Université de Montréal

LE THÉORÈME DE LEBESGUE SUR LA DÉRIVABILITÉ DES FONCTIONS À VARIATION BORNÉE

par

Patrick Landry Mombo Mingandza

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Mathématiques pures

février 2012

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

LE THÉORÈME DE LEBESGUE SUR LA DÉRIVABILITÉ DES FONCTIONS À VARIATION BORNÉE

présenté par

Patrick Landry Mombo Mingandza

a été évalué par un jury composé des personnes suivantes :

Qazi Ibadur Rahman
(président-rapporteur)

André Giroux
(directeur de recherche)

Richard Fournier
(membre du jury)

Mémoire accepté le:
25 Janvier 2012

SOMMAIRE

Dans ce mémoire, nous traiterons du théorème de Lebesgue, un des plus frappants et des plus importants de l'analyse mathématique ; à savoir qu'une fonction à variation bornée est dérivable presque partout. Le but de ce travail est de fournir, à part la démonstration souvent proposée dans les cours de la théorie de la mesure, d'autres démonstrations élaborées avec des outils mathématiques plus simples. Ma contribution a consisté essentiellement à détailler et à compléter ces démonstrations, puis à inclure la plupart des figures pour une meilleure lisibilité. Nous allons maintenant, pour ce théorème qui se présente sous d'autres variantes, en proposer l'historique et trois démonstrations différentes.

MOTS CLÉS : dérivée ; fonction ; intervalle ; mesure ; monotone ; variation bornée

SUMMARY

In this dissertation, we will be handling a theorem of Lebesgue, one of the most stricking and ultimate of mathematical analysis ; namely a function with bounded variation has a derivative almost everywhere. The aim of our research is to provide, apart from the proof usually offered in measure theory courses, other demontrations achieved with more simple mathematical tools. My contribution was primarily to simplify and to complete these demonstrations, to include the most of the drawings in order to visualize what is being said. For this theorem, which has other presentations, we will give now the history and three different demonstrations.

KEYWORDS : derivative ; function ; interval ; measure ; monotonic ; bounded variation

TABLE DES MATIÈRES

LISTE DES FIGURES

DÉDICACE

Je dédie ce mémoire à mon père Mingandza Jean Baptiste, à titre posthume, à ma mère Moutsinga Nicole, à mes frères et soeurs charnels, tous vivant au Gabon, et à tous les Gabonais. Je le dédie aussi à mes frères et soeurs spirituels chrétiens Témoins de Jéhovah du monde entier.

REMERCIEMENTS

Je remercie du fond du coeur mon directeur de recherche, Monsieur André Giroux, qui a fait preuve d'une disponibilité remarquable tout au long de ce travail. Son soutien moral et matériel qui n'a pas défaut m'a toujours permis d'aller de l'avant et de parachever cette oeuvre.
Ma gratitude va également à l'endroit de Guillaume Provencher, informaticien au DMS, qui m'a apporté une aide sans faille pour ce qui était de me familiariser avec des logiciels comme Latex et Inkscape que je ne connaissais même pas auparavant.
Je remercie aussi tous les autres professeurs du département qui m'ont dispensé leurs cours.

INTRODUCTION

Notre exposé nous amènera à parler en quelques mots de la thèse de Lebesgue et à donner l'énoncé de son théorème. Ensuite nous ferons un tour d'horizon sur les fonctions à variation bornée. Puis nous proposerons une première démonstration qui fait appel à la théorie générale de la mesure, telle que souvent donnée dans les cours de théorie de la mesure. La deuxième, celle de Riesz mais reprise par Boas englobe aussi bien le cas des fonctions continues que celui des fonctions discontinues, car il faut rappeler que Lebesgue avait établi son théorème avec l'hypothèse additionnelle de la continuité. La troisième repose uniquement sur les connaissances plus élémentaires de l'analyse réelle.

C'est dans les *Comptes Rendus des Séances de l'Académie* du 29 avril 1901 que Henri Lebesgue (1875-1941) expose les aspects essentiels d'une nouvelle définition de l'intégrale. Nous apprécierons cette nouvelle définition en faisant ressortir les différences avec les définitions antérieures des intégrales de Cauchy et de Riemann.

Lebesgue a su tirer profit de la nouvelle théorie de la mesure établie par Emile Borel (1871-1956) trois ans plus tôt et qu'il va généraliser, même si des années plus tard des querelles vont surgir entre les deux mathématiciens sur la paternité scientifique de la définition des ensembles de mesure nulle.

L'idée de mesure en mathématiques correspond à une généralisation de l'idée de longueur, ou, à plusieurs dimensions, à celles d'aire, de volume etc...La mesure de Borel est une extension à des ensembles plus généraux que des segments. Elle «mesure» ces ensembles et garde le principe d'additivité : la mesure d'une réunion dénombrable d'ensembles disjoints est la somme des mesures de ces ensembles. C'est à partir de cette idée de Borel que Lebesgue va définir son intégrale : dans une intégrale, il y a deux quantités qui jouent un rôle de premier plan, la mesure de l'intervalle de variation de la fonction et les valeurs que la fonction prend sur cet «intervalle».

Lebesgue voulait intégrer des «fonctions» plus sauvages que les fonctions habituelles (les fonctions continues). Parmi les fonctions discontinues, l'une des plus pathologiques (à l'époque) est la fonction de Dirichlet définie sur le segment $[0, 1]$ et qui vaut 1 sur les rationnels et 0 sur les irrationnels.

L'intégrale de Riemann (1826-1866) est une généralisation de l'intégrale de Cauchy (1789-1857). Le problème initial est de calculer l'aire située sous une courbe définie par une fonction $f(x)$. Cauchy remplace l'aire située sous la courbe de $f(x)$ par la somme des aire des rectangles jointifs (voir la figure 0.1). La base de chacun de ces rectangles (un intervalle) est une portion du segment $[a, b]$ et la hauteur est égale à la valeur de la fonction au «début» du segment. Cauchy montre (1) que la valeur de l'aire des rectangles additionnés est égale à celle sous la courbe continue quand la longueur de la plus grande base des rectangles tend vers zéro, et (2) que cette valeur limite ne dépend pas du découpage en rectangles. Cauchy«montre» plus qu'il ne démontre, car il suppose implicitement que la fonction f est uniformément continue. Cauchy ne sait pas intégrer des fonctions qui sont discontinues en un nombre infini de points, c'est-à-dire qui font des sauts d'une valeur r à une valeur s.

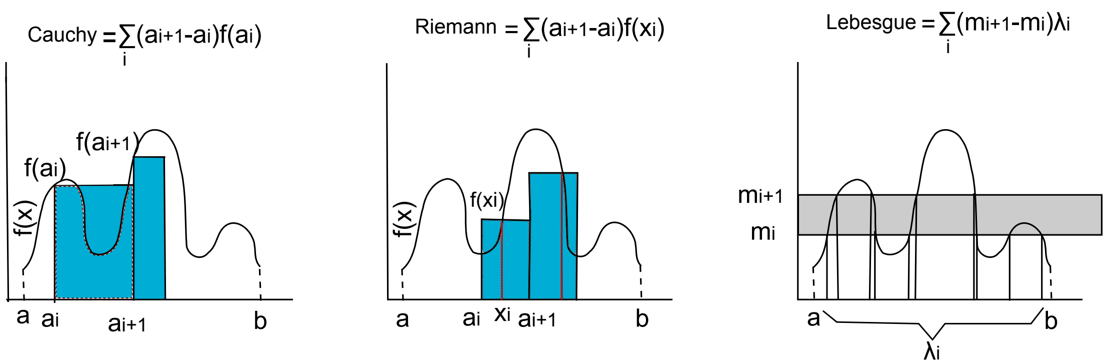

Fig. 0.1. Définitions des différentes intégrales de Cauchy, Riemann et Lebesgue.

La définition de l'intégrale de Riemann est une variation subtile de l'intégrale de Cauchy. Riemann prend pour hauteur du rectangle défini sur un intervalle, non plus la valeur du début de l'intervalle, mais une valeur de la fonction quelconque à l'intérieur de chaque intervalle (voir la figure 0.1). Quand on fait tendre la longueur maximale des intervalles vers zéro, et que la limite de la somme ne dépend ni du découpage en intervalles ni de la valeur des hauteurs, l'intégrale de Riemann existe. Il est évident que les valeurs des deux intégrales sont égales pour des fonctions continues, mais, avec la définition de Riemann, les mathématiciens savent intégrer des fonctions plus générales incluant par exemple des fonctions qui ont un nombre infini de discontinuités.

