Zeitschrift: Commentarii Mathematici Helvetici
Herausgeber: Schweizerische Mathematische Gesellschaft
Band: 17 (1944-1945)

Artikel: Sur la théorie ergodique.
Autor: Riesz, Frédéric
DOI: https://doi.org/10.5169/seals-16339

Sur la théorie ergodique

Par Frédéric Riesz, Szeged [1])

I.

La théorie ergodique découle, comme vous le savez certainement, d'une hypothèse hardie et ingénieuse dont on s'est servi pendant une longue période sans la justifier, et sous des formes variées, dans la théorie cinétique de la matière et plus généralement en mécanique statistique. Le premier résultat définitif, de valeur pour le pur mathématicien, était peut-être le célèbre théorème de recurrence de Poincaré datant de l'année 1890; et même les idées intuitives par lesquelles l'illustre géomètre se laissait guider, devaient attendre encore une douzaine d'années avant d'être légitimées par la théorie de Lebesgue. Mais la renaissance de la théorie ne date que de 1931, année où, avec les théorèmes de Neumann et de Birkhoff, elle entre définitivement dans le cadre de l'analyse fonctionnelle. C'est de ces théorèmes et de quelques généralisations que je vous parlerai.

Pour nous orienter, rappelons d'abord le modèle dont se sert Poincaré; à part la dimension, c'est-à-dire le nombre des variables qui est très élevé dans les problèmes corpusculaires, ce modèle représente, on peut le dire, le cas général. Considérons un liquide enfermé dans un vase de forme invariable et qu'il remplit complètement. Supposons notre liquide en mouvement stationnaire. Alors les molécules qui se trouvent dans une certaine région à l'instant t_0, rempliront, à chaque instant t, une région de même volume; c'est-à-dire que le passage d'un instant à l'autre n'est qu'une transformation ponctuelle qui conserve le volume ou plus généralement la mesure. Il s'agit donc d'une famille de transformations dépendant du temps t et formant un groupe ou (si nous ne sommes pas assez historiens pour nous intéresser au passé) formant un semi-groupe tel que, en désignant par P_t ce que devient le point P après un temps t, on a, avec une écriture évidente, $(P_s)_t = P_{s+t}$; ou aussi, en désignant par T_t la transformation ponctuelle qui correspond à un intervalle de temps égal à t, on a $T_s T_t = T_{s+t}$. Alors le théorème de Poincaré affirme que pour presque tout point P, c'est-à-dire sauf peut-être pour les points d'un ensemble de mesure nulle, les trajectoires de $P_t = T_t(P)$ retournent une infinité de fois dans chaque voisinage du point P, et que ce fait sub-

[1]) L'Université de Genève avait invité Monsieur Frédéric Riesz à faire deux conférences au printemps 1944. L'auteur n'ayant pu s'y rendre en personne, nous sommes heureux de pouvoir en publier le texte. *La rédaction.*

siste lorsque, au lieu de t continu, on ne regarde que la succession des multiples $t_n = nt_1$ d'une unité de temps t_1, choisie à volonté. Or, la théorie moderne de l'intégration suggère de remplacer le vase et les voisinages par des ensembles mesurables et il ne serait pas difficile d'adapter l'argument de Poincaré à ces nouvelles conditions.

Mais il y a encore un problème plus profond qui se pose. En effet, le théorème ne nous dit rien concernant la fréquence des passages de notre trajectoire dans le voisinage en question pendant que t croît indéfiniment ou, quand il ne s'agit que des instants discrets $t_n = nt_1$, de la valeur moyenne du nombre des cas de passage. Rappelons l'exemple le plus simple, celui où une circonférence tourne avec une vitesse angulaire constante autour de son centre. Alors, ce que dit le théorème de Poincaré dans ce cas particulier est évident pour t continu et il en est de même quant au problème de la fréquence. D'autre part, dans le cas discontinu, le théorème de Poincaré n'est qu'une conséquence immédiate du fameux „Schachtelprinzip" de Dirichlet, tandis que le problème de la fréquence se résout par un théorème beaucoup plus caché, concernant la répartition uniforme des points $P_n = P_{t_n}$ dans tous les cas où t_1 est une fraction incommensurable du temps de parcours de la circonférence entière. Il s'agit du théorème de répartition uniforme, démontré autour de 1910 par plusieurs auteurs et qui nous dit, que pour un arc quelconque de la circonférence, la moyenne ν/n où ν désigne le nombre de ceux des n premiers P_k qui sont situés sur l'arc envisagé, converge vers le rapport des longueurs de cet arc et de la circonférence entière. M. Weyl qui s'est occupé à plusieurs reprises, du théorème et aussi de quelques généralisations et leurs applications, entre autre dans une conférence faite, il y a justement 30 années, à la *Société mathématique suisse* et imprimée dans *L'Enseignement mathématique*, en a donné une démonstration dans les Göttinger Nachrichten de 1914 et dans les Mathematische Annalen, t. 77, qui nous intéresse particulièrement au point de vue de notre sujet actuel. En voici l'idée principale.

On voit immédiatement que le théorème peut se mettre sous la forme suivante.

Soit $f(x)$ la fonction escalier de période 2π qui est égale à 1 et à 0 respectivement dans les intervalles $(0, a)$ et $(a, 2\pi)$, et soit ξ un nombre irrationnel quelconque. Alors

$$\lim \frac{1}{n} \sum_{k=1}^{n} f(x + k\,\xi) = \int_0^{2\pi} f(t) \quad dt \ . \tag{1}$$

Ce n'est pas par manque d'attention que j'ai écrit l'intégrale au lieu de mettre simplement sa valeur a. En effet, la relation (1) est valable pour une classe étendue de fonctions, entre autres pour toutes les fonctions de période 2π qui sont intégrables au sens de Riemann sur les intervalles finis. Cela vient de ce qu'elle se vérifie, par un calcul évident, pour les fonctions $e(mx)$ où $e(x) = e^{2\pi i x}$ et m est un entier, donc aussi pour les polynomes trigonométriques de $2\pi x$ et enfin pour les fonctions dont nous venons de parler, ce qu'on voit en les intercalant entre deux polynomes trigonométriques dont les intégrales sur $(0, 2\pi)$ diffèrent aussi peu qu'on voudra.

