montre alors, et ce n'est pas du tout évident, que c'est une mesure de Gibbs.

À la suite d'un article de Lanford et moi-même, nous avons été prévenus par Jean Lascoux d'un travail semblable de Roland Dobrushin. Les deux implications ont été démontrées par Lanford et moi-même et une des deux par Dobrushin un peu avant. Il restait quand même l'autre direction, d'où le fait que l'on a gardé les initiales DLR.

Ceci est donc un théorème mathématique suggéré par la mécanique statistique, où l'on peut, si l'on veut, oublier la mécanique statistique. Si l'on prend la mécanique statistique au sérieux, on considère aussi des réseaux à plusieurs dimensions, plus difficiles à étudier. Pourquoi n'est-il pas intéressant de se restreindre à une seule dimension? Pourquoi est-ce que l'on fait de la mécanique statistique? La mécanique statistique est une théorie développée pour comprendre des phénomènes comme le changement de phase. Or, les physiciens savent depuis longtemps qu'il n'y a pas de changement de phase en une dimension. Cependant, ce qui peut paraître assez trivial pour les physiciens ne l'est pas pour les mathématiciens parce que le modèle reste facile à étudier et cela peut être une qualité lorsqu'on veut comprendre en détail.

Réseau à une dimension

Occupons-nous du cas à une dimension. Dans ce cas, l'espace de configurations devient $F^{\mathbb{Z}}$. Mais on va introduire une petite complication : deux spins successifs ne peuvent pas être arbitraires. On va donc imposer une condition sur les paires de spins successifs qui sont autorisées. (Si le spin est une variable qui vaut 0 quand le site est inoccupé et 1 quand il est occupé, on peut par exemple imposer que deux sites contigus ne peuvent pas être simultanément occupés.) C'est une condition de type markovien et il y a intérêt à imposer de telles conditions. On introduit donc une matrice (t_{ij}) dont les éléments sont soit 0, soit 1 : 0 si la paire $(i, j) \in F^2$ est interdite, 1 si elle est autorisée. On va aussi demander que, si l'on prend cette matrice (t_{ij}) à une puissance suffisamment grande, elle a tous ses éléments différents de zéro :

$$(t^N)_{ij} > 0 \text{ pour N assez grand.}$$

C'est une propriété de *mixing* : cette matrice (t_{ij}) définit ce qu'on appelle une chaîne de Markov mélangeante.

Ainsi, au lieu de considérer l'ensemble de Cantor $F^{\mathbb{Z}}$, on va considérer un sous-ensemble Σ qui est formé des configurations telles que toutes les paires de spins consécutives soient autorisées :

$$\Sigma = \{\underline{\sigma} = (\sigma_x)_{x\in\mathbb{Z}} : t_{\sigma_x, \sigma_{x+1}} = 1 \text{ pour tout } x\}.$$

C'est à nouveau un ensemble de Cantor. Il y a des raisons (que vous allez comprendre dans un moment) qui nous amènent à regarder ce cas un peu plus général. Cet ensemble de Cantor Σ avec la translation $\tau : \Sigma \to \Sigma$ donnée par

$$(\tau\underline{\sigma})_x = \sigma_{x+1}$$

forme un système dynamique (Σ, τ) qui s'appelle un sous-shift de type fini. On dit « de type fini » parce que, au lieu d'imposer des conditions sur les plus proches voisins, on pourrait garder en mémoire un nombre fini de voisins (c'est une expression traditionnelle). Il y a même sans doute moyen de traduire shift en français, mais je ne me souviens plus comment[3].

On peut définir une métrique sur cet ensemble de Cantor en disant que la distance entre une configuration $\underline{\sigma}$ et une configuration $\underline{\xi}$ est λ^n, où n est choisi de telle sorte que les spins à distance moindre que n de l'origine doivent être tous égaux, à savoir

$$n = \inf\{|x| : \sigma_x \neq \xi_x\},$$

et λ est un nombre compris entre 0 et 1. L'idée est que, si deux configurations coïncident sur une région assez grande autour de l'origine, elles sont en fait proches, exponentiellement proches l'une de l'autre. Pour définir exactement la métrique, il faudrait choisir en plus λ. Ce

3. Le terme français est « décalage », et l'expression « sous-décalage de type fini » est effectivement employée en français. (N.d.R.)

que pour l'instant je ne veux pas faire. De toute façon, c'est une métrique naturelle sur cet ensemble de Cantor et elle respecte la topologie de Cantor qui est introduite par le produit.

On a maintenant un théorème assez détaillé que je vais vous asséner sans donner trop d'explications sur sa preuve. J'essaierai de revenir à la fin de mon exposé sur la raison pour laquelle c'est vrai. Avant de l'énoncer, je dois faire une remarque. On a hésité un petit peu entre l'usage de Φ (l'interaction) et A_Φ (la contribution d'un point du réseau à l'énergie). En fait, dire que Φ décroît exponentiellement à l'infini et dire que A est une fonction continue höldérienne pour la métrique précédente revient à dire la même chose. Ainsi, cela nous permet d'oublier les interactions et de parler uniquement de fonctions höldériennes sur un ensemble de Cantor.

Théorème (Formalisme thermodynamique). *Soit* $A = A_\Phi : \Sigma \to \mathbb{R}$ *höldérienne. Alors, il existe une unique mesure d'équilibre* ρ_A *associée à la fonction* A*, laquelle est aussi l'unique mesure de Gibbs donnée par* A*. La correspondance* $P : \mathit{Hölder}(\Sigma, \mathbb{R}) \to \mathbb{R}$ *définie par*

$$P(A) = \max_{\rho \in I} \left(h(\rho) + \int A \, d\rho \right)$$

est analytique réelle convexe. La mesure ρ_A *est la dérivée de* P *en* A *et satisfait à :* $\mathit{supp}(\rho_A) = \Sigma$*. En outre, on vérifie la décroissance exponentielle des corrélations, c'est-à-dire, il existe* $a > 0$ *tel que*

$$\left| \int B \cdot C \circ \tau^k \, d\rho_A - \int B \, d\rho_A \int C \, d\rho_A \right| \leqslant \mathit{constante}\, e^{-a|k|}$$

pour $B, C \in \mathit{Hölder}(\Sigma, \mathbb{R})$.

Comme nous sommes en une dimension et, comme je vous l'ai dit, en une dimension il n'y a pas de changement de phase, la situation est triviale du point de vue de la physique. Il est cependant remarquable d'observer qu'il existe une unique mesure d'équilibre pour une fonction höldérienne donnée en dimension 1. Il existe aussi une unique mesure de Gibbs et les deux sont évidemment équivalentes. C'est une situation un peu désespérante du point de vue des physiciens (qui voudraient voir des changements de phases), mais elle a l'avantage de la simplicité du point de vue des mathématiciens.