Avec l'intégrale de Riemann, on ne sait pas intégrer la fonction de Dirichlet, car sur chaque petit intervalle la fonction f oscille entre 0 et 1 : quelque soit l'intervalle, et aussi petit soit-il, il y a une infinité de valeurs rationnelles de l'abscisse pour lesquelles la valeur de la fonction vaut 1 et une infinité plus grande encore de valeurs irrationnelles pour lesquelles la valeur de la fonction vaut 0.

Lebesgue a un point de vue différent : au lieu de découper le segment $[a, b]$ en parties, il découpe en tranches l'intervalle des valeurs de la fonction $f(x)$ et réunit dans un même lot les «intervalles» pour lesquels la fonction est comprise entre deux valeurs (voir la figure 0.1 ; où λ_i désigne la mesure des ensembles des points x de $[a, b]$ pour lesquels $f(x)$ est compris entre m_i et m_{i+1}). Puis Lebesgue définit deux sommes et quand ces deux sommes ont la même limite (alors que la valeur du plus grand des intervalles de valeurs de la fonction tend vers zéro), la fonction f est «intégrable».

Dans sa thèse de doctorat qui s'intitule *Intégrale, longueur, aire* Lebesgue voulait élargir les ensembles de fonctions qu'il pouvait intégrer par son découpage. C'est ainsi que le problème de l'intégration de la fonction de Dirichlet fut résolu grâce à l'intégrale de Lebesgue. En effet, l'intégrale de Lebesgue de la fonction de Dirichlet revient à additionner deux valeurs sur des ensembles différents, l'une égale à 1 pour les nombres rationnels sur un ensemble de mesure nulle (donc cette contribution des rationnels est égale à zéro), et l'autre égale à 0 sur l'ensemble des irrationnels dont la mesure est égale à 1, mais cela n'a guère d'importance, car cette contribution est évidémment nulle, la fonction étant nulle en tous ces points. En conclusion, l'intégrale de la fonction de Dirichlet est nulle.

Depuis Riemann et Weierstrass, les mathématiciens savent déjà que les fonctions dérivables ne forment qu'une classe particulière dans l'ensemble des fonctions continues ; mais dès qu'on ne veut plus faire usage de la dérivée, les difficultés rencontrées sont telles qu'on se rend compte que les fonctions dérivables peuvent être utilisées presque uniquement en analyse.

Il fallait donc rechercher quand on pouvait affirmer qu'une fonction est dérivable, s'était dit Lebesgue, c'est-à-dire il fallait rechercher des conditions suffisantes pour l'existence de la dérivée.

Mais au lieu de rechercher quelles sont les fonctions dérivables, Lebesgue a pensé plutôt à étendre cette notion de dérivée. Toutefois, les mathématiciens du Bois-Reymond et Dini y avaient songé avant lui par l'introduction des nombres dérivés. Relativement à ces nombres, on pouvait se proposer des questions semblables à celles que l'on résout pour la dérivée ; en particulier, on peut rechercher $f(x)$ connaissant l'un de ses nombres dérivés, $\Lambda_d[f(x)]$[1] par exemple. Si Λ_d est intégrable au sens de Riemann, l'intégration de Λ_d fournit $f(x)$. Lebesgue qui avait déjà défini l'intégrale qui porte son nom dans sa thèse de doctorat et qui généralise celle de Riemann, va tout simplement utiliser ce résultat à l'aide de son intégrale. Ce résultat sera donc étendu à des cas plus généraux. En clair : ***Si*** $\Lambda_d[f(x)](ou\ \Lambda_g, \lambda_d, \lambda_g)$ ***est fini pour chaque valeur de x, pour que*** Λ_d ***ait une intégrale, il faut et il suffit que f(x) soit à variation bornée, et alors on a :***

$$f(x) - f(a) = \int_a^x \Lambda_d[f(x)]\, \mathrm{dx} \tag{0.0.1}$$

Si λ_d est aussi fini, de l'égalité analogue à (0.0.1) relative à λ_d, on déduit que les points où Λ_d et λ_d sont différents, c'est-à-dire ceux où $f(x)$ n'a pas de dérivée à droite, forment un ensemble de mesure nulle. En raisonnant ainsi, on voit que : ***si l'un des nombres dérivés d'une fonction continue à variation bornée f(x) est toujours fini, l'ensemble des valeurs de x, pour lesquelles f(x) n'a pas de dérivée, est de mesure nulle, et l'on a :***

$$\int_a^b f'(x)\, \mathrm{dx} = f(b) - f(a), \tag{0.0.2}$$

[1]Cette notation signifie la dérivée à droite supérieure de la fonction f en x. Les quatre dérivées d'une fonction donnée en un point quelconque sont définies à la page 17.

l'intégrale n'étant étendue qu'à l'ensemble des valeurs où la dérivée existe.

Un exemple des fonctions qui remplissent les conditions énoncées sont les fonctions à nombres dérivés bornés, ou fonctions satisfaisant à la condition de Lipschitz :

$$|f(x+h) - f(x)| < M\,|h|$$

Enfin, Lebesgue ira encore plus loin en affirmant que : ***Si une fonction continue f(x) est monotone ou plus généralement est à variation bornée, l'ensemble des valeurs de x, pour lesquelles f'(x) n'existe pas, est de mesure nulle ;*** sauf que l'égalité (0.0.2) n'est plus nécessairement vraie.

La notion de mesure linéaire est une généralisation de la notion de longueur d'un segment, une autre généralisation conduit à la définition de la longueur d'un arc de courbe. C'est en étudiant les questions relatives à la rectification des courbes que Lebesgue aura l'occasion d'appliquer quelques uns des résultats obtenus sur l'intégrale et les nombres dérivés pour ensuite établir son théorème.

Chapitre 1

LES FONCTIONS À VARIATION BORNÉE

Une fonction f définie sur un intervalle $I \subseteq \mathbb{R}$ et à valeurs dans $\mathbb{R}$ est dite monotone si elle est soit décroissante soit croissante. C'est-à-dire, f est monotone si pour x et y dans I : soit $y > x$ entraîne $f(y) \leq f(x)$, soit $y > x$ entraîne $f(y) \geq f(x)$. Si une de ces conditions tient avec une inégalité stricte, on dit que f est *strictement monotone*. Les fonctions usuelles qu'on utilise en analyse, si elles ne sont pas monotones, sont au moins monotones par morceaux. D'ailleurs Lejeune-Dirichlet a démontré la convergence des séries de Fourier pour ces fonctions ; nous y reviendrons plus loin. Par exemple, si $f(x) = x^2$, f est décroissante lorsque $x < 0$ et croissante lorsque $x > 0$; si $f(x) = cos\ x$, f est alternativement croissante et décroissante dans les intervalles $[-\pi, 0]$,$[0, \pi]$, et ainsi de suite ; si $f(x) = e^x$, f est croissante sur $\mathbb{R}$ tout entier. Toutes ces fonctions sont continues. Par contre, la fonction f définie par $f(x) = [x]$ (partie entière de x) est croissante et a une discontinuité à chaque entier x.

On dit qu'une fonction f a un *saut* à un point x de son domaine si f a des limites à la fois à gauche et à droite de x, mais n'est pas continue en x. On peut montrer facilement que les seules discontinuités d'une fonction monotone sont les sauts. Les fonctions monotones les plus faciles à visualiser sont celles qui ont un nombre fini de sauts, mais une fonction monotone peut avoir une structure beaucoup plus compliquée que cela. Par exemple, si $f(x) = 2^{-n}$ dans l'intervalle $[1/(n+1), 1/n[$, f est une fonction croissante avec des sauts qui ont un point limite à 0. Dans la dernière démonstration de ce travail, celle de W. Botsko, nous montrerons que les points de discontinuité d'une fonction monotone sont au plus dénombrables.