En réalité, comme l'a observé M. Khintchine, la relation (1) reste encore valable, cette fois pour presque tous les x, sous l'hypothèse plus large de l'intégrabilité au sens de Lebesgue, même pour les fonctions non bornées. Mais, dans ce cas général, malgré l'énoncé extrêmement simple, la seule voie par laquelle je peux l'approcher, et c'est ce que fait en réalité aussi M. Khintchine, c'est de la considérer comme corollaire du plus achevé des théorèmes ergodiques, savoir de celui de Birkhoff.

Avant d'en parler, rappelons encore brièvement un second exemple, choisi dans la théorie des probabilités dénombrables. Il s'agit du fait, découvert par M. Borel, que pour presque tout nombre réel, quand on l'écrit en forme de fraction décimale infinie, les divers chiffres se répartissent également à la limite, chaque chiffre admettant la fréquence moyenne 1/10, et la moyenne arithmétique des chiffres successifs convergeant vers $4\frac{1}{2}$. Le premier de ces faits, dont le second n'est qu'un corollaire, est impliqué par la formule

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \to \int_0^1 f(x)\, dx \,, \tag{2}$$

analogue à (1), où l'on définit la transformation T, pour l'intervalle $(0, 1)$, par $x = 0, a_1 a_2 a_3 \ldots$, $Tx = 0, a_2 a_3 a_4 \ldots$, c'est-à-dire en supprimant le premier chiffre décimal de la fraction x et où T^k désigne les itérés de T. Pour fonction $f(x)$, on n'aura qu'à choisir successivement les fonctions caractéristiques des parties dixièmes de l'intervalle $(0, 1)$, c'est-à-dire les fonctions égales à 1 sur l'un de ces dixièmes et s'annulant ailleurs. En réalité, la formule (2) est valable presque partout non seulement pour ces fonctions, mais pour toute fonction intégrable au sens de Lebesgue. C'est M. Raïkov qui a établi ce résultat dans un travail en langue russe où, à ce que je peux juger en regardant les formules, il ne fait en réalité qu'adapter à ce cas particulier une des démonstrations du théorème de Birkhoff.

La raison principale pour laquelle je viens de rappeler ce second exemple, c'est que la transformation T qui y intervient, n'est pas biunivoque comme dans l'autre exemple et cela non seulement à cause de la double écriture des fractions décimales finies, mais puisque son inverse admet dix déterminations différentes.

Je pourrais encore citer, à titre d'analogie, un des premiers résultats de la théorie des fonctions presque périodiques, c'est que toute fonction de ce genre $f(x)$ admet une valeur moyenne, ou d'une façon précise, que les moyennes

$$\frac{1}{T}\int_0^T f(x+t)\,dt$$

tendent, pour T infini, vers une valeur constante déterminée. Mais il faut me dépêcher pour arriver au théorème général qui est au bout de notre ordre d'idées et que M. G. D. Birkhoff a établi en 1931, à la suite des recherches de MM. Koopman et de Neumann, recherches auxquelles je reviendrai dans ma seconde conférence.

Voici le théorème.

Soit donné un ensemble mesurable Ω, de mesure finie ou infinie, la mesure et l'intégrale correspondante étant définies d'après Lebesgue, ou plus généralement, par rapport à une distribution de masses positives.

Cela étant, désignons par T une transformation ponctuelle univoque (mais non nécessairement biunivoque) de Ω en soi-même et supposons que T conserve la mesure au sens que, E étant un ensemble mesurable, TE son transformé et E' l'ensemble des points P dont les images appartiennent à TE, les ensembles E' et TE admettent la même mesure. Alors, en partant d'une fonction intégrable $f_1(P)$ et en posant $f_k(P) = f_1(T^{k-1}P)$, la moyenne arithmétique des fonctions $f_1, f_2, \ldots, f_n$ converge presque partout, pour $n \to \infty$, vers une fonction intégrable $\varphi(P)$, invariante (presque partout) par rapport à T.

J'ai énoncé le théorème sous une forme un peu plus générale que l'a fait son auteur, en particulier en ne me bornant pas à des transformations biunivoques. D'ailleurs je pense qu'il ne serait pas difficile d'adapter le raisonnement de M. Birkhoff à cette hypothèse plus générale qui m'était suggérée par le théorème de M. Borel.

M. Birkhoff base la démonstration de son théorème, en substance, sur le *lemme* suivant, qu'il applique sous une forme un peu différente et auquel il arrive par un démembrement très subtil des ensembles qui interviennent. *Soit, pour une fonction donnée $f_1(P)$, E l'ensemble des points P*

pour lesquels la limite supérieure des moyennes arithmétiques figurant dans le théorême est positive. Alors

$$\int_E f_1(P) > 0 \ . \tag{3}$$

Il y a quelque temps, MM. Yosida et Kakutani ont réussi à remplacer, dans ce lemme, la limite supérieure par la borne supérieure; d'une façon précise, l'ensemble E devra être caractérisé par le fait que, en ses points, l'une au moins des sommes $f_1 + f_2 + \cdots + f_n$ admet une valeur positive. C'est cette proposition (ou plutôt une qui lui est équivalente) qu'ils appellent le „maximal ergodic theorem".

Ce qui est nouveau dans la présente démonstration, c'est que le lemme en question se déduit, par un artifice simple, du lemme tout élémentaire que voici. *Etant donnés n quantités réelles* $a_1, a_2, \ldots, a_n$ *et un entier* $m < n$, *considérons toutes les sommes* $a_k + a_{k+1} + \cdots a_l$ *de valeur positive, formées d'éléments successifs dont le nombre des termes ne dépasse pas l'entier m. Alors les* a_k *figurant comme premiers termes dans l'une au moins de ces expressions, ont leur somme positive.*

Ce dernier lemme se démontre en quelques lignes. Pour faciliter le langage, appelons termes favorables les a_k en question et soit a_{k_1} le premier d'entre eux. Soit $a_{k_1} + \cdots + a_{l_1}$ la plus courte parmi les sommes issues de ce terme et de valeur > 0; alors les termes de cette somme sont tous des termes favorables. En effet, s'il n'en était pas ainsi pour un de ces termes, soit $a_{k'}$, alors on aurait $a_{k'} + \cdots + a_{l_1} \leqq 0$ et par conséquent $a_{k_1} + \cdots + a_{k'-1} > 0$, contrairement à la convention faite. Pour continuer, soit a_{k_2} le premier terme favorable parmi ceux qui restent et $a_{k_2} + \cdots + a_{l_2}$ la somme la plus courte commençant par a_{k_2}. En continuant de la même façon, nos sommes se composeront à la fin précisément de tous les termes favorables. Notre lemme est donc démontré. Voici comment on passe alors au „maximal ergodic theorem".