On rappelle que la pression P(A) est égale au maximum, parmi les mesures de probabilité invariantes par translation, de la somme de l'entropie $h(\rho)$ et de la moyenne de la fonction A par rapport à ρ. Il se trouve qu'en une dimension cette fonction P est analytique réelle et convexe sur l'espace de Banach des fonctions höldériennes. Malheureusement, la preuve est vraiment très difficile.

L'espace de Banach sur lequel cette fonction est définie est inclus dans l'espace des fonctions continues sur Σ. On peut dériver cette fonction, puisqu'elle est analytique réelle donc différentiable. Sa dérivée est un élément du dual de l'espace des fonctions höldériennes qui s'étend, en fait, en un élément du dual de l'espace des fonctions continues, autrement dit une mesure. Et cette mesure est précisément ρ_A, l'unique mesure d'équilibre ou mesure de Gibbs. On a commencé par introduire une notion de mesure de Gibbs ou d'équilibre, on a défini ensuite une notion de pression P et on s'est rendu compte que cette mesure décrit le plan tangent du graphe de cette fonction. Il est, en plus, utile de savoir que, lorsque A est höldérienne, le support de cette mesure ρ_A est tout l'espace Σ, ou encore que ρ_A donne une masse positive à tous les ouverts.

Je voudrais faire encore une dernière remarque. On appelle fonction de corrélation entre deux fonctions höldériennes B et C la quantité

$$\int B \cdot C \circ \tau^k \, d\rho_A - \int B \, d\rho_A \int C \, d\rho_A.$$

Le théorème nous assure que la valeur absolue de cette quantité décroît à l'infini de manière exponentielle.

Avant de donner une idée des techniques employées dans la preuve de ce résultat, je voudrais, dans la seconde partie de cet exposé, vous parler des systèmes dynamiques différentiables et vous montrer d'abord l'intérêt de cette théorie.

Commentaires et références

J'ai réuni quelques références de base (voir la bibliographie à la fin de la leçon). Je commence par citer l'article original de Dobrushin : [9] ; et celui de Lanford et moi-même : [11].

J'ai rencontré Bowen qui, comme je vous l'ai dit, est malheureusement décédé jeune. Il avait le génie de la simplicité. Face à une question confuse et compliquée, il trouvait toujours des réponses simples. Il a écrit un petit fascicule où il explique la théorie du formalisme thermodynamique : [7]. Il expliquait les concepts de manière extrêmement claire. Il parlait très lentement et allait droit au but. Une fois Lanford m'a invité à aller donner un cour de niveau *graduate* à Berkeley. Dans l'auditoire se trouvait Rufus Bowen qui a écrit par la suite ce petit fascicule. J'ai écrit une version plus complète mais plus tardive : [15]. Parmi les nombreuses publications plus récentes sur ce sujet, on peut citer encore le livre de Gallavotti, Bonetto et Gentile : [5].

Dynamique hyperbolique

Ensemble hyperbolique, variétés stable et instable

Maintenant je vais oublier pour un moment la mécanique statistique et je veux parler de systèmes dynamiques différentiables. Ainsi, je prends une variété compacte M et une application f de cette variété dans elle-même. On va supposer que cette fonction a r dérivées continues, ou même qu'il s'agit d'un difféomorphisme C^r (i.e. elle est bijective et son inverse a aussi r dérivées continues). Éventuellement on pourra parler de flot, c'est-à-dire qu'au lieu de temps discret, au lieu de considérer les itérés f^n de f, on va supposer qu'il y a un temps continu, donc une famille $(f^t)_{t\in\mathbb{R}}$ (à un paramètre t) d'applications de la variété dans elle-même. Cela est un cadre tout à fait général, que l'on l'appelle la *smooth dynamics* (la dynamique différentiable).

Et la première chose que l'on comprend, si l'on essaye d'étudier la dynamique différentiable, est qu'il y a une telle multitude de possibilités que l'on est complètement perdu. Donc, il faut essayer de se restreindre à quelques situations contrôlées, où certaines conditions sont imposées. Cela a été fait d'une part par Anosov, d'autre part par Steve Smale, en supposant l'hyperbolicité.

Quelle est l'hypothèse d'hyperbolicité? Eh bien, on va supposer qu'il existe un sous-ensemble $\Lambda \subset M$, compact et invariant pour le difféomorphisme f, et qui satisfait aux conditions suivantes. D'abord, l'espace tangent à la variété restreinte à Λ doit être la somme de deux sous-fibrés

$$T_\Lambda M = E^s \oplus E^u \quad \text{(somme de Whitney)}$$

et on exige que ces sous-fibrés dépendent continûment du point où on se trouve. Les sous-fibrés E^s et E^u s'appellent respectivement le fibré stable et le fibré instable (ou contractant et dilatant). On demande en outre que, si l'on prend un vecteur dans la direction stable, l'effet de l'application tangente du difféomorphisme $T_x f$ va être de contracter uniformément, c'est-à-dire, qu'il doit exister des constantes $C > 0$ et

$\theta > 1$ telles que, pour tout point $x \in \Lambda$ et pour chaque entier $n \geqslant 0$,

$$\|T_x f^n v\| \leqslant C\theta^{-n}\|v\| \qquad \text{si } v \in E^s.$$

Par contre, dans la direction instable, l'effet sera de dilater uniformément, en ce sens qu'en fait les inverses contractent uniformément

$$\|T_x f^{-n} v\| \leqslant C\theta^{-n}\|v\| \qquad \text{si } v \in E^u.$$

L'ensemble Λ est dit, alors, un ensemble *hyperbolique.*

Il faut remarquer que c'est une définition qui a le grand avantage de donner lieu à des théorèmes intéressants. Cela n'est pas immédiatement évident, mais cette particularité se révèle être vraie et c'est, sans doute, le génie d'Anosov et de Smale que de s'en être rendu compte.

La situation que l'on a, quand on suppose qu'il y a l'hyperbolicité, est donc qu'en chaque point de l'ensemble hyperbolique on a une direction qui est dilatante et une direction qui est contractante et, en plus, ces directions dépendent continûment du point où on se trouve. Ce qui n'est pas évident, mais qui est très utile, est qu'il y a une forme non linéaire de cette condition : les variétés stable et instable. Plus spécifiquement, si l'on suppose les conditions linéaires que j'ai indiquées là, il existe aussi des variétés non linéaires qui sont tangentes en tout point $x \in \Lambda$ aux sous-fibrés respectifs et qui sont aussi dilatées ou contractées exponentiellement. Ces variétés sont, en fait, uniques. Donc, à partir de la condition que l'on a imposée, c'est-à-dire, la condition linéaire qui définit l'ensemble hyperbolique, on peut obtenir l'existence des variétés stable et instable, qui sont des variétés non linéaires.