Il existe des fonctions monotones continues, non constantes, dont la dérivée est nulle presque partout. De telles fonctions sont appelées fonctions monotones *singulières*. Pour donner un exemple, construisons d'abord une fonction singulière f sur l'ensemble de Cantor de la façon suivante : on écrit tous les points x de l'ensemble de Cantor dans la base 3, en n'utilisant aucun chiffre 1, en divisant tous les chiffres par 2 , et en posant $f(x)$ égale au nombre obtenu interprété en base 2. Une telle fonction définie sur l'ensemble de Cantor a la même valeur aux extrémités des intervalles complémentaires de l'ensemble de Cantor. On étend f sur tout $[0,1]$ en lui donnant la même valeur dans tout l'intervalle complémentaire que celle qu'elle prend aux deux extrémités de celui-ci. Cette fonction est continue sur $[0,1]$. À gauche d'une extrémité d'un intervalle complémentaire, la dérivée à droite f'_+[1] existe et est 0 ; pareillement la dérivée à gauche $f'_-(x) = 0$ à droite d'une extrémité x. Si on appelle C cette même fonction étendue à $]-\infty,+\infty[$ en posant $C(x) = 0$ pour $x \leq 0$ et $C(x) = 1$ pour $x \geq 1$ (cette fonction est monotone singulière et continue sur $\mathbb{R}$), et si $\{r_n\}$ est une suite de nombres rationnels, alors la fonction g définie par

$$g(x) = \sum_{n=1}^{\infty} 2^{-n} C(2^n(x - r_n))$$

est continue et strictement croissante non constante et dont la dérivée est nulle presque partout. En effet, comme la série $\sum_{n=1}^{\infty} 2^{-n} C(2^{-n}(x - r_n))$ est uniformément convergente (la preuve est évidente par le M-test de Weierstrass, voir le lemme 1.0.1), et comme ses termes sont des fonctions continues alors g est continue . Aussi g est strictement croissante puisque si $x < y$, il y a un nombre rationnel r entre x et y. Donc $C(2^n(x - r)) = 0 < C(2^n(y - r))$. Pour chaque $m \neq n$, on a :

$$C(2^m(x - r)) \leq C(2^m(y - r))$$

[1]Pour une fonction donnée f, on note $f^+(x)$, $f_+(x)$, $f^-(x)$, $f_-(x)$ (ou Λ_d, λ_d, Λ_g, λ_g) respectivement la dérivée à droite supérieure, la dérivée à doite inférieure, la dérivée à gauche supérieure, la dérivée à gauche inférieure en un point x (voir à la page 17 les définitions de ces dérivées). Si $f^+(x) = f_+(x)$, on dit qu'il y a une dérivée à droite en x, et on la note par $f'_+(x)$, de même on note par $f'_-(x)$ la dérivée à gauche en x.

parce que C est croissante. Ainsi $g(x) < g(y)$.

Enfin, par le théorème de Fubini sur la différentiation des séries de fonctions croissantes (voir le lemme 1.0.2), on a que : $g'(x) = \sum_{n=1}^{\infty} C'(2^n(x - r_n)) = 0$ pour presque tout x .

Une autre application de cette fonction singulière C est de construire deux fonctions qui ont la même dérivée (infinie en certains points) dans tout un intervalle, mais ne diffèrent pas par une constante. Il suffit de trouver une fonction h qui soit continue et croissante ayant une dérivée finie à chaque point n'appartenant pas à l'ensemble de Cantor et une dérivée $+\infty$ à chaque point de l'ensemble de Cantor (voir la construction d'une telle fonction aux pages 10 et 11). Une fois qu'on obtient une telle fonction, on pose $k(x) = g(x) + h(x)$; alors $k'(x) = h'(x) = +\infty$ en tous les points de l'ensemble de Cantor ; et $k'(x) = h'(x)$ en tous les points n'appartenant pas à l'ensemble de Cantor, car $g'(x) = 0$ en de tels points. Cependant, h et k diffèrent de g, qui n'est pas une constante.

Les fonctions à variation bornée forment une classe de fonctions définies par Camille Jordan de la façon suivante :

Soit une fonction $f(x)$ bornée[2] définie dans un intervalle fermé borné $[a, b]$. Partageons $[a, b]$ à l'aide des points

$$x_0 = a \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n = b;$$

la somme

$$v = |f(x_1) - f(x_0)| + |f(x_2) - f(x_1)| + \cdots + |f(x_n) - f(x_{n-1})|$$

est ce que l'on appelle la variation de $f(x)$ pour le système de points $x_0, x_1, \ldots, x_n$. Si, quel que soit le système des points de division, v reste bornée, la fonction est dite à *variation totale finie* ou, simplement, à *variation bornée* ; la variation totale étant, par définition, la borne supérieure de v quand le maximum λ de la longueur des intervalles partiels employés tend vers zéro. Ce qui précède nous permet de citer des fonctions à variation bornée. Une fonction croissante est, en effet, une fonction à variation totale finie et égale à $f(b) - f(a)$; de même, une fonction

[2] Il est d'ailleurs évident qu'une fonction non bornée ne peut satisfaire aux définitions qui suivent.

décroissante est à variation bornée.

Toutefois, voici comment, à l'origine, Camille Jordan a présenté ces fonctions pour la première fois : Soit $y = f(x)$ une fonction d'une variable réelle x, bornée dans un intervalle $[a, b]$ contenant des valeurs particulières x_0 et $X > x_0$. Donnons à x une suite de valeurs croissantes $x_0, x_1, \ldots, x_{n-1}, X$, et soient $y_0, y_1, \ldots, y_{n-1}, Y$, les valeurs correspondantes de y. On aura :

$$Y - y_0 = \sum (y_k - y_{k-1}) = p - n \tag{1.0.3}$$

p désignant la somme des termes positifs, $-n$ celle des termes négatifs de la somme ci-dessus.

Nous dirons que p est la *variation positive* de y et n sa *variation négative* pour le système de valeurs $x_0, x_1, \ldots, X$.

La somme

$$t = \sum |y_k - y_{k-1}| = p + n \tag{1.0.4}$$

est sa *variation totale*.

En changeant le nombre et la position des valeurs intermédiaires $x_1, \ldots, x_{n-1}$, on pourra faire varier ces trois sommes. En particulier, si entre x_{k-1} et x_k on intercale une nouvelle valeur η, ces sommes conserveront leur valeur initiale si $f(\eta)$ est compris entre y_{k-1} et y_k ; sinon p et n seront augmentés tous deux de la différence entre $f(\eta)$ et le plus proche des nombres y_{k-1}, y_k, et t sera augmentée du double de cette différence.

Cela dit, supposons qu'un des trois systèmes de sommes p, n, t admette un maximum. Il en sera de même de chacun des deux autres puisque tous les trois sont liés par les équations (1.0.3) et (1.0.4). On dira, dans ce cas , que $y = f(x)$ est une fonction à variation bornée entre x_0 et X.

Un peu plus tard, Jordan établira deux autres résultats importants sur les fonctions à variation bornée à savoir : (1) ***Une fonction f(x) à variation bornée dans un intervalle*** $[a, b]$ ***est intégrable dans cet intervalle***, (2) ***Une fonction f(x) continue et à variation bornée dans l'intervalle*** (x_0, X) ***est la différence de deux fonctions continues et croissantes***.

Dans la recherche des conditions suffisantes sur la convergence des séries de Fourier, le mathématicien allemand Lejeune-Dirichlet, avait montré que les fonctions ayant un nombre fini de maxima et de minima dans l'intervalle $[0, 2\pi]$ ont leurs séries de Fourier qui convergent (théorème de Dirichlet). Mais Camille Jordan qui va plus tard découvrir les fonctions à variation bornée montrera à son tour que ce type de fonctions ont leurs séries de Fourier qui convergent également. Or une fonction admettant un nombre fini de maxima et de minima est une fonction à variation bornée (voir le lemme 1.0.3 un peu plus loin). Autrement dit, les conditions de Jordan seront une extension de celles de Dirichlet. Par conséquent, Jordan aura généralisé le théorème de Dirichlet. C'est pourquoi il sera finalement appelé le théorème de Dirichlet-Jordan.

Lemmes sans démonstrations

Lemme 1.0.1. *(ou **M-test de Weierstrass**) Si $\{f_n\}_n$ est une suite de fonctions sur un domaine E, et $|f_n(x)| \leq M_n$ pour tout x dans E (M_n étant indépendant de x) alors, si la série $\sum M_n$ converge alors la série $\sum f_n$ converge uniformément.*

Lemme 1.0.2. ***(théorème de Fubini : voir à la page 12 du livre de Riesz)*** *Soit $f_1 + f_2 + \ldots$ une série convergente point par point de fonctions croissantes sur un intervalle $[a, b]$, ayant pour somme s. Alors pour presque tout x dans $[a, b]$ on a : $f_1'(x) + f_2'(x) + \ldots = s'(x)$.*

Construction d'une fonction h continue et croissante ayant une dérivée finie à chaque point n'appartenant pas à l'ensemble de Cantor et une dérivée $+\infty$ à chaque point de l'ensemble de Cantor.