Désignons par $E^{(m)}$ l'ensemble des points P pour lesquels une au moins des sommes

$$\sum_1^l f_k(P) \qquad (l \leqq m)$$

est plus grande que 0. Les ensembles $E^{(m)}$ épuisent successivement l'ensemble E qui figure au théorème. Donc, au lieu de (3), nous n'avons qu'à vérifier l'inégalité analogue

$$\int_{E^{(m)}} f_1(P) \geqq 0 \ . \tag{4}$$

A cet effet, envisageons, pour chaque P, la somme des termes favorables (au sens qui précède) de la suite finie

$$f_1(P), \ldots, f_{n+m}(P)$$

(où nous avons remplacé n par $n + m$). Cette somme étant une fonction positive et sommable de P, on obtient par intégration:

$$\sum_{k=1}^{n+m} \int_{E_k} f_k(P) \geqq 0 \quad , \tag{5}$$

E_k désignant l'ensemble des points P pour lesquels $f_k(P)$ est un terme favorable.

D'autre part, pour $k \leqq n$, on a évidemment $TE_k = E_{k-1}$, $E_k = T^{-1}E_{k-1}$. Comme de plus T conserve la mesure et donc aussi l'intégrale, les n premières intégrales dans (5) admettent des valeurs égales, et comme $E_1 = E^{(m)}$, leur valeur commune est égale à l'intégrale qui figure dans (4). Il n'en est pas nécessairement ainsi pour les m intégrales qui restent, cepandant la valeur d'aucune d'elles ne dépasse celle de l'intégrale de $|f_1(P)|$ sur Ω entier. Il s'ensuit que

$$n \int_{E^{(m)}} f_1(P) + m \int_{\Omega} |f_1(P)| \geqq 0$$

et enfin, en divisant par n et en faisant le passage à l'infini, on obtient l'inégalité (4) qu'il fallait démontrer.

La démonstration du théorème de Birkhoff s'achève maintenant comme il suit.

Soit $E_{\alpha\beta}$ où $\alpha > \beta$, l'ensemble des points P pour lesquels on a à la fois

$$\overline{\lim} \frac{1}{n} \sum_1^n f_k(P) > \alpha \tag{6}$$

et

$$\underline{\lim} \frac{1}{n} \sum_1^n f_k(P) < \beta \quad . \tag{7}$$

Supposons pour l'instant que Ω soit de mesure finie, alors il en sera de même de l'ensemble $E_{\alpha\beta}$ qui est évidemment mesurable. Or ce dernier est invariant par rapport à T; en effet, en écrivant dans les formules (6) et (7) TP à la place de P et $n - 1$ au lieu de n, les deux limites d'indétermination ne seront pas changées.

L'ensemble $E_{\alpha\beta}$ étant invariant, il pourra jouer le rôle de Ω dans le lemme que je viens de prouver et en appliquant celui-ci à la fonction $f_1(P) - \alpha$ au lieu de $f_1(P)$, l'ensemble E du théorème coïncidera aussi avec $E_{\alpha\beta}$. Par conséquent

$$\int_{E_{\alpha\beta}} [f_1(P) - \alpha] \geqq 0 \quad .$$

De même, en remplaçant $f_1(P)$ par $\beta - f_1(P)$, on obtient

$$\int_{E_{\alpha\beta}} [\beta - f_1(P)] \geqq 0 \quad .$$

et par addition

$$\int_{E_{\alpha\beta}} (\beta - \alpha) \geqq 0 \quad .$$

Mais cela contredit à l'hypothèse que $\alpha > \beta$, excepté si $E_{\alpha\beta}$ est de mesure nulle.

Faisons maintenant (α, β) parcourir les couples des nombres rationnels, alors $E^* = \Sigma E_{\alpha\beta}$ étant la réunion d'une infinité dénombrable d'ensembles de mesure nulle, il sera lui-même de mesure nulle. C'est-à-dire que la moyenne arithmétique des fonctions $f_k(P)$ converge presque partout, savoir partout à l'exception de E^*, vers une limite déterminée $\varphi(P)$, finie ou infinie. Or, comme

$$\int_{\Omega} \left| \frac{1}{n} \sum_1^n f_k(P) \right| \leq \frac{1}{n} \int_{\Omega} \sum_1^n \left| f_k(p) \right| = \int_{\Omega} \left| f_1(P) \right| \ ,$$

la fonction $\varphi(P)$ sera intégrable grâce au théorème de Fatou; elle sera donc finie presque partout.

Quand Ω est de mesure infinie, nous n'avons qu'à montrer, avant d'intégrer sur l'ensemble $E_{\alpha\beta}$, que ce dernier est toujours de mesure finie. Nous avons besoin de ce fait pour pouvoir intégrer $f_1(P) - \alpha$ sur $E_{\alpha\beta}$. Supposons que $\alpha > 0$; en cas contraire on aurait $\beta < 0$ et on pourrait opérer avec $-f_1(P) - (-\beta)$ au lieu de $f_1(P) - \alpha$.