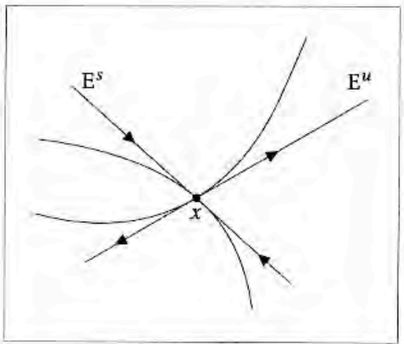

Fig. 3. Les variétés stable et instable.

La variété stable de $x \in \Lambda$ peut être décrite localement comme

$$\mathcal{V}_x^s = \{y \in \mathrm{M} : d(f^n x, f^n y) < \delta \text{ pour } n \geqslant 0\}.$$

Je remarque que la définition donnée ici est de nature locale et cela donne un morceau de variété stable qu'on appelle variété stable locale. Pour la variété instable, notée $\mathcal{V}_x^u$, on remplace simplement le difféomorphisme f par son inverse. Et on pourrait aussi définir des variétés globales. Cependant, elles ont le désavantage de se replier sur elles-mêmes. D'autre part, les variétés locales ont le désavantage de ne pas être uniques parce qu'elles dépendent du fait qu'on va les prendre grandes ou petites. Les variétés globales sont uniques, mais ne sont pas des variétés bien plongées. Donc, en général, on se contente de regarder les variétés stables et instables locales. La contraction dans la variété stable est exponentielle. Cela nous fournit des outils très intéressants.

Il y a quelques aspects qu'il faut garder présents à l'esprit. Premièrement, les variétés stable et instable ne sont définies à priori que sur cet ensemble hyperbolique Λ et pas en dehors. Pourtant, on peut supposer que toute la variété est hyperbolique. C'est le cas des systèmes d'Anosov. D'ailleurs, par définition, si toute la variété M est hyperbolique, on a un difféomorphisme d'Anosov. La variété dans ce cas-là est remplie par des variétés stables et des variétés instables. Donc, on a des sortes de feuilletages stable et instable. En fait, ce n'est qu'à moitié vrai, car ces variétés stables et instables sont différentiables, mais ne forment pas un feuilletage différentiable : les variétés dépendent de manière continue mais pas différentiable du point où on se trouve.

Deuxièmement, je note qu'au lieu de considérer des systèmes uniformément hyperboliques comme je l'ai indiqué, on pourrait peut-être essayer de considérer des systèmes non uniformément hyperboliques. Bah, c'est beaucoup plus difficile ! Là je ne vais pas en parler parce que c'est vraiment une autre histoire. Pour les gens qui ont entendu parler de la théorie de Pesin, c'est une des approches au cas non uniforme, qui a donné lieu à des études par François Ledrappier, dont Philippe Thieullen ici présent pourrait vous parler.

Difféomorphismes d'Anosov et axiome A

Nous allons donc revenir aux choses simples et uniformément hyperboliques. Je l'ai dit à un moment : un difféomorphisme d'Anosov est un difféomorphisme pour lequel la variété M toute entière est un ensemble hyperbolique. Steve Smale a introduit une autre notion, et je veux prendre une partie de ce qu'il a fait en définissant ce que l'on appelle un attracteur satisfaisant à un axiome A.

D'abord, on va supposer que notre ensemble hyperbolique est un attracteur, c'est-à-dire que Λ a un voisinage ouvert U qui, sous l'application du difféomorphisme f, se contracte,

$$f\left(\overline{\mathrm{U}}\right) \subset \mathrm{U},$$

et, en plus, tel que l'intersection de tous les itérés positifs de f appliqués à U donne l'ensemble Λ,

$$\Lambda = \bigcap_{n \geqslant 0} f^n \mathrm{U}.$$

Intuitivement, cela veut dire que l'on a bien un attracteur, car on a un voisinage qui se contracte sur Λ.

On va supposer aussi qu'il existe une orbite dense dans l'ensemble Λ, c'est-à-dire qu'il existe un point $x \in \Lambda$ tel que $\{f^n x\}$ soit dense dans l'ensemble hyperbolique. Cette propriété s'appelle « transitivité topologique ».

Ces deux hypothèses sont tout à fait naturelles. Par contre, il n'est pas évident que ce soit une bonne idée d'introduire la notion selon laquelle l'ensemble des points f-périodiques est dense dans Λ. C'est Smale qui a eu cette vision géniale en percevant que ce concept était la bonne idée à introduire.

Bien, résumons brièvement les propriétés qui définissent un attracteur satisfaisant à un axiome A : on demande que l'ensemble hyperbolique Λ soit un attracteur avec une orbite dense et tel que

$$\Lambda = \overline{\{x : x \text{ est périodique}\}} \cap \Lambda. \qquad (*)$$

Pourquoi cette dernière hypothèse est-elle une bonne idée ? Smale a vu que les variétés stables et instables ne couvrent pas nécessairement toute la variété. En particulier, si l'on a la variété stable d'un point $x \in \Lambda$ et la variété instable d'un autre point $y \in \Lambda$, l'intersection des deux n'appartient peut-être pas à l'ensemble hyperbolique.

Toutefois, moyennant la condition $(*)$, de Smale, c'est bien le cas. Si cette hypothèse est vérifiée, l'intersection d'une variété stable et d'une variété instable de points dans l'attracteur sera encore un point de l'attracteur ou d'un ensemble hyperbolique. Techniquement c'est

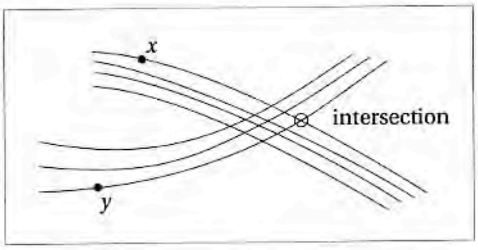

Fig. 4.

une propriété très utile qui s'appelle « structure produit locale ». Mais, comme je ne vais pas rentrer maintenant dans les détails, je laisse cela un peu à côté.