Énumérons les intervalles complémentaires (a_n, b_n) de l'ensemble de Cantor en ordre des longueurs décroissant. Soit φ_n une fonction continue et croissante comme indiqué sur la figure ci-dessous.

$\varphi_n(x) = 0$ pour tout $x < a_n$, $\varphi_n(x) = 1$ pour $x > b_n$, $\varphi'_{n+}(a_n) = \varphi'_{n-}(b_n) = +\infty$ (Un exemple explicite est $\varphi_n(x) = (2/\pi)\tan^{-1}\left\{(x - a_n)^{1/2}(b_n - x)^{-1/2}\right\}$.). On peut constater que les longueurs des intervalles $]a_n, b_n[$ sont les puissances entières négatives de 3 ; soit $\alpha_n = (\frac{2}{5})^m$ si $]a_n, b_n[$ a pour longueur 3^{-m}. Définissons $h(x) = \sum \alpha_n \varphi_n(x)$ avec la somme (infinie) étendue sur tous les n. En d'autres termes, la valeur $h(x)$ est la somme des α_n sur tous les intervalles $]a_n, b_n[$ à gauche

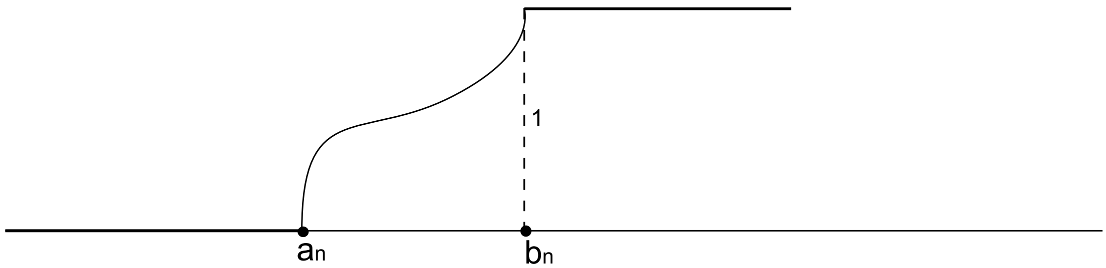

FIG. 1.1. Fonction φ_n définie sur $[a_n, b_n]$.

de x, plus $\alpha_k\varphi_k(x)$ si x est dans $]a_k, b_k[$. Or il y a 2^{m-1} intervalles $]a_n, b_n[$ de longueur 3^{-m}, et $\alpha_n = (\frac{2}{5})^m$ sur chaque, donc $\sum \alpha_n$ converge. Par conséquent, la série définissant h converge uniformément (M-test de Weierstrass). Par construction, h est croissante. Soit x un point de l'ensemble de Cantor, autre que a_n, et soit $\delta > 0$. Le rapport

$$\begin{aligned}\Delta =& \delta^{-1}\{h(x+\delta) - h(x)\}\\ =& \delta^{-1}\sum \alpha_n \{\varphi_n(x+\delta) - \varphi_n(x)\}\end{aligned}$$

excédera α_k si l'intervalle $]x, x+\delta[$ contient l'intervalle complémentaire $]a_k, b_k[$. Il est facile de voir que $]x, x+\delta[$ contient toujours un intervalle complémentaire de longueur $3^{-m} \geq \delta/9$. Alors $\Delta \geq (\frac{2}{5})^m/\delta \geq \frac{1}{9}3^m(\frac{2}{5})^m$ tend vers ∞, donc $h'_+(x) = +\infty$. Si, toutefois, x est égal à a_n alors $h'_+(x) = +\infty$ par observation. De même $h'_-(x) = +\infty$ en tout point de l'ensemble de Cantor. C'est-à-dire, $h'(x) = +\infty$ en tout point de l'ensemble de Cantor.

Lemme 1.0.3. *Si une fonction admet un nombre fini de maximums et de minimums sur l'intervalle* $[0, 2\pi]$, *alors elle est à variation bornée.*

Preuve :

Soit une fonction $f : [a, b] \to \mathbb{R}$. À toute partition $P = (a = x_0 < x_1 < \ldots < x_n = b)$ de $[a, b]$, on associe la *variation totale* de f sur P :

$$V(f, P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|$$

On pose

$$V_a^b(f) = \sup_P V(f, P) \quad \in [0, \infty]$$

que l'on appelle la *variation* de f sur $[a,b]$. Si f est croissante sur $[a,b]$, on obtient, pour toute partition P de $[a,b]$,

$$V(f,P) = \sum_{k=1}^{n}(f(x_k) - f(x_{k-1})) = f(b) - f(a).$$

Donc, une fonction croissante f est à variation bornée sur $[a,b]$ et $V_a^b(f) = f(b) - f(a)$. De même, en posant $g = -f$ et en se servant de ce qui précède, on montre qu'une fonction f décroissante sur $[a,b]$ est à variation bornée, de variation $V_a^b(f) = f(a) - f(b)$. Ainsi, une fonction *monotone* est à variation bornée.
Soit f définie sur $[0,2\pi]$ et ayant un nombre fini de maximums et de minimums sur cet intervalle. Soit $m_1, m_2, \ldots, m_s$ les abscisses de ces extrémums prises dans cet ordre. Posons :

$$\begin{gathered} f_1 \text{ la restriction de } f \text{ sur } [0,m_1] \\ f_2 \text{ la restriction de } f \text{ sur } [m_1,m_2] \\ \vdots \\ f_s \text{ la restriction de } f \text{ sur } [m_{s-1},m_s]. \end{gathered}$$

Donc $f(x) = \sum_{j=1}^{s} f_j(x)$, et f est une somme finie de fonctions à variation bornée, puisque sur chacun des intervalles $[0,m_1],[m_1,m_2],\ldots,[m_{s-1},m_s]$ les f_j sont monotones respectivement. Nous allons montrer que si f est la somme de f_1 et f_2 qui sont à variation bornée alors elle-même est à variation bornée, et on pourra montrer de proche en proche que c'est aussi vrai avec une somme de s fonctions.
Soit P une partition de $[0,2\pi]$, $P' = (0 = x_0 < x_1 < \ldots < x_n = 2\pi)$ la partition obtenue en rajoutant à P le point c s'il n'y est pas. Notons p l'entier tel que $x_p = c$; $P_1 = (0 = x_0 < x_1 < \ldots < x_p = c)$ est une partition de $[0,c]$ et $P_2 = (x_p < \ldots < x_n = 2\pi)$ une partition de $[c,2\pi]$ On a $V(f,P) \leq V(f,P')$: on rajoute au plus un point dans la partition et l'inégalité résulte de l'inégalité triangulaire. D'autre part, $V(f,P') = V(f_1,P_1) + V(f_2,P_2)$. On en déduit

$$V(f,P) \leq V(f_1,P_1) + V(f_2,P_2) \leq V_0^{m_1}(f_1) + V_{m_1}^{m_2}(f_2)$$

On a donc $V_0^{2\pi}(f) \leq V_0^{m_1}(f_1) + V_{m_1}^{m_2}(f_2)$. Ce qui veut dire que f est à variation bornée. En fait, on peut montrer qu'il y a égalité. En effet, f est à variation bornée sur $[0, c]$ et sur $[c, 2\pi]$. Soit $\epsilon > 0$. On peut trouver des partitions P_1 et P_2 de $[0, c]$ et $[c, 2\pi]$ respectivement telles que : $V(f_1, P_1) \geq V_0^c(f_1) - \frac{\epsilon}{2}$ et $V(f_2, P_2) \geq V_c^{2\pi}(f_2) - \frac{\epsilon}{2}$. Considérons la partition P de $[0, 2\pi]$ obtenue par réunion de P_1 et de P_2. On a alors $V(f, P) = V(f_1, P_1) + V(f_2, P_2) \geq V_0^c(f_1) + V_c^{2\pi}(f_2) - \epsilon$. ϵ étant arbitraire, on a finalement : $V_0^{2\pi}(f) = V_0^c(f_1) + V_c^{2\pi}(f_2)$.