Envisageons un sous-ensemble E' de $E_{\alpha\beta}$, de mesure finie, d'ailleurs quelconque. Soit $e'(P)$ sa fonction caractéristique. Appliquons le „maximal ergodic theorem" à la fonction $g_1(P) = f_1(P) - \alpha e'(P)$ au lieu de $f_1(P)$; alors, pour l'ensemble E qui correspond à $g_1(P)$, on aura

$$\int_E g_1(P) \geqq 0 \ ,$$

et comme $E' \subset E_{\alpha\beta} \subset E$, il vient

$$\int_E f_1(P) \geqq \alpha \int_E e'(P) = \alpha \text{ mes } E' \quad .$$

Donc, à plus forte raison,

$$\alpha \text{ mes } E' \leqq \int_{\Omega} | f_1(P) | \quad .$$

Cela nous dit que la mesure d'aucun sous-ensemble de mesure finie de $E_{\alpha\beta}$ ne dépasse la borne

$$\frac{1}{\alpha}\int_{\Omega} |f_1(P)| ,$$

et par conséquent $E_{\alpha\beta}$ lui-même doit être de mesure finie.

L'invariance de la fonction $\varphi(P)$, $\varphi(TP) = \varphi(P)$, en découle immédiatement en remplaçant, comme nous l'avons déjà fait, P par TP, c'est-à-dire $f_k(P)$ par $f_{k+1}(P)$ et de plus n par $n - 1$ dans les moyennes arithmétiques considérées, ce qui ne change pas la limite de ces moyennes.

Ajoutons enfin que dans le cas où Ω est de mesure finie, il vient encore par intégration terme à terme (ce qui est permis dans ce cas grâce à l'intégrabilité uniforme des termes)

$$\int_{\Omega} \varphi(P) = \int_{\Omega} f_1(P) .$$

Disons encore quelques mots sur la variante „intégrale" du théorème, celle où l'on introduit un paramètre continu t et où les itérés T^k sont remplacés par une famille T_t de tranformations avec $T_s T_t = T_{s+t}$. Le théorème correspondant que je n'énonce pas en détail pour ne pas vous fatiguer, se réduit aisément au cas discontinu et cela par un artifice dû à MM. E. Hopf et Khintchine. Mais peut-être y a-t-il quelque intérêt d'observer que, au lieu de cet artifice, les détails de la démonstration que je viens de vous présenter, s'adaptent aussi immédiatement au cas continu. Qu'il nous suffise aujourd'hui de formuler le lemme qui correspond, pour t continu, à notre lemme élémentaire.

Etant donnée, sur l'intervalle (a, b), une fonction intégrable $g(t)$, de plus une longueur $d \leqq b - a$, envisageons l'ensemble e des valeurs t_0 telles que l'intégrale de $g(t)$ prise de t_0 jusqu'à $t_0 + h$, soit positive pour une au moins des valeurs $h < d$. Alors l'intégrale de $g(t)$ sur l'ensemble e est aussi positive ou 0.

Qu'il me soit permis d'attirer votre attention sur la relation intime entre ce lemme et un autre dont je me suis servi, précisément au temps de la découverte de M. Birkhoff, pour en déduire l'existence de la dérivée des fonctions monotones ainsi qu'une inégalité importante de MM. Hardy et Littlewood. En effet, soit $G(t)$ l'intégrale de la fonction $g(t)$ qui figure dans notre lemme ci-dessus; alors celui-ci et l'autre ne sont que des corollaires du théorème suivant.

Soit $G(t)$ une fonction continue, définie dans l'intervalle (a, b) et envisageons, pour d positif donné, l'ensemble E des points t intérieurs à cet inter-

valle et tels qu'il existe un t' avec $t < t' < t + d$, de sorte que $G(t) < G(t')$. Alors l'ensemble E est ouvert et pour les intervalles (a_k, b_k) dont il se compose, on a $G(a_k) \leqq G(b_k)$.

Les diverses démonstrations de mon lemme d'autrefois dont je viens de parler et vous en trouverez une dans ma conférence faite au *Congrès de Zurich* en 1932, s'adaptent aussitôt au lemme actuel.

Enfin, en ce qui concerne les généralisations du théorème de Birkhoff, on n'en connaît que très peu et je ne cite aujourd'hui que celle de M. E. Hopf. Je me contente de l'énoncer pour le cas discontinu d'une succession $\{T^k\}$ et cela même sous une forme légèrement modifiée. Envisageons, avec des notations évidentes, l'expression

$$\frac{\sum_1^n f_k(P)}{\sum_1^n g_k(P)} \tag{8}$$

(dans le cas de Birkhoff $g_k \equiv 1$) où la fonction $g_1(P)$ et avec elle les $g_k(P)$ sont encore supposées essentiellement positives; quant à leur intégrabilité, on ne les suppose intégrables que sur les sous-ensembles de Ω qui sont de mesure finie. Soit Ω' l'ensemble des points pour lesquels la série $\Sigma g_k(P)$ diverge. Alors le théorème affirme que l'expression (8) converge presque partout sur Ω' vers une fonction $\Psi(P)$, invariante par rapport à T et que de plus, lorsque $g_1(P)$ est intégrable sur Ω', on a

$$\int_{\Omega'} g_1(P)\ \Psi(P) = \int_{\Omega'} f_1(P) \quad .$$

Le théorème s'établit par la même méthode que celui de Birkhoff et c'est seulement la formule finale qui exige d'être vérifiée séparément. Cela se fait par un calcul élémentaire de la théorie de Lebesgue et qui consiste à décomposer Ω' en des ensembles E_n sur lesquels

$$(n - 1)\,\delta \leqq \Psi(P) < n\,\delta\ ,$$

d'observer que les E_n sont invariants par rapport à T et d'en conclure, par le „maximal ergodic theorem", que

$$(n-1)\,\delta \int_{E_n} g_1(P) \leqq \int_{E_n} f_1(P) \leqq n\,\delta \int_{E_n} g_1(P) \quad .$$

Comme l'intégrale

$$\int_{E_n} g_1(P)\ \Psi(P)$$

est comprise évidemment entre les mêmes bornes, elle ne diffère de celle de $f_1(P)$ que par

$$\delta \int_{E_n} g_1(P)$$

au plus; notre formule finale s'ensuit, après addition par rapport à n, pour δ infiniment petit.

Observons d'ailleurs, que sous des conditions faciles à préciser et en particulier, lorsque Ω est de mesure finie, le théorème se réduit à un corollaire de celui de Birkhoff.