Pour les gens qui n'ont jamais vu de difféomorphismes d'Anosov, je veux donner un exemple : le *cat map* d'Arnold. On prend comme variété compacte un tore à deux dimensions, formé des points du plan modulo 1, et usuellement noté $\mathbb{T}^2$. On considère, alors, une application donnée par la matrice

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

dont les éléments sont des entiers. Les éléments de la matrice inverse sont aussi des entiers parce que le déterminant est 1. Eh bien, cette application du tore $\mathbb{T}^2$ dans lui-même est un automorphisme hyperbolique. Si l'on regarde les directions propres de cette matrice qui sont orthogonales, il y a une valeur propre qui est plus grande que 1 et l'autre valeur propre qui est plus petite que 1. Donc, on a contraction dans une direction et une expansion dans la direction perpendiculaire, et tout cela uniformément, puisque visiblement c'est le même phénomène partout sur ce tore.

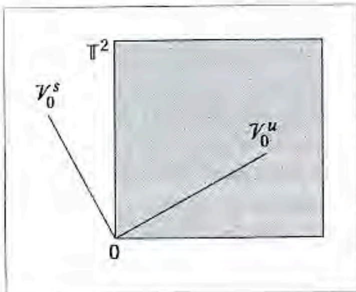

Fig. 5.

Je crois qu'en fait c'est René Thom qui a été le premier à considérer ces automorphismes hyperboliques du tore. Mais, c'est Arnold qui a fait le dessin du chat dans un livre, celui d'Arnold et Avez [3] : *Problèmes ergodiques de la mécanique classique*. Celui-ci reste un très, très bon bouquin d'introduction à ce genre de sujet. Bien, maintenant vous avez au moins un exemple de difféomorphisme d'Anosov. Et il y a des exemples de flots d'Anosov qui sont bien connus. C'est le cas des flots géodésiques sur une variété à courbure négative (non nécessairement constante). Donc, il existe aussi des exemples de systèmes dynamiques d'Anosov à temps continu.

Commentaires et références

Je vous donne quand même les deux références fondamentales sur le sujet : celle d'Anosov [2], qui est un relativement gros fascicule, et celle de Smale [20], qui est un article extrêmement court. L'article de Smale est, à mon avis, une des œuvres d'art de la littérature mathématique. C'est écrit de manière, je dirais, presque désinvolte, mais cet article se lit avec une grande facilité. En plus, à chaque fois qu'on le relit, on trouve des choses nouvelles et intéressantes. C'est vraiment un très, très bon article qui a eu d'ailleurs une influence absolument énorme et fondatrice sur le sujet.

Dynamique symbolique

Structure produit locale

Je veux passer à présent à la dynamique symbolique. D'une manière claire, je veux essayer de coller ensemble les deux théories dont je vous ai parlé : d'une part, les systèmes de spins sur un réseau et, d'autre part, les systèmes dynamiques différentiables hyperboliques. Je considère ainsi soit un attracteur satisfaisant à un axiome A, soit un difféomorphisme d'Anosov supposé mélangeant. C'est simplement une question de simplification de dire que c'est mélangeant, on peut toujours se ramener à ce cas-là.

Je l'ai mentionné, l'hypothèse que l'on a fait en admettant que l'ensemble des points périodiques était dense a pour conséquence que, si l'on prend des points qui appartiennent à l'ensemble Λ, alors l'intersection des variétés stable et instable appartient également à Λ. En termes plus exacts, pour tous $x, y \in \Lambda$, si l'on considère les variétés stable et instable locales suffisamment petites, $\mathcal{V}_x^s \cap \mathcal{V}_y^u$ se réduit à un unique point appartenant à Λ, lequel est noté $[x, y]$.

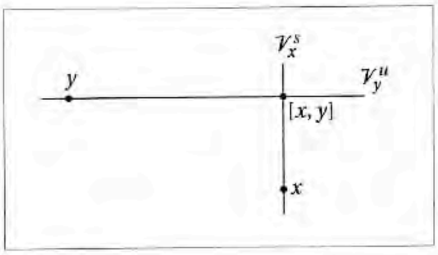

Fig. 6. La structure produit locale.

Ayant une structure produit locale, il y a un concept naturel qui est celui de « rectangle », défini comme suit. On prend, dans l'ensemble Λ, un bout de variété instable et un bout de variété stable. Il y a avantage à prendre la variété instable horizontale et la variété stable verticale, parce que comme ça l'effet de la gravité va aplatir la variété stable sur la variété instable. (Bon, c'est simplement un aide-mémoire !) On considère ensuite un ensemble C qui est l'intersection de cette variété

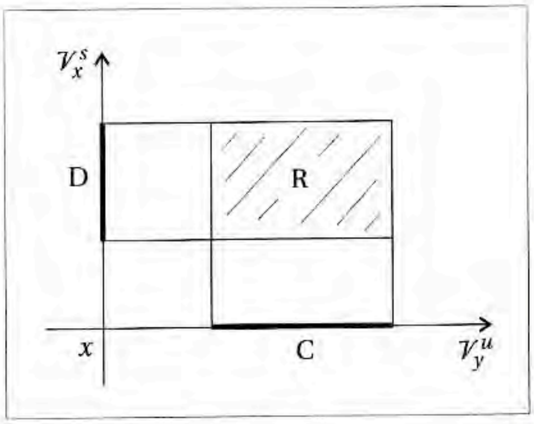

Fig. 7.

instable locale avec l'ensemble Λ et D qui est l'analogue pour la direction stable. Si l'on prend par les points de D les variétés instables et par les points de C les variétés stables, on regarde les intersections. Elles sont uniques parce qu'il y a une seule intersection localement, et elles appartiennent à l'ensemble Λ. On a donc une structure qui est localement un produit, $R = [C, D]$, et qu'on est tenté d'appeler un rectangle.

Toutefois, on introduit une condition plus technique. On demande encore que l'ensemble C soit la fermeture de son intérieur dans la variété instable intersectée avec l'ensemble hyperbolique, $\mathcal{V}_x^u \cap \Lambda$, et que l'ensemble D vérifie la même propriété par rapport à $\mathcal{V}_y^s \cap \Lambda$. Cela donne la véritable notion de rectangle, et on peut définir ses frontières stable et instable de manière naturelle. Il faut prendre les frontières de C et de D multipliées par D et C respectivement :

$$\partial^s R = [\partial C, D] \quad \text{et} \quad \partial^u R = [C, \partial D].$$

Enfin, on constate que $\partial R = \partial^s R \cup \partial^u R$.