Chapitre 2

LA DÉMONSTRATION STANDARD

Comme déjà dit dans l'introduction, cette démonstration est tributaire de la théorie de la mesure. L'un des éléments clé sera la mesurabilité des dérivées de Dini. Nous précisons que λ est la mesure de Lebesgue.

Théorème 2.1. *Une fonction à variation bornée $\phi : [a,b] \to \mathbb{R}$ est dérivable presque partout.*

Pour démontrer ce théorème nous nous servirons du lemme de Vitali que voici :

Lemme 2.0.4. *Soit $E \subseteq \mathbb{R}$ un ensemble de mesure extérieure finie. Supposons que $\{I_\alpha\}_{\alpha\in A}$ est une famille (pas nécessairement dénombrable) d'intervalles d'intérieur non vide ayant la propriété suivante : pour tout $x \in E$ et pour $\epsilon > 0$ arbitrairement petit, on peut trouver $\alpha \in A$ tel que $x \in I_\alpha$ et $\lambda(I_\alpha) < \epsilon$. Alors à chaque $\epsilon > 0$ correspond un ensemble fini d'intervalles deux à deux disjoints $\{I_{\alpha_1}, I_{\alpha_2}, \ldots, I_{\alpha_N}\}$ tels que*[1]

$$\lambda^* \left(E \cap \left(\sum_{k=1}^{N} I_{\alpha_k} \right)^c \right) < \epsilon$$

Démonstration.

En considérant si nécessaire leur adhérence, on peut supposer que les intervalles I_α sont fermés. Soit $O \subseteq \mathbb{R}$ un ensemble ouvert de mesure finie contenant E. En ne considérant si nécessaire que ceux qui le sont, on peut supposer que tous les

[1]Uniquement dans ce chapitre, on convient de noter une intersection de deux ensembles A et B par AB au lieu de $A \cap B$. Et une réunion disjointe d'ensembles $A_1, A_2, \ldots, A_n$ par $\sum_{k=1}^{n} A_k$.

intervalles I_α sont contenus dans O.

Formons alors une suite d'intervalles disjoints $\{I_{\alpha_k}\}_k$ de la façon suivante.I_{α_1} est choisi arbitrairement. Si les intervalles disjoints $I_{\alpha_1}, I_{\alpha_2}, \ldots, I_{\alpha_n}$ ont déjà été choisis, posons

$$\Lambda_n = \sup_{\alpha \in A} \{\lambda(I_\alpha)/I_\alpha I_{\alpha_k} = \varnothing \text{ pour } 1 \leq k \leq n\}$$

et choisissons un intervalle I_{α_n+1} tel que

$$I_{\alpha_n+1} I_{\alpha_k} = \varnothing \text{ pour } 1 \leq k \leq n \text{ et } \lambda(I_{\alpha_n+1}) > \frac{\Lambda_n}{2}.$$

Si ce processus s'arrête après N étapes (faute d'intervalles I_α remplissant la condition), on a nécessairement

$$E \subseteq \sum_{k=1}^{N} I_{\alpha_k}$$

car si

$$x \in E \cap \left(\sum_{k=1}^{N} I_{\alpha_k}\right)^c,$$

il existe, par hypothèse, un intervalle I_α contenant x et tel que $I_\alpha I_{\alpha_k} = \varnothing$ pour $1 \leq k \leq N$. S'il ne s'arrête jamais, la relation

$$\sum_{k=1}^{+\infty} \lambda(I_{\alpha_k}) \leq \lambda(O) < +\infty$$

implique que

$$\lim_{k \to +\infty} \lambda(I_{\alpha_k}) = 0$$

et aussi que l'on peut trouver N tel que

$$\sum_{k>N} \lambda(I_{\alpha_k}) < \frac{\epsilon}{5}.$$

Montrons dans ce cas, que

$$\lambda^* \left(E \cap \left(\sum_{k=1}^{N} I_{\alpha_k}\right)^c\right) < \epsilon.$$

Soit en effet

$$x \in E \cap \left(\sum_{k=1}^{N} I_{\alpha_k}\right)^c.$$

Il existe par hypothèse, un intervalle $I_{\alpha'}$ contenant x et tel que $I_{\alpha'} I_{\alpha_k} = \varnothing$ pour $1 \leq k \leq N$. D'autre part, il doit aussi exister des intervalles de la suite $\{I_{\alpha_k}\}_{k \in \mathbb{N}}$ tels que $I_{\alpha'} I_{\alpha_k} \neq \varnothing$. Autrement on aurait $\lambda(I_{\alpha'}) \leq \Lambda_k < 2\lambda(I_{\alpha_k+1})$ pour tout

$k \in \mathbb{N}$, ce qui impliquerait $\lambda(I_{\alpha'}) = 0$. Soit donc I_{α_n} le premier des intervalles de la suite $\{I_{\alpha_k}\}_{k\in\mathbb{N}}$ tel que $I_{\alpha'}I_{\alpha_n} \neq \varnothing$. Alors $n > N$ et

$$\lambda(I_{\alpha'}) \leq \Lambda_{n-1} < 2\lambda(I_{\alpha_n}).$$

Ainsi la distance de x au centre x_n de l'intervalle I_{α_n} est au plus

$$\lambda(I_{\alpha'}) + \frac{1}{2}\lambda(I_{\alpha_n}) < \frac{5}{2}\lambda(I_{\alpha_n})$$

et

$$x \in [x_n - \frac{5}{2}\lambda(I_{\alpha_n}), x_n + \frac{5}{2}\lambda(I_{\alpha_n})].$$

Donc

$$\lambda^*\left(E \cap \left(\sum_{k=1}^{N} I_{\alpha_k}\right)^c\right) \leq \sum_{n>N} 5\lambda(I_{\alpha_n}) < \epsilon.$$

$\square$

Lemme 2.0.5. *Si un ensemble $E \subseteq \mathbb{R}$ est mesurable alors à chaque $\epsilon > 0$ correspond un ensemble ouvert $O_\epsilon \subseteq \mathbb{R}$ tel que $E \subseteq O_\epsilon$ et que $\lambda(O_\epsilon E^c) < \epsilon$*

Preuve :

Cas où $\lambda(E) < +\infty$

Soit O_ϵ un ensemble ouvert tel que $E \subseteq O_\epsilon$ et que

$$\lambda(O_\epsilon) < \lambda(E) + \epsilon.$$

Si E est mesurable, on a

$$\lambda(O_\epsilon) = \lambda(E) + \lambda(O_\epsilon E^c) > \lambda(O_\epsilon) - \epsilon + \lambda(O_\epsilon E^c)$$

et donc

$$\lambda(O_\epsilon E^c) < \epsilon.$$

Cas où $\lambda(E) = +\infty$

Posons $I_n =]-n, n[, \quad E_n = EI_n$. Si E est mesurable, chaque ensemble E_n l'est. Si donc O_m est un ouvert tel que

$$\lambda(O_m {E_m}^c) < \frac{\epsilon}{2^{m+1}}$$

on aura , en posant

$$O = \bigcup_{m\in\mathbb{N}} O_m,$$

que

$$\lambda(OE^c) = \lambda\Big(\bigcup_{m\in\mathbb{N}}\bigcap_{n\in\mathbb{N}} O_m E_n^c\Big) \leq \sum_{m\in\mathbb{N}} \lambda(O_m E_m^c) < \epsilon.$$

Démonstration du théorème.