Pour terminer cette première conférence, disons encore quelques mots de ce qu'on appelle la *transitivité métrique*. Dans les applications, il est parfois d'une grande importance de savoir si notre fonction limite est constante ou d'une façon générale, de prouver que la transformation qui change $f(P)$ en $f(TP)$, n'admet pas d'autre fonction invariante que des constantes, bien entendu des constantes presque partout. On voit aussitôt que pour qu'il en soit ainsi, il faut et il suffit que l'ensemble Ω ne puisse être décomposé en deux ensembles invariants presque partout par rapport à T sans que l'un de ces ensembles soit de mesure nulle. C'est cette propriété qu'on appelle la transitivité métrique de la transformation T. En général, il n'est pas facile de la constater et dans la plupart des applications, c'est une tâche qui attend encore d'être accomplie.

II.

Dans ma première conférence, je vous ai présenté le théorème ergodique de M. G. D. Birkhoff; aujourd'hui je vous parlerai de celui de M. J. de Neumann qui l'a devancé de quelques semaines. Dans sa forme originale, ce théorème envisage, tout comme l'autre, les moyennes arithmétiques $\varphi_n(P)$ des fonctions $f_k(P)$ qui viennent de $f_1(P)$ en itérant la transformation ponctuelle T et en posant $f_k(P) = f_1(T^{k-1}P)$. Le théorème envisage encore et tel n'était pas le cas chez M. Birkhoff, les moyennes plus générales

$$\varphi_{m,n}(P) = \frac{1}{n-m} \sum_{m+1}^{n} f_k(P) \quad . \tag{9}$$

Cette fois-ci, la fonction $f_1(P)$ est supposée appartenir à la classe L^2, c'est-à-dire être intégrable ainsi que son carré; ou plutôt, comme on admet aussi des fonctions à valeurs complexes, que le carré de son module. Alors le théorème en question affirme que, pour $n - m \to \infty$, les fonctions φ_n et $\varphi_{m,n}$ convergent „en moyenne" vers une limite $\varphi(P)$ déterminée presque partout; en formules

$$\int_{\Omega} |\varphi(P) - \varphi_n(P)|^2 \to 0 \ , \quad \int_{\Omega} |\varphi(P) - \varphi_{m,n}(P)|^2 \to 0 \ .$$

Sans doute, le théorème, sous sa présente forme, paraît dire beaucoup moins que celui de Birkhoff; en effet il ne dit rien quant à la convergence effective au sens ordinaire. Cependant, à part ce que le théorème suffit pour les applications en Mécanique statistique, il présente aussi, comme nous allons le voir, beaucoup d'intérêt au point de vue purement mathématique. Tout d'abord, il ne faut pas oublier qu'il embrasse aussi le cas général des moyennes du type (9), c'est-à-dire qu'il implique une sorte de convergence uniforme par rapport aux indices. Bien plus, déjà la première démonstration par Neumann ainsi que celle de M. Hopf qui la suivit immédiatement, ne se servent que d'une seule conséquence de la conservation de la mesure, conséquence observée indépendamment par MM. Koopman et Carleman. C'est que le passage de $f(P)$ à $f(TP)$ conserve l'intégrale du carré ou se qui revient au même, la norme $||f|| = [\int |f|^2]^{\frac{1}{2}}$. Avec ce fait, le rôle dont jouit la transformation ponctuelle est épuisé et les fonctions ne sont que des éléments d'un espace vectoriel du type euclidien ou unitaire à un nombre fini ou une infinité de dimensions, disons brièvement d'un espace de Hilbert réel ou complexe; le passage de $f(P)$ à $f(TP)$ ne sera qu'une transformation linéaire qui conserve la norme $||f||$ ou ce qui revient au même, la distance $||f - g||$, en d'autres termes ce sera une transformation *isométrique.* Aussi n'était-il pas difficile de s'apercevoir que cette hypothèse permet d'être grandement élargie. Bientôt après les premières publications sur le sujet, en 1933 ou 1934, je me m'en souviens plus, la lecture du travail de M. Carleman m'a suggéré une méthode de démonstration qui s'appliquait à une classe étendue de contractions; ce sont les transformations linéaires qui ne font pas augmenter la distance. Mais à ce temps là, ma méthode me semblait ne pas réussir pour toutes les contractions et comme d'autre part j'étais presque sûr que les hypothèses pouvaient encore être élargies, je ne me hâtais pas de publier ma démonstration. D'ailleurs elle se propageait par des conversations et par correspondance et à la fin M. Hopf l'a incorporée, seulement pour le cas isométrique, dans l'excellent facicule qu'il a rédigé pour les „Ergebnisse der Mathematik". Aujourd'hui je sais que *la méthode embrasse toutes les contractions* et qu'elle permet encore d'autres généralisations. Je vous en parlerai tout à l'heure.

C'est en 1938, avant de compléter la démonstration dont je viens de parler, que j'ai réussi, par une autre voie, à étendre le théorème à toutes les contractions et à passer même aux espaces fonctionnels L^p pour lesquels le rôle de l'exposant 2 est joué par un nombre quelconque $p > 1$

ou même, sous certaines hypothèses additionnelles, pour $p=1$. Indépendamment et en même temps, les deux mathématiciens japonais: MM. Yosida et Kakutani ont découvert la même méthode et l'ont même formulée pour une classe d'espaces abstraits. Depuis lors, divers auteurs se sont occupés du problème, donnant des démonstrations et des généralisations nouvelles; citons entre autres MM. Carathéodory, Wiener et Birkhoff fils, Garrett Birkhoff. L'argument, déjà très simple de ce dernier m'en a suggéré, en 1941, un autre, d'ailleurs essentiellement différent, d'une simplicité, à ce que je pense, difficile à surpasser. C'est de cette démonstration que je veux tout d'abord vous parler.

Voici l'énoncé du théorème.