Partition de Markov

Maintenant qu'on a défini un rectangle, on peut introduire l'idée de partition de Markov. En fait, une partition de Markov n'est pas une partition, c'est un recouvrement de l'ensemble Λ par un nombre fini de rectangles $R_i = [C_i, D_i]$. Mais on demande quand même que cela soit presque une partition, c'est-à-dire que les intérieurs des rectangles ne se rencontrent pas : $\mathrm{int}R_i \cap \mathrm{int}R_j = \emptyset$ si $i \neq j$. On va alors exiger une propriété que j'écris ici

$$x \in \mathrm{int}R_i \cap f^{-1}\mathrm{int}R_j \Rightarrow f[C_i, x] \supset [C_j, fx] \quad \text{et} \quad f[x, D_i] \subset [fx, D_j],$$

mais je vais vous expliquer de quoi il s'agit. Quand on applique le difféomorphisme f à un rectangle R_i, on obtient un ensemble sans nouvelles frontières dans la direction stable. Comme indiqué par la figure 8, on va élargir le rectangle dans la direction instable mais contracter dans la direction stable et on remarque un petit bout, l'intersection, qui sera un morceau d'un rectangle R_j préexistant. En somme, quand on applique f, on crée de nouvelles lignes horizontales mais pas de nouveaux segments verticaux.

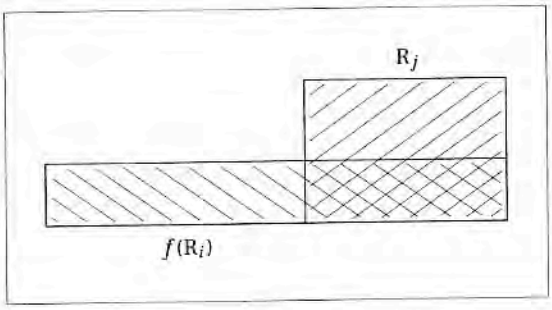

Fig. 8.

Il suffit de lire attentivement ce qui est écrit avant. C'est un tout petit peu différent, mais cela vous donne de manière formelle l'idée intuitive de la partition de Markov que je viens de vous décrire.

Je voudrais faire quelques commentaires. On peut, bien sûr, donner des définitions comme celle-ci. Cependant, cela ne présente d'intérêt que si l'on montre qu'il existe des objets de cette sorte-là. Et c'est le génie, je dirais, de Sinai d'avoir vu que l'on pouvait démontrer l'existence de partitions de Markov dans le cas des difféomorphismes d'Anosov. Puis, Bowen est venu, a étendu le résultat de Sinai au cas des difféomorphismes satisfaisant à l'axiome A de Smale, en changeant les définitions. Au lieu de parler de frontières d'un rectangle comme je viens de le faire et comme l'avait fait Sinai, il a introduit une nouvelle définition de partition de Markov en imposant que

$$f\partial^s \subset \partial^s \text{ où } \partial^s = \bigcup_i \partial^s R_i \text{ et}$$

$$f^{-1}\partial^u \subset \partial^u \text{ où } \partial^u = \bigcup_i \partial^u R_i.$$

Cette définition-ci est valable dans le cas « axiome A » et pas seulement « Anosov ». Bowen a ainsi donné une démonstration plus simple. Il en a donné, en fait, plusieurs successives dont la plus récente, qui utilise la notion de *shadowing*, est une démonstration relativement facile à comprendre. Bon, je formule le théorème dû à Sinai et à Bowen.

Théorème. *Il existe des partitions de Markov de diamètre arbitrairement petit pour un difféomorphisme satisfaisant à l'axiome A de Smale ou un difféomorphisme d'Anosov.*

J'avais donné un exemple de difféomorphisme d'Anosov : le *cat map* d'Arnold. En fait, avant Sinai, il avait été découvert par Roy Adler et Benjamin Weiss qu'il existait pour ce système dynamique particulier une partition de Markov. Il s'agit de la partition en deux rectangles (à des translations près du réseau, évidemment) décrite par la figure 9.B. C'est en effet une partition de Markov, mais elle n'est pas arbitrairement petite. Le fait d'avoir un petit diamètre pour les rectangles est demandé parce que lorsqu'on prend f appliquée à un rectangle R et qu'on l'intersecte avec un autre rectangle R′, on veut avoir un seul morceau : on ne veut pas que fR se recouvre et soit réintersecté comme dans la figure 9.A. Cette situation serait désagréable pour les applications. Malheureusement cela se passe pour la partition de Markov en deux rectangles. Alors, au lieu de prendre deux rectangles, on en prend cinq comme indiqué aussi par la figure 9.B, et on n'a plus cette ambiguïté.

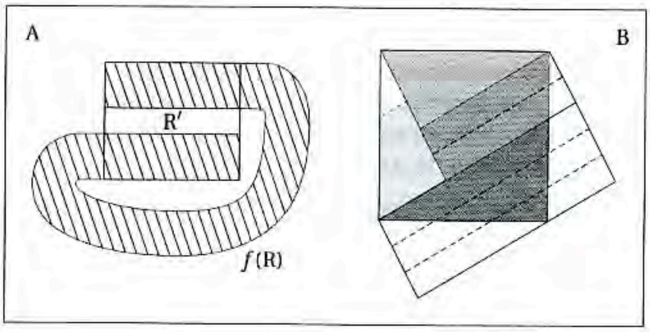

Fig. 9.

Représentation symbolique

Voilà, on a vu le théorème de Sinai et Bowen. Cela donne lieu à ce que l'on appelle une dynamique symbolique. D'abord, on va prendre un ensemble fini : l'ensemble des rectangles de la partition de Markov. On va définir une matrice de transition $(t_{R_iR_j})$ dont les éléments sont déterminés de la façon suivante

$$t_{R_iR_j} = \begin{cases} 1 & \text{si int } R_i \cap f^{-1}\text{int } R_j \neq \emptyset \\ 0 & \text{sinon.} \end{cases}$$

Ayant introduit un ensemble fini et une matrice de transition, nous avons un sous-shift de type fini comme celui qui correspondrait à un système de spins sur un réseau, sauf qu'ici il n'est plus question de spins, bien entendu. Le sous-shift sera le système dynamique formé par l'ensemble

$$\Sigma = \{\underline{\sigma} = (\sigma_k)_{k\in\mathbb{Z}} : t_{R_{\sigma_k}R_{\sigma_{k+1}}} = 1 \text{ pour tout } k\}$$

avec la translation

$$\tau : \Sigma \to \Sigma \text{ telle que } (\tau\underline{\sigma})_k = (\underline{\sigma})_{k+1}. \qquad (\dagger)$$

Sinai connaissait bien les travaux de Dobrushin sur les systèmes de spins, et il connaissait bien aussi les travaux d'Anosov sur les difféomorphismes d'Anosov. Il a alors compris que l'on pouvait appliquer l'un à l'autre. Apparemment des idées de ce type-là préexistaient, comme le montre l'exemple d'Adler et Weiss. Il paraît même qu'il y a des codifications de ce genre dans Hadamard, mais je ne peux pas le garantir, je n'ai pas regardé moi-même. De toute façon, ce qui existait avant n'est pas comparable à ce qui a existé après. La raison de cette révolution est que Sinai a vu une approche qui permettait de traduire la dynamique différentiable d'un difféomorphisme d'Anosov en une dynamique symbolique pour des suites de symboles.