On peut supposer que ϕ est croissante et, en posant $\phi(x) = \phi(a)$ si $x < a$ et $\phi(x) = \phi(b)$ si $x > b$, que l'intervalle $[a, b]$ est symétrique par rapport à l'origine (voir la figure 2.1). Considérons les dérivées de Dini de ϕ en un point $x \in]a, b[$:

$$D_-\phi(x) = \liminf_{h\downarrow 0} \frac{\phi(x-h)-\phi(x)}{-h} = \lambda_g$$

$$D^-\phi(x) = \limsup_{h\downarrow 0} \frac{\phi(x-h)-\phi(x)}{-h} = \Lambda_g$$

$$D_+\phi(x) = \liminf_{h\downarrow 0} \frac{\phi(x+h)-\phi(x)}{h} = \lambda_d$$

$$D^+\phi(x) = \limsup_{h\downarrow 0} \frac{\phi(x+h)-\phi(x)}{h} = \Lambda_d$$

Il s'agit de montrer que l'on a

$$D_-\phi(x) = D^-\phi(x) = D_+\phi(x) = D^+\phi(x) < +\infty \tag{2.0.5}$$

presque partout sur $]a, b[$. Il suffit en fait de montrer que l'on a :

$$1^\circ D^+\phi(x) \leq D_-\phi(x)$$

$$2^\circ D_-\phi(x) < \infty$$

presque partout sur $]a, b[$. On sait que : $D_-\phi(x) < D^-\phi(x)$ et $D_+\phi(x) < D^+\phi(x)$. En supposant que le 1° et le 2° sont démontrés, et si on considère la fonction

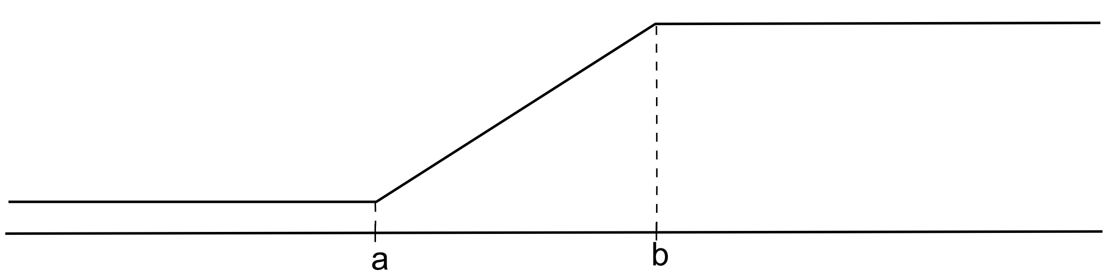

Fig. 2.1. Croissance de la fonction ϕ sur $[a, b]$.

croissante $\psi(x) = -\phi(-x)$, elle admet pour dérivées :

$$D^+\psi(-x) = \overline{\lim_{h\to 0^+}} \frac{\psi(-x+h) - \psi(-x)}{h} = \overline{\lim_{h\to 0^+}} \frac{\phi(x-h) - \phi(x)}{-h} = D^-\phi(x)$$

$$D_-\psi(-x) = \underline{\lim_{h\to 0^+}} \frac{\psi(-x-h) - \psi(-x)}{-h} = \underline{\lim_{h\to 0^+}} \frac{\phi(x+h) - \phi(x)}{h} = D_+\phi(x)$$

Grâce à 1°, on aura $D^+\psi(-x) \leq D_-\psi(-x)$, c'est-à-dire : $D^-\phi(x) \leq D_+\phi(x)$, puis que :

$$D_-\phi(x) \leq D^-\phi(x) \leq D_+\phi(x) \leq D^+\phi(x) \leq D_-\phi(x) < +\infty$$

presque partout sur $]a, b[$. Ce qui équivaudra à l'égalité (2.0.5) .

Puisqu'on peut toujours placer deux nombres rationnels s et t tels que $s > t$ entre deux nombres réels quelconques, on a alors :

$$\Big\{x/D^+\phi(x) > D_-\phi(x)\Big\} = \bigcup_{s,t\in\mathbb{Q}} \Big\{x/D^+\phi(x) > s > t > D_-\phi(x)\Big\}$$

et il s'agit de vérifier que, quels que soient $s, t \in \mathbb{Q}$,

$$\lambda(\Big\{x/D^+\phi(x) > s > t > D_-\phi(x)\Big\}) = 0.$$

Posons

$$E = \Big\{x/D^+\phi(x) > s > t > D_-\phi(x)\Big\} = \Big\{x/D^+\phi(x) > s\Big\} \cap \{x/D_-\phi(x) < t\}\,.$$

ϕ est croissante. Elle est donc mesurable. Donc $\dfrac{\phi(x-h) - \phi(x)}{-h}$ et $\dfrac{\phi(x+h) - \phi(x)}{h}$ le sont aussi. Donc $D_-\phi(x),\quad D^-\phi(x),\quad D_+\phi(x),\quad D^+\phi(x)$ sont mesurables comme lim inf et lim sup de fonctions mesurables. E est mesurable, car c'est l'intersection de deux ensembles mesurables.

Soit $O \subseteq \mathbb{R}$ un ensemble ouvert contenant E et tel que $\lambda(O) < \lambda(E) + \epsilon$. En effet, E est mesurable, donc il existe O ouvert tel que $\lambda(OE^c) < \epsilon$ (voir lemme 2.0.5 plus haut)

$$E \subset O$$

$$\lambda(O) = \lambda(OE) + \lambda(OE^c)$$

$$\lambda(O) < \lambda(OE) + \epsilon < \lambda(E) + \epsilon$$

Si $x \in E$,

$$D_-\phi(x) = \liminf_{h\downarrow 0} \frac{\phi(x-h) - \phi(x)}{-h} < t.$$

Donc pour chaque $x \in E$ et pour $h > 0$ arbitrairement petit, il existe un intervalle $[x-h, x] \subseteq O$ tel que

$$\phi(x) - \phi(x-h) < th.$$

E est alors recouvert par les $[x-h, x]$ pour chaque $x \in E$. En vertu du théorème de Vitali, on peut trouver

$$\{[x_1 - h_1, x_1], [x_2 - h_2, x_2], \ldots, [x_N - h_N]\}$$

disjoints tels que,

$$\lambda\left(E\left(\sum_{k=1}^{N}[x_k - h_k, x_k]\right)^c\right) < \epsilon.$$

En posant $D = \sum_{k=1}^{N}]x_k - h_k, x_k[$ qui est ouvert, donc mesurable, on aura d'une part

$$\lambda(ED^c) < \epsilon$$

et d'autre part,

$$\lambda(E) = \lambda(ED) + \lambda(ED^c) < \lambda(ED) + \epsilon$$

c'est-à-dire,

$$\lambda(ED) > \lambda(E) - \epsilon. \tag{2.0.6}$$

De plus, pour chaque k :

$$\phi(x_k) - \phi(x_k - h_k) < th_k \quad \text{entraîne} \quad \sum_{k=1}^{N}(\phi(x_k) - \phi(x_k - h_k)) < t\sum_{k=1}^{N} h_k < t\lambda(O) < t(\lambda(E) + \epsilon).$$

car

$$]x_k - h_k, x_k[\subseteq O \quad \text{entraîne} \quad D = \bigcup_k]x_k - h_k, x_k[\subseteq O, \quad \text{donc} \quad \lambda(D) = \sum_{k=1}^{N} h_k.$$

Si $y \in ED$,

$$D^+\phi(y) = \limsup_{h \downarrow 0} \frac{\phi(y+h) - \phi(y)}{h} > s.$$

Donc pour chaque $y \in ED$ et pour $H > 0$ arbitrairement petit, il existe un intervalle $[y, y+H] \subseteq D$ tel que

$$\phi(y+H) - \phi(y) > sH.$$

En vertu du théorème de Vitali, on peut trouver

$$\{[y_1, y_1 + H_1], [y_2, y_2 + H_2], \ldots, [y_M, y_M + H_M]\}$$

disjoints tels que, $\lambda\left(ED\left(\sum_{j=1}^{M}[y_j, y_j+H_j]\right)^c\right) < \epsilon$.

On pose

$$G = \sum_{j=1}^{M}]y_j, y_j + H_j[.$$

C'est un ouvert, donc mesurable. Nous avons que : $\lambda(EDG^c) < \epsilon$ et $\lambda(ED) = \lambda(EDG) + \lambda(EDG^c)$, puis :

$$\begin{aligned}
\lambda(ED) &< \lambda(EDG) + \epsilon \\
\lambda(EG) &> \lambda(EDG) \\
&> \lambda(ED) - \epsilon \\
&> \lambda(E) - 2\epsilon \qquad \text{(voir l'inégalité (2.0.6))}
\end{aligned}$$

De plus, pour chaque j :

$$\phi(y_j+H_j)-\phi(y_j) > sH_j \quad \text{implique} \quad \sum_{j=1}^{M}(\phi(y_j+H_j)-\phi(y_j)) > s\sum_{j=1}^{M} H_j > s(\lambda(E)-2\epsilon).$$

Maintenant remarquons que chaque intervalle $]y_j, y_j+H_j[$ doit être contenu dans un des intervalles $]x_k - h_k, x_k[$ (voir la figure 2.2). En regroupant les termes de la somme

$$\sum_{j=1}^{M}(\phi(y_j + H_j) - \phi(y_j))$$

suivant les intervalles $]x_k - h_k, x_k[$ auxquels ils correspondent et en utilisant le fait que la fonction ϕ est croissante, on obtient

$$\sum_{j=1}^{M}(\phi(y_j + H_j) - \phi(y_j)) \leq \sum_{k=1}^{N}(\phi(x_k) - \phi(x_k - h_k)).$$

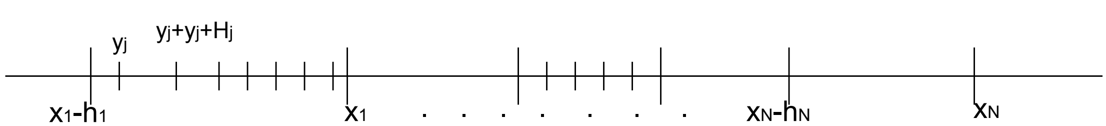

Fig. 2.2. Emboîtement de nouveaux intervalles dans les anciens.