Etant donnés, dans l'espace de Hilbert, un élément f_1 et une contraction S, la suite des moyennes arithmétiques

$$\varphi_n = \frac{1}{n} \sum_1^n f_k$$

où $f_k = S^{k-1} f_1$, converge vers un élément invariant φ, c'est-à-dire qu'il existe un élément $\varphi = S\varphi$ tel que

$$\|\varphi - \varphi_n\| \to 0 .$$

Plus généralement on a

$$\| \varphi - \varphi_{m,n} \| \to 0 \qquad (n - m \to \infty) ,$$

où

$$\varphi_{m,n} = \frac{1}{n-m} \sum_{m+1}^n f_k .$$

Je suppose que vous connaissez les notations et les principaux faits qui sont à la base de la théorie de l'espace de Hilbert et je ne rappelle que deux d'entre eux dont nous aurons besoin. Le premier c'est l'inégalité du triangle

$$\| f + g \| \leqq \| f \| + \| g \|$$

et le second la relation

$$\| f + g \|^2 + \| f - g \|^2 = 2 \| f \|^2 + 2 \| g \|^2 . \tag{10}$$

Cette relation s'établit par un calcul évident, mais je pense qu'il n'est pas sans intérêt d'en indiquer aussi la source géométrique. La voici. L'ensemble des combinaisons linéaires de f et g n'est, en réalité, qu'un plan ordinaire de vecteurs et la relation (10) exprime tout simplement le fait bien connu que dans un parallélogramme, la somme des carrés des diagonales égale celle des carrés des côtés.

Après ces préliminaires, envisageons l'ensemble convexe G formé par toutes les combinaisons de la forme

$$g = \sum_1^{\nu} c_k \, f_k \tag{11}$$

à coefficients non-négatifs et de somme 1. Soit μ la borne inférieure des normes $||g||$.

Les moyennes φ_n et $\varphi_{m,n}$ appartiennent évidemment à l'ensemble G. Nous allons voir que $||\varphi_n|| \to \mu$ et $||\varphi_{m,n}|| \to \mu$ lorsque $n - m \to \infty$, ce que nous exprimons aussi en disant que $\{\varphi_n\}$ et $\{\varphi_{m,n}\}$ sont des *suites minimisantes.* Comme

$$|| \varphi_{m,n} || = || S^m \varphi_{n-m} || \leqq || \varphi_{n-m} || \;,$$

il suffit de considérer la suite $\{\varphi_n\}$. Il faut donc montrer que, pour ε arbitraire > 0 et pour n suffisamment grand,

$$|| \varphi_n || < \mu + \varepsilon \, .$$

A cet effet, soit d'abord g_1 un élément de G tel que

$$|| g_1 || < \mu + \frac{\varepsilon}{2} \tag{12}$$

et en partant de g_1 formons les analogues g_k et ψ_n de f_k et φ_n. Dans

$$\psi_n = \frac{1}{n} \sum_1^n g_k \;,$$

écrivons pour g_1 le second membre de (11) et pour les autres g_k substituons les expressions qui résultent de ce second membre lorsqu'on y applique les itérés de S. Comme $\sum c_k = 1$, il est manifeste que l'expression de ψ_n ainsi formée et celle de φ_n s'accordent dans leurs termes à partir du rang ν jusqu'au rang n et que par conséquent, les termes se détruisent dans la différence $\varphi_n - \psi_n$. Or les autres termes de cette différence, en nombre $2(\nu - 1)$, ont des coefficients ne dépassant pas $\frac{1}{n}$ en module. En somme

$$|| \varphi_n - \psi_n || \leqq 2(\nu - 1) \frac{1}{n} \, || f_1 || < \frac{\varepsilon}{2} \tag{13}$$

pour n suffisamment grand. D'autre part, S étant une contraction, l'hypothèse (12) relative à g_1 donne l'inégalité analogue pour les g_k, donc aussi pour ψ_n et enfin, par (13),

$$\|\varphi_n\| \leqq \|\varphi_n - \psi_n\| + \|\psi_n\| < \mu + \varepsilon ,$$

ce qu'il fallait prouver.

Cela étant, on achève la démonstration en rappelant le fait que toute suite minimisante converge vers un élément φ, le même pour toute telle suite. Ce fait, bien connu et dont plusieurs auteurs se sont servis pour établir le célèbre principe de Dirichlet, découle immédiatement de la relation (10). En effet, soit $g_1, g_2, \ldots$ une telle suite; alors comme la moyenne arithmétique de g_m et g_n appartient évidemment à l'ensemble G, on a $\|g_m + g_n\| \geqq 2\mu$ et de là, en appliquant encore la relation (10) à g_m et g_n on obtient

$$\|g_m - g_n\|^2 \leqq 2\|g_m\|^2 + 2\|g_n\|^2 - 4\mu^2 ;$$

comme de plus $\|g_m\| \to \mu$, $\|g_n\| \to \mu$, il en résulte que

$$\|g_m - g_n\| \to 0 ,$$

ce qui assure la convergence de la suite. Bien plus, la réunion $g_1, g'_1, g_2, g'_2, \ldots$ de deux suites minimisantes $\{g_k\}$ et $\{g'_k\}$ étant manifestement du même type, la limite φ est nécessairement la même pour les deux suites. Appliqué en particulier aux suites $\{\varphi_n\}$ et $\{\varphi_{n+1} = S\varphi_n\}$, ce fait assure aussi l'invariance de φ et avec cela, le théorème est complètement démontré.

La même méthode permet d'établir deux résultats plus généraux. Le premier auquel la méthode s'applique d'une façon évidente, indique que notre théorème reste valable quand on remplace l'espace de Hilbert par un espace linéaire complet *uniformément convexe*, d'ailleurs quelconques. On appelle ainsi les espaces linéaires pour lesquels la norme $\|f\|$, obéissant en outre aux règles usuelles comme $\|cf\| = |c|\,\|f\|$ et l'inégalité triangulaire satisfait encore à l'hypothèse que les suppositions $\|f\| \leqq 1 + \varepsilon$, $\|g\| \leqq 1 + \varepsilon$ et $\|f+g\| \geqq 2$ entraînent que $\|f - g\|$ devient infiniment petit en même temps ~~que~~ ε. Comme l'a montré M. Clarkson, les espaces fonctionnels L^p appartiennent, pour $p > 1$, à cette catégorie d'espaces; mais il n'en est pas ainsi lorsque $p = 1$. Il convient de dire que sous certaines conditions additionnelles, le théorème reste valable pour l'espace L^1, mais on le démontre par une méthode différente.