Théorème. *Si* $\underline{\sigma} = (\sigma_k)_{k\in\mathbb{Z}} \in \Sigma$, *alors* $\bigcap f^{-k}R_{\sigma_k}$ *se réduit à un point, noté* $\pi\underline{\sigma}$. *En plus,*

(a) $\pi : \Sigma \to \Lambda$ *est continue (et même lipschitzienne) et surjective ;*

(b) $\pi \circ \tau = f \circ \pi$ *;*

(c) $\pi^{-1}x$ *est un unique point pour* x *appartenant à l'ensemble résiduel*

$$\Lambda - \bigcup_{k\in\mathbb{Z}} f^k(\partial^s \cup \partial^u);$$

(d) *pour tout* $x \in \Lambda$, $|\pi^{-1}x| < |F|^2$.

Les éléments de Σ sont ces suites de spins, si vous voulez ; nous les appelons des symboles. La première conséquence du théorème est que, si l'on a une telle suite de symboles qui est permise, qui est dans Σ, et si l'on considère l'intersection des rectangles auxquels on va appliquer f^{-k}, on a un point unique. Cela arrive simplement parce qu'appliquer f à un rectangle l'aplatit ou le contracte dans une direction et le dilate dans l'autre, avec pour résultat qu'il y a un unique point d'intersection. Ce qui définit une application π des suites de symboles dans les points de la variété.

Alors, quelles sont les propriétés de cette application π ? La première est qu'elle est continue. En fait, la fonction π est continue et surjective sur l'ensemble Λ. On peut mettre, en vérité, continue höldérienne ou lipschitzienne. Vous vous souvenez que, dans la définition de la métrique sur l'ensemble Σ, il y avait un petit λ qui avait été laissé arbitraire. Alors, si on le laisse arbitraire, on a une application qui est continue höldérienne, mais, si on le choisit bien, la fonction π devient lipschitzienne.

Ensuite on a cette relation

$$\pi \circ \tau = f \circ \pi,$$

où, je vous le rappelle, τ est la translation sur le sous-shift de type fini (†) et f est le difféomorphisme. C'est la représentation symbolique : l'application π remplace un difféomorphisme f par un shift sur un sous-type de symboles. On a, d'un côté, une structure différentiable et, d'autre part, quelque chose de très discret, mais simple à analyser.

Malgré le fait que l'application π est surjective, elle n'est pas inversible en ce sens que $\pi^{-1}x$ n'est pas unique partout. Cependant, $\pi^{-1}x$ est unique pour x dans un ensemble résiduel, c'est-à-dire dans un ensemble qui contient une intersection dénombrable d'ouverts denses. En plus, on peut montrer que le nombre de préimages est borné par $|F|^2$. Pourtant c'est l'item (c) du théorème qui est la propriété importante. En fait, je ne vais pas entrer dans les détails, mais on a raffiné les résultats au point où on peut compter les points périodiques pour un difféomorphisme satisfaisant à l'axiome A ou pour un difféomorphisme d'Anosov en termes des points périodiques du shift, lesquels sont faciles à analyser. Et cela nous permet, en particulier, de montrer qu'une certaine fonction zêta (qui compte les points périodiques pour un axiome A) est rationnelle, ce qui avait été conjecturé mais non démontré par Smale. Enfin, ce qui est important est de se souvenir de ce diagramme-ci qui échange le difféomorphisme f et le shift τ :

$$\begin{array}{ccc} \Sigma & \xrightarrow{\tau} & \Sigma \\ \pi\downarrow & & \pi\downarrow \\ \Lambda & \xrightarrow{f} & \Lambda \end{array}$$

Mesure de Gibbs et mesure SRB

On est presque arrivé à pouvoir étudier la théorie ergodique des systèmes hyperboliques. Ce qu'il faut est utiliser cette dynamique symbolique pour transporter les mesures entre la variété et le sous-shift de type fini. Alors, comment peut-on définir des états de Gibbs sur la variété ? Je vous rappelle que les états de Gibbs ou d'équilibre dépendent d'une fonction A continue höldérienne. Donc, je vais supposer que l'on a, sur la variété dans le cas Anosov ou sur l'ensemble hyperbolique pour un axiome A, une fonction continue höldérienne réelle

$$\mathrm{A} : \Lambda \to \mathbb{R}.$$

Et, alors, j'affirme (d'autres gens l'ont affirmé avant moi) qu'il existe une mesure de probabilité unique ρ_{A} qui est invariante par le difféomorphisme et qui rend maximum la somme de l'entropie et de la valeur moyenne de A, c'est-à-dire

$$h_f(\rho_{\mathrm{A}}) + \int \mathrm{A}\, d\rho_{\mathrm{A}} = \max_{\rho \in \mathrm{I}_f} \left(h_f(\rho) + \int \mathrm{A}\, d\rho \right).$$

C'est une chose miraculeuse pour la théorie ergodique sur une variété différentiable, et on la démontre en se ramenant au cas moins miraculeux où l'on a un système de spins sur un réseau à une dimension. Ce que l'on va faire est, d'abord, utiliser π pour transférer la fonction A sur la variété en la fonction $\mathrm{A} \circ \pi$ sur le sous-shift de type fini. Cette fonction sera à nouveau une fonction continue höldérienne. Pour $\mathrm{A} \circ \pi$ on va trouver un état de Gibbs unique, $\rho_{\mathrm{A}\circ\pi}$. Puis on va envoyer cet état de Gibbs unique sur la variété par l'application π en déterminant

$$\rho_{\mathrm{A}} = \rho_{\mathrm{A}\circ\pi} \circ \pi^{-1}.$$

La seule chose qui pourrait aller mal est qu'il y ait des problèmes là où l'application π n'est pas inversible. On s'en tire en remarquant que les points où elle n'est pas inversible sont sur les frontières. Par suite de l'invariance de la mesure, on voit que, s'il y avait un ensemble difficile, son complémentaire serait un ouvert non vide. Et on utilise, alors, que les mesures de Gibbs pour un sous-shift de type fini ont pour support tout l'ensemble Σ. Donc, s'il y avait un problème quelque part, par ergodicité, l'ensemble où il y a éventuellement un problème aurait pour mesure soit 0, soit 1. Dans ce cas-ci, il aurait pour mesure 0.