Ceci entraîne

$$s(\lambda(E) - 2\epsilon) \leq t(\lambda(E) + \epsilon)$$

donc

$$s\lambda(E) \leq t\lambda(E)$$

et enfin

$$\lambda(E) = 0$$

.

Il ne reste plus qu'à voir que l'on a

$$\liminf_{h\to 0} \frac{\phi(x+h)-\phi(x)}{h} < +\infty$$

presque partout $x \in]a,b[$. Or le lemme de Fatou suggère que si on veut établir que $\underline{\lim}\, f_n < +\infty$ presque partout sur un ensemble E , il suffit de voir que $\underline{\lim} \int_E f_n < +\infty$.

$$\text{A-t-on} \qquad \lim_{n\to\infty} \int_a^b \frac{\phi(x+\frac{1}{n}-\phi(x)}{\frac{1}{n}} < +\infty \quad ?$$

On a

$$n\int_a^b (\phi(x+\frac{1}{n})-\phi(x))\,\mathrm{d}x \quad = \quad n\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} \phi(y)\,\mathrm{d}y - n\int_a^b \phi(y)\,\mathrm{d}y$$

On fait un changement de variables en posant : $y = x + \frac{1}{n}$ et on obtient :

$$\begin{aligned} n\int_a^b (\phi(x+\frac{1}{n})-\phi(x))\,\mathrm{d}x &= n\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} \phi(y)\,\mathrm{d}y - n\int_a^b \phi(y)\,\mathrm{d}y \\ &= n\int_{a+\frac{1}{n}}^{b} \phi(y)\,\mathrm{d}y + n\int_b^{b+\frac{1}{n}} \phi(y)\,\mathrm{dy} - n\int_a^{a+\frac{1}{n}} \phi(y)\,\mathrm{dy} \\ &\quad - n\int_{a+\frac{1}{n}}^{b} \phi(y)\,\mathrm{dy} \\ &= n\int_b^{b+\frac{1}{n}} \phi(y)\,\mathrm{dy} - n\int_a^{a+\frac{1}{n}} \phi(y)\,\mathrm{dy} \end{aligned}$$

Sachant que : $\phi(y) = \phi(b)$ pour $y \geq b$ et que pour : $a \leq y \leq a + \frac{1}{n} \quad \phi(a) \leq \phi(y)$, on en déduit :

$$\int_a^{a+\frac{1}{n}} \phi(a)\,\mathrm{dy} \leq \int_a^{a+\frac{1}{n}} \phi(y)\,\mathrm{dy},$$

alors

$$\begin{aligned} n\int_a^b (\phi(x+\frac{1}{n})-\phi(x))\,\mathrm{d}x \quad &= \quad \phi(b) - n\int_a^{a+\frac{1}{n}} \phi(y)\,\mathrm{dy} \\ &\leq \quad \phi(b)-\phi(a) \end{aligned}$$

$\square$

Chapitre 3

LA DÉMONSTRATION DE F.RIESZ

Nous allons démontrer la dérivabilité presque partout des fonctions monotones et cela sans nous reporter à la théorie de l'intégration. Nous précisons que nous nous sommes inspiré de Boas [RP] pour présenter cette démonstration.

Théorème 3.1. *Toute fonction monotone f(x) admet une dérivée finie et déterminée en tout point x, sauf peut-être aux points x d'un ensemble de mesure nulle, ou comme on dit encore, presque partout.*

Rappels utiles

Nous indiquerons le sens des expressions ci-dessus. Signalons que lorsque Lebesgue avait donné son théorème, ni l'idée de l'intégrale, ni celle de la mesure n'intervenaient dans l'énoncé du théorème. En effet, l'idée d'ensemble de mesure nulle ne dépend pas essentiellement de la théorie générale de la mesure, et les propriétés principales de ces ensembles s'établissent en quelques mots.

Par *ensemble de mesure nulle*, on entend d'après Lebesgue les ensembles des valeurs de x qu'on peut enfermer en un nombre fini ou une suite dénombrable d'intervalles de manière que leur longueur totale, c'est-à-dire la somme de leurs longueurs, soit arbitrairement petite. Il découle immédiatement de cette définition que tout sous-ensemble d'un tel ensemble est aussi de mesure nulle. Il en est de même quant à la réunion d'un nombre fini ou d'une suite dénombrable de tels ensembles ; en effet, on n'aura qu'à enfermer ces ensembles respectivement en des systèmes d'intervalles dont la longueur totale ne dépasse pas $\frac{\epsilon}{2^n}$; la longueur totale de tous ces intervalles, enfermant la réunion de nos intervalles, ne dépassera pas alors la quantité ϵ. En particulier, chaque ensemble fini ou dénombrable de valeurs x est de mesure nulle.

Lemme 3.0.6. *Soit E un sous-ensemble de $\mathbb{R}$ ayant la propriété suivante :*
Il existe un nombre positif q $(0 < q < 1)$ tel que, pour tout intervalle $]a, b[$, l'ensemble $E\cap]a, b[$ peut être recouvert par une suite dénombrable d'intervalles dont la longueur totale est au plus égale à $q(b-a)$. Alors E est de mesure nulle.(Intuitivement, ceci signifie qu'un ensemble qui couvre au plus une fraction de tout intervalle ne couvre presqu'aucun intervalle)

Preuve :
Il suffit de montrer que $E\cap]a, b[$ est de mesure nulle pour tout intervalle $]a, b[$, puisque $\mathbb{R}$ peut être recouvert par une suite dénombrable d'intervalles. On recouvre $E\cap]a, b[$ par des intervalles $]a_n, b_n[$ de longueur totale au plus égale à $q(b-a)$. De la même façon, on recouvre chaque $E\cap]a_n, b_n[$. Nous aurons recouvert $E\cap]a, b[$ par des intervalles de longueur totale au plus égale à $q(b_1 - a_1) + q(b_2-a_2)+\ldots = q[(b_1-a_1)+(b_2-a_2)+\ldots] \leq q^2(b-a)$. En répétant le processus n fois, on obtiendra une suite $\Sigma_1, \Sigma_2, \ldots$ de systèmes d'intervalles[1], chacun emboîté dans les précédents (voir la figure 3.1 à la page suivante) et l'on aura d'une façon générale :

$$\Sigma_{2n} \leqq q\Sigma_{2n-1}.$$

En effet, si $\Sigma_j << \Sigma_i$ signifie que le système d'intervalles Σ_j est enfermé dans le système Σ_i, on a :

$$\begin{array}{ccc}
\Sigma_2 << \Sigma_1 & : & \Sigma_2 \leq q\Sigma_1 \\
\Sigma_3 << \Sigma_2 & : & \Sigma_3 \leq \Sigma_2 \\
\Sigma_4 << \Sigma_3 & : & \Sigma_4 \leq q\Sigma_3 \leq q\Sigma_2 \leq q^2\Sigma_1 \\
 & \vdots & \\
\Sigma_{2n} & \leq & q^n\Sigma_1
\end{array}$$

Il s'ensuit que

$$\Sigma_{2n} \leqq q^n\Sigma_1 \longrightarrow 0.$$

[1] On convient que ces notations désignent aussi leurs longueurs totales correspondantes.