La seconde généralisation dont je veux parler, donnée par M. Dunford dans le Duke Journal en 1939, envisage, au lieu d'une seule contraction S, plusieurs contractions permutables entre elles. Pour fixer les idées, je me bornerai au cas de deux contractions, S et U.

En partant d'un élément donné $f_{1,1}$ de l'espace L^2, formoms les éléments $f_{i,k} = S^{i-1} U^{k-1} f_{1,1}$ et leurs moyennes

$$\varphi_{m,n\,;\,m',n'} = \frac{1}{(n-m)\,(n'-m')} \sum_{i=m+1}^{n} \sum_{k=m'+1}^{n'} f_{i,k} \,.$$

Lorsque $n - m \to \infty$, $n' - m' \to \infty$, ces moyennes *convergent vers un élément* φ,

$$\| \varphi - \varphi_{m,n\,;\,m',n'} \| \to 0 \ ,$$

et cet élément φ *est invariant par rapport à* S *et* U.

La démonstration ne diffère presque en rien de la précédente et bien entendu, elle reste aussi valable pour les espaces uniformément convexes. Je ne veux ajouter qu'une seule remarque. Observons d'abord que, en passant au langage géométrique, les $\varphi_{m,n\,;\,m',n'}$ sont les moyennes des $f_{i,k}$ qui correspondent aux points (i, k) à coordonnées entières, comprises dans certains rectangles que l'on fait croître indéfiniment. Ce que je voudrais ajouter c'est que le rôle des rectangles pourra aussi être joué par d'autres figures. Ainsi par exemple, le théorème subsiste pour chaque suite de figures convexes C_n dont l'aire croît indéfiniment, tandis que le quotient du périmètre par l'aire converge vers 0. La démonstration reste toujours la même; c'est seulement l'évaluation (13) qui devra être remplacée par la remarque évidente que le nombre des points (i, k) dont la distance au contour est inférieur à une quantité fixée, divisé par l'aire de la figure respective, devient infiment petit avec $1/n$.

C'est le lieu ici de revenir au sujet de ma première conférence et de vous indiquer en passant une seconde généralisation importante du théorème de Birkhoff, due à M. N. Wiener. Il s'y agit d'étendre le théorème au cas de plusieurs transformations ponctuelles du même type que celle qui figure dans le théorème original et qui de plus sont permutables entre elles, c'est-à-dire de la généraliser de la même façon que le théorème de Dunford le fait pour celui de Neumann. Wiener commence par baser la démonstration du théorème de Birkhoff sur un lemme qui n'est qu'un simple corollaire du „maximal ergodic theorem" dont je vous ai parlé dans ma première conférence, mais qui n'y est nullement équivalent et qu'il établit indépendamment. Pour parfaire ce lemme, il le combine avec le théorème de Neumann. Or, grâce peut-être à son imperfection, ce lemme de Wiener a l'avantage de pouvoir être étendu au cas de plusieurs transformations et en le combinant avec le théorème de Dunford, la généralisation en vue sera accomplie. Je n'entrerai pas dans les détails, vous les trouverez dans le Duke Journal de l'année 1939.

Permettez moi d'esquisser encore ma première démonstration du théorème de Neumann, celle qui est reproduite, comme je l'ai déjà dit, dans la monographie de M. Hopf, seulement pour le cas isométrique. J'ai réussi à la perfectionner, il y a peu de temps, en collaboration avec un de mes anciens élèves, M. Béla de Sz. Nagy qui d'ailleurs vient de rédiger, dans les mêmes „Ergebnisse", un excellent facicule embrassant les parties principales de la théorie de l'espace de Hilbert.

Nous envisageons deux sous-espaces de ce dernier. Le premier, soit E', se compose des éléments du type $f - Sf$ et de leurs limites. Lorsque f_1 appartient à cette classe, on pourra poser $f_1 = f - Sf + h$, avec $\| h \| < \varepsilon$ où ε est arbitrairement petit. De là découle

$$\varphi_{m,n} = \frac{S^{m+1} f - S^{n+1} f}{n - m} + \frac{1}{n - m} \sum_{k=m+1}^{n} S^k h ,$$

donc

$$\| \varphi_{m,n} \| \leqq \frac{2 \| f \|}{n - m} + \varepsilon$$

et par conséquent

$$\| \varphi_{m,n} \| \leqq 2\varepsilon$$

pour $n - m$ suffisamment grand. Donc $\varphi_{m,n} \to 0$.

Le second sous-espace, E'', se forme des invariants g de la contraction S, $Sg = g$. Lorsque $f_1 = g$, on a évidemment $\varphi_{m,n} = g \to g$.

Le théorème sera donc démontré si l'on réussit à décomposer tout élément f_1, du type général, en $f_1 = f_1' + f_1''$, ces derniers appartenant respectivement aux sous-espaces E' et E''. Or, *dans le cas où les invariants de S coïncident avec ceux de la transformation adjointe S^*, les deux sous-espaces sont orthogonaux et complémentaires.* En effet, l'identité générale

$$(f - Sf, g) = (f, g - S^* g)$$

qui vient immédiatement de la définition des transformations adjointes, fait aussitôt voir que l'ensemble des éléments g qui sont orthogonaux à E', c'est-à-dire à tous les $f - Sf$, est formé précisément par les invariants de S^*, ou alors, dans le cas actuel, par ceux de S. Par conséquent, dans ce cas, la décomposition exigée est évidente.

Ce que j'ai soupçonné depuis quelque temps et que nous avons réussi, à la fin, à prouver par un calcul extrêmement simple qui aurait dû sauter aux yeux dès le début, c'est que *le cas envisagé a lieu, sans exception, pour toute contraction S.*

En effet, avec les notations habituelles, l'hypothèse $g = S^*g$ entraîne que

$$(Sg, g) = (g, S^*g) = (g, g) = \| g \|^2$$

et

$$(g, Sg) = (S^*g, g) = \| g \|^2 ,$$

donc

$$\| Sg - g \|^2 = (Sg - g, Sg - g) =$$

$$= (Sg, Sg) - (Sg, g) - (g, Sg) + (g, g) = \| Sg \|^2 - \| g \|^2 .$$

Alors comme, par hypothèse, $\| Sg \| \leqq \| g \|$, on aura nécessairement

$$\| Sg - g \| = 0 , \quad g = Sg ,$$

ce qu'il fallait prouver.