Voilà la raison pour laquelle on peut passer de la théorie ergodique sur une variété à la théorie ergodique sur ce sous-shift de type fini.

Dans cet esprit, on remarque encore que le maximum atteint par ρ_A, la pression si l'on veut, vérifie

$$P_f(A) = P_\tau(A \circ \pi).$$

Je ne vais plus continuer indéfiniment, mais je vais essayer de mentionner quelques faits importants dans ce cadre-là. En premier lieu, je note qu'il existe une mesure particulièrement intéressante. Intuitivement, c'est ceci. Si l'on prend un axiome A ou un difféomorphisme d'Anosov, il n'y a pas en général de mesure absolument continue par rapport à la mesure de Lebesgue qui soit invariante par ce difféomorphisme. Alors, une idée que l'on peut avoir est de prendre la mesure de Lebesgue, d'appliquer les itérés du difféomorphisme à celle-ci et de voir ce qui se passe à la limite. La limite, dans ce cas-ci, existe

$$\rho = \lim \text{vague}_{n\to\infty} f^n \varphi,$$

où $\varphi \in L^1$ (Lebesgue sur un voisinage de Λ) indique une mesure de probabilité. Cette limite ρ s'appelle mesure SRB (Sinai, Ruelle, Bowen) et possède des propriétés intéressantes. Elle a des mesures conditionnelles sur les variétés instables (notées ρ^u) qui sont absolument continues par rapport à la mesure de Lebesgue sur les variétés instables. Plus formellement, on a

$$\rho^u \in L^1(\text{Lebesgue sur } \mathcal{V}^u)$$

et la densité de ρ^u s'exprime en termes du jacobien instable J^u. C'est encore une mesure de Gibbs par rapport à une certaine fonction continue höldérienne. En fait, ρ est l'unique mesure d'équilibre pour

$$A = -\log J^u.$$

Et, il est hautement non évident, mais vrai, que les fonctions de corrélation

$$\int B \cdot C \circ f^k \, d\rho - \int B \, d\rho \int C \, d\rho$$

décroissent exponentiellement si B, C sont höldériennes sur Λ, parce que c'est vrai pour une mesure de Gibbs en général.

Commentaires et références

Voici quelques références. On commence par l'article d'Adler et Weiss [1] qui est ancien, de 1967. Citons aussi les travaux de Sinai, qui sont de l'année d'après : [17, 18].

Sinai a été invité en 1970 au Congrès International des Mathématiciens à Nice, cependant la politique de l'Union Soviétique de l'époque était telle qu'il n'a pas pu venir ; mais, enfin, il a un article dans le compte rendu. Une version plus complète n'a été publiée qu'en 1972 : [19]. Finalement, je mentionne Bowen [6] qui vient un peu plus tard, et puis les articles sur la mesure SRB : [14], [8].

Opérateurs de transfert

Outils pour le formalisme thermodynamique

J'ai couvert plusieurs aspects de la théorie, ce que je n'ai pas couvert (dont je voudrais simplement dire deux mots) est comment on peut espérer démontrer des choses pareilles. Ce que j'ai plus ou moins expliqué est comment on pouvait passer de la théorie ergodique sur un sous-shift de type fini à la théorie ergodique pour les systèmes différentiables hyperboliques. Cela utilise le concept de partition de Markov, et là je ne vous ai pas expliqué comment on démontre l'existence d'une partition de Markov.

Mais il y a d'autres données que je ne vous ai pas expliquées. Par exemple, comment on obtient des propriétés comme le fait que la pression est une fonction différentiable sur l'espace des fonctions continues höldériennes ; autrement dit, quelle est la stratégie adoptée pour démontrer que la correspondance

$$\mathrm{A} \mapsto \mathrm{P}(\mathrm{A})$$

est analytique réelle sur $C^{\alpha}(\Lambda)$? En outre, comment peut-on trouver des propriétés qui assurent la décroissance exponentielle de fonctions de corrélation :

$$\left| \int \mathrm{B} \cdot \mathrm{C} \circ \tau^{k} \, d\rho_{\mathrm{A}} - \int \mathrm{B} \, d\rho_{\mathrm{A}} \int \mathrm{C} \, d\rho_{\mathrm{A}} \right| \leq \text{constante } e^{-a|k|} \text{ ?}$$

Revenons, alors, à notre réseau à une dimension. Au lieu de considérer un réseau infini (donc tous les entiers), je vais considérer seulement les entiers positifs (strictement).

Je peux définir sur les entiers positifs une mesure de Gibbs. Sans aucune surprise, c'est une mesure de probabilité qui a cette propriété : si l'on regarde la mesure conditionnelle par rapport à ce qui se passe à l'extérieur, alors cette mesure conditionnelle est donnée simplement par un exposant moins β multiplié par une certaine énergie et cela,

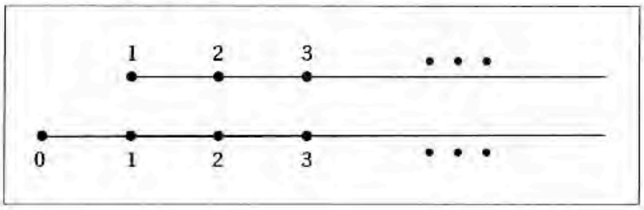

Fig. 10.

bien sûr, normalisé. Donc, j'ai donné une définition d'un état de Gibbs qui ne fait pas intervenir l'invariance par translation et qui s'applique aussi à un système semi-infini comme celui-là.

Si vous croyez qu'il va y avoir un état de Gibbs unique dans cette situation, alors il est intéressant de voir comment on peut comparer l'état de Gibbs pour ce système semi-infini à l'état de Gibbs pour le système auquel on a rajouté un point. Dénotons l'état de Gibbs initial par ν. L'état de Gibbs quand on a rajouté un point, disons 0, pourra s'obtenir à partir de ν simplement en translatant, c'est-à-dire que la mesure correspondante sur $\mathbb{N}$ sera $\tau\nu$. Néanmoins, il y a une autre approche que l'on peut adopter. Considérons l'état de Gibbs initial et multiplions-le par l'exponentielle de l'opposé de l'énergie d'interaction du point 0 avec tous les points qui sont à droite :

$$e^{-W_\Phi(\sigma_0,\underline{\sigma})}\nu(d\underline{\sigma}) \;/\; \text{normalisation.}$$

Cela devrait donner la même chose, à la normalisation près. Donc, on a une équation avec un facteur de normalisation qui n'est pas donné : ce sera une équation aux valeurs propres.