Donc $E\cap]a,b[$ pouvant être recouvert par des intervalles dont la longueur totale est arbitrairement petite, est de mesure nulle. En particulier :
$E\cap]-n,n[$ est de mesure nulle, donc

$$\bigcup_n (E\cap]-n,n[) = E\cap(\bigcup_n]-n,n[) = E\cap\mathbb{R} = E,$$

est de mesure nulle.CQFD

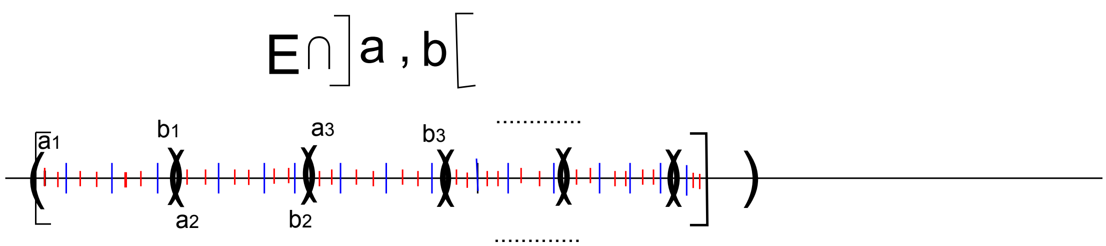

FIG. 3.1. Recouvrement de $E\cap]a,b[$ par des systèmes d'intervalles emboîtés.

La démonstration sera fondée sur le lemme de F.Riesz, connu sous l'appelation du lemme de «l'eau qui coule» ou le lemme du «soleil levant». Si g est une fonction continue d'un intervalle I dans $\mathbb{R}$, si le graphe de g est une section du lit d'un ruisseau dans le sens de la longueur, et si nous considérons l'ensemble E des points où l'eau coule, il est clair par l'intuition que E consiste en une réunion d'intervalles ouverts dont aux extrémités de chacun g a la même valeur. Si le graphe de g est une chaîne de montagnes vue de profil, si le soleil se lève dans la direction de l'axe [Ox), et si E est l'ensemble des points qui sont dans l'ombre, il est encore clair de façon intuitive que E sera la réunion d'intervalles ouverts pour lesquels g prend la même valeur aux extrémités de chacun (sauf dans le cas exceptionnel de l'intervalle situé à l'extrême gauche).

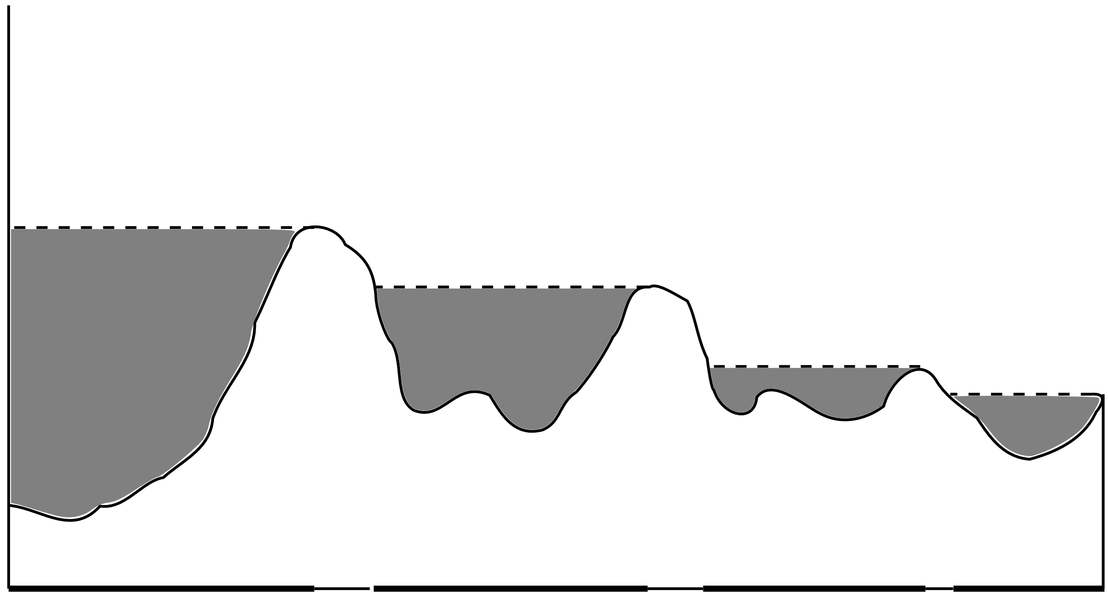

FIG. 3.2. Illustration du lemme du «soleil levant».

Rappels-Notations

On rappelle que pour les fonctions à valeurs dans $\mathbb{R}$, $\lim\limits_{x \to x_0} f(x)$ est définie comme la valeur commune de $\limsup_{x \to x_0} f(x)$ et de $\liminf_{x \to x_0} f(x)$. Si le domaine de f inclut un voisinage à droite de x_0, on définit $\limsup_{x \to x_0^+} f(x)$ ($\liminf_{x \to x_0^+} f(x)$) comme étant la limite supérieure (inférieure) de la restriction de f sur le voisinage à droite de x_0. La valeur commune (si elle existe) de ces limites à droite supérieure et inférieure est notée $\lim\limits_{x \to x_0^+} f(x)$, ou plus simplement $f(x_0^+)$. Pareillement, $f(x_0^-)$ sera la valeur commune (si elle existe) des limites à gauche de x_0 supérieure et inférieure de f.

Maintenant énonçons le lemme en des termes plus abstraits :

Lemme 3.0.7. *Soit g(x) une fonction continue sur un intervalle I, sauf à un ensemble au plus dénombrable de points, et soit G tel que :*

$$G(x) = max(g(x^-), g(x), g(x^+)).$$

L'ensemble E des points x intérieurs à I tels qu'il existe un ξ situé à droite de x, de sorte que $g(\xi) > G(x)$ est ouvert. Et si $]a_k, b_k[$ est l'un des intervalles ouverts et disjoints dont se compose E, alors $g(a_k^+) \leq G(b_k)$.

Remarque 3.0.1.

Si on fait une permutation entre x et ξ et si on définit E' comme l'ensemble des points x tels qu'il existe un ξ situé à gauche de x, de sorte que $g(\xi) > G(x)$, alors, de même, si E' est la réunion des intervalles ouverts $]a', b'[$, on a : $G(a') \geq g(b'^{-})$.

Démonstration.

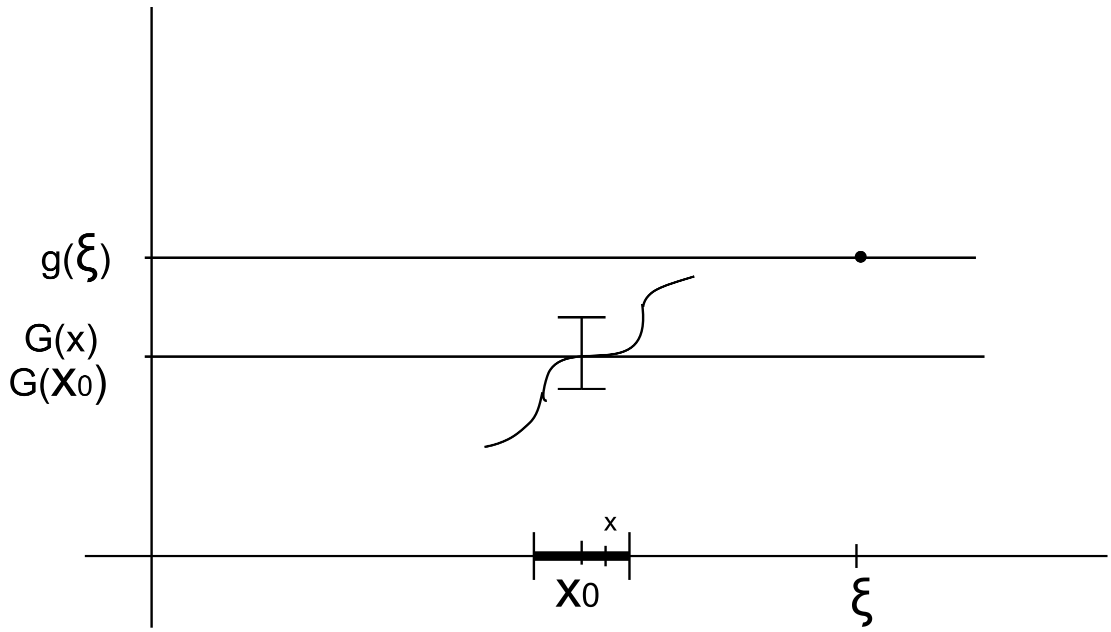

FIG. 3.3. $g(\xi) > G(x)$ si x est dans un voisinage