En échangeant les rôles de S et S^*, il résulte que les deux catégories d'invariants coïncident et avec cela le théorème est démontré.

Le raisonnement que je viens d'esquisser, s'étend entre autres aux espaces fonctionnels L^p où $p > 1$. Mais au point de vue général, sa portée est beaucoup plus limitée que celle de la plupart des autres méthodes. Le temps nous manque de les énumérer toutes et de comparer leur efficacité. Je voudrais en mentionner une qui s'attache immédiatement à l'ordre d'idées de tout à l'heure. Oublions pour l'instant que nous avons réussi à prouver l'identité des invariants de S et S^*. Cette proposition serait même en défaut, si au lieu de nous borner à des contractions nous nous contentions de supposer seulement que les transformations S^k restent bornées dans leur ensemble, c'est-à-dire que $\| S^k f \| \leqq C \| f \|$ pour tous les f et tous les k. En particulier, cette hypothèse plus large est souvent adoptée dans la théorie des probabilités et dans celle des équations intégrales; consultez par exemple le Mémoire de M. Fréchet dans les Quaterly Journal de 1934. La méthode développée par MM. Yosida et Kakutani et par moi-même dans nos travaux de 1938 s'applique encore sous cette hypothèse plus générale. Mais c'est d'une autre méthode que je veux parler.

Oublions donc, je le répète, l'identité des invariants de S et de S^* et envisageons, outre E' et E'', les deux sous-espaces analogues, disons F' et F'', qui correspondent à la transformation S^*. Nous avons vu — et notre calcul s'adapte aussi à l'hypothèse plus générale dont nous venons de parler — que la limite φ existe lorsque f_1 appartient soit à E' soit à E'' et que $\varphi = 0$ dans le premier cas et $\varphi = f_1$ dans le second; les mêmes faits subsistent pour F' et F'' par rapport à S^*. De là il découle aussitôt que F' et F'' n'ont aucun élément commun sauf $f = 0$. Aussi, comme nous l'avons vu, les sous-espaces E' et E'' sont orthogonaux et complémentaires; de même F' et F''. Pour achever la démonstration, nous n'avons qu'à montrer que tout élément f est de la forme $f = f' + f''$ où f', f'' (orthogonaux ou non, cela ne fait rien) appartiennent respectivement à E' et E''. Or supposons le contraire. Il existerait, comme nous le savons, un élément $g \neq 0$, orthogonal à la fois à E' et E'' et par conséquent, compris à la fois dans F' et F'', contrairement au fait que, comme nous venons de le voir, ces deux sous-espaces n'ont, outre 0, aucun élément commun. Ainsi le théorème est démontré.

Cette façon de compléter mon essai de démonstration est dû à un autre de mes collaborateurs, M. Lorch[2]), de l'Université Columbia à New-York. Sa démonstration de notre théorème ou plutôt d'un théorème qui le généralise, se trouve dans le Bulletin de la même société, de l'année 1939. Dans ce qui précède, je viens de l'adapter au cas de l'espace de Hilbert; grâce à la structure particulière de cet espace le raisonnement devient plus simple que dans le cas général envisagé par Lorch. Là il s'agit des *espaces réflexifs*; ce sont les espaces de Banach qui sont les conjugués de leurs espaces conjugués; au lieu d'expliquer ces termes, qu'il nous suffise de dire que les espaces uniformément convexes ne sont qu'un cas bien particulier des espaces réflexifs. Observons encore que, dans le cas général, Lorch est obligé de se reporter à un théorème de M. Banach dont la vérification fait appel au principe de Zermelo d'après lequel tout ensemble peut être bien ordonné.

Quand on veut aller encore plus loin, une des tâches les plus intéressantes est de regarder ce que devient notre théorème quand on passe du rôle des itérés S^k à un semi-groupe de contractions et cela *sans les supposer permutables*. Déjà le problème le plus simple dans cette direction, à savoir: si le théorème de Dunford reste valable lorsque S et U ne sont plus permutables, montre les difficultés que présente la voie dans laquelle nous voulons entrer. Il sera intéressant de voir, avec M. Garrett Birkhoff, comment ces difficultés, sans être vaincues, peuvent néanmoins être tournées.

Etant données des contractions, en nombre fini ou infini, permutables ou non, parmi lesquels l'identité, envisageons l'ensemble des produits, à

[2]) D'ailleurs, si cela vous intéresse, Lorch est suisse romand par sa mère et c'est ici à Genève qu'il a reçu son éducation primaire. Boursier de voyage pendant l'année 1934/35, il vint en Europe pour travailler sous ma direction et le principal produit de cette collaboration est un Mémoire sur la résolution spectrale des transformations autoadjointes non bornées, inséré en 1936 dans les Transactions de la société mathématique d'Amérique.

un nombre quelconque de facteurs, que l'on en peut déduire. Formons toutes les „moyennes" au sens général de ces produits S, c'est-à-dire toutes les transformations du type

$$U = c_1 S_1 + c_2 S_2 + \cdots + c_n S_n \qquad (c_i \geqq 0, \, \Sigma c_i = 1)$$

et convenons de dire que g est un *conséquent* de f toujours s'il existe une moyenne U qui transforme f en g. Alors, pour chaque élément fixé f_0, ses conséquents convergent vers un élément déterminé φ la convergence ayant alors le sens suivant: à tout conséquent f_1 et à tout ε positif, on peut assigner un conséquent f_2 de f_1 de sorte que $\| \varphi - f \| < \varepsilon$ pour tout conséquent f de f_2.

La démonstration qu'en donne Garrett Birkhoff n'envisage que l'espace de Hilbert, cependant, par une modification légère, on pourra la faire embrasser tous les espaces L^p où $p > 1$. Faute de temps je n'entrerai pas dans les détails. D'ailleurs, le théorème cité n'est qu'une seule pièce dans la riche collection de généralisations pénétrantes que M. Garrett Birkhoff, à lui seul et en collaboration avec M. Alaoglu, viennent de recueillir au cours de ces dernières années.

(Reçu le 5 août 1944.)