Cependant, il se trouve que travailler sur les mesures signifie travailler sur un espace de Banach qui n'est pas très sympathique. Il y a avantage à travailler sur les fonctions continues et, en réalité, à regarder les fonctions continues höldériennes. Donc, on va considérer l'opérateur défini par

$$(\mathscr{L}\Phi)(\underline{\sigma}) = \sum_{\sigma_0} e^{-W_\Phi(\sigma_0,\underline{\sigma})}\Phi(\sigma_0,\underline{\sigma}).$$

La motivation de sa définition montre que cet opérateur, appelé opérateur de transfert, est en fait le dual d'un opérateur.

Si l'on regarde, alors, les choses de près, on se rend compte que cet opérateur de transfert qui agit sur les fonctions höldériennes sur un demi-réseau est quasi-compact (c'est-à-dire que ses itérés s'approchent d'opérateurs compacts) et a une valeur propre isolée. Et cette valeur propre maximale, par nécessité, va dépendre de manière analytique de tous les paramètres qui interviennent dans le problème. En outre, le logarithme de cette valeur propre est, en fait, la pression. Eh bien, il va se trouver que la pression sera une fonction analytique de tous les paramètres que l'on peut vouloir. On peut de cette manière-là aussi obtenir la décroissance exponentielle de corrélations. Fondamentalement, il faut remarquer que le fait que la valeur propre soit isolée détermine un trou (*gap*) spectral.

Bon, là je crois que je vais m'arrêter. Merci.

Références

Pour ces outils, voir les références [13], [12], [4], [16], [10].

Questions

M. Mendès France. — Pour les systèmes de spins sur un réseau, quand vous dites que l'ensemble Λ tend vers l'infini, c'est dans toutes ses dimensions quand même, non ?

D. Ruelle. — Oui. Alors, là il faudrait préciser. Mais si l'on prend un parallélépipède dont toutes les dimensions tendent vers l'infini, ça va. Si l'on prend une sphère ou un cube, tout cela est bon. Si l'on commence à prendre des choses très, très filamenteuses, cela risque de ne pas marcher.

P. Thieullen. — Qu'est-ce que l'on peut dire de la mécanique statistique hors de l'équilibre ?

D. Ruelle. — La mécanique statistique hors de l'équilibre... Il se trouve que la mécanique statistique de l'équilibre est un sujet naturel, parce que l'on commence par la donnée d'une mesure de probabilité : la mesure de Gibbs. La mécanique statistique de l'équilibre est l'étude de cette mesure de Gibbs dans les situations plus ou moins compliquées. Hors de l'équilibre, il ne s'agit plus de faire la dynamique qui correspond aux translations du réseau, mais une dynamique différente. Et cela pose, alors, des tas de problèmes, lesquels sont étudiés pour l'instant ; c'est un sujet d'étude actuel. Cependant, il faut bien dire que cela n'a pas le même caractère évident et naturel que la mécanique statistique de l'équilibre où, à partir du moment où l'on s'est dit « on veut mettre sur une base rigoureuse ce qu'ont fait Boltzmann et Gibbs », on a trouvé des bonnes mathématiques. C'était un domaine à mathématiser en priorité, tandis que le non-équilibre est beaucoup plus difficile et, pour l'instant, cela paraît moins naturel.

Bibliographie

[1] R. Adler et B. Weiss : *Entropy, a complete metric invariant for automorphisms of the torus*, Proceedings of the National Academy of Sciences of the United States of America **57** (1967), p. 1573-1576.

[2] D. V. Anosov : *Geodesic flows on compact Riemann manifolds of negative curvature*, Proceedings of the Steklov Institute of Mathematics **90** (1967), p. 1-209.

[3] V. I. Arnold et A. Avez : *Problèmes ergodiques de la mécanique classique*, Gauthier-Villars, 1967.

[4] V. Baladi : *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics **16**, World Scientific, Singapore, 2000.

[5] F. Bonetto, G. Gallavotti et G. Gentile : *Aspects of ergodic, qualitative and statistical theory of motion*, Springer, Berlin, 2004.

[6] R. Bowen : *Markov partitions for Axiom A diffeomorphisms*, American Journal of Mathematics **92** (1970), p. 725-747.

[7] R. Bowen : *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics **470**, Springer, Berlin, 1975.

[8] R. Bowen et D. Ruelle : *The ergodic theory of Axiom A flows*, Inventiones Mathematicae **29** (1975), p. 181-202.

[9] R. L. Dobrushin : *Gibbsian random fields for lattice systems with pairwise interactions*, Funktsional'nyi analiz i ego prilozheniya **2** (1968), n^{o} 4, p. 31-43. [Traduction anglaise : Functional Analysis and Its Applications **2** (1968), p. 292-301.]

[10] S. Gouëzel and C. Liverani : *Banach spaces adapted to Anosov systems*, Ergodic Theory and Dynamical Systems **26** (2006), p. 189-217.

[11] O. E. Lanford et D. Ruelle : *Observables at infinity and states with short range correlations in statistical mechanics*, Communications in Mathematical Physics **13** (1969), p. 194-215.

[12] W. Parry et M. Pollicott : *Zeta fonctions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187-188**, Société Mathématique de France, Paris, 1990.

[13] D. Ruelle : *Statistical mechanics of a one-dimensional lattice gas*, Communications in Mathematical Physics **9** (1968), p. 267-278.

[14] D. Ruelle : *A measure associated with Axiom A attractors*, American Journal of Mathematics **98** (1976), p. 619-654.

[15] D. Ruelle : *Thermodynamic formalism*, Encyclopedia of Mathematics and Its Applications **5**, Addison-Wesley, Reading, 1978.

[16] D. Ruelle : *Dynamical zeta functions and transfer operators*, Notices of the American Mathematical Society **49** (2002), p. 887-895.

[17] Ya. G. Sinai : *Markov partitions and C-diffeomorphisms*, Funktsional'nyi analiz i ego prilozheniya **2** (1968), n^{o} 1, p. 64-89. [Traduction anglaise : Functional Analysis and Its Applications **2** (1968), p. 61-82.]

[18] Ya. G. Sinai : *Construction of Markov partitions*, Funktsional'nyi analiz i ego prilozheniya **2** (1968), n^{o} 3, p. 70-80. [Traduction anglaise : Functional Analysis and Its Applications **2** (1968), p. 245-253.]

[19] Ya. G. Sinai : *Gibbsian measures in ergodic theory*, Uspekhi Matematicheskikh Nauk **27** (1972), n^{o} 4, p. 21-64. [Traduction anglaise : Russian Mathematical Surveys **27** (1972), n^{o} 4, p. 21-69.]

[20] S. Smale : *Differentiable dynamical systems*, Bulletin of the American Mathematical Society **73** (1967), p. 747-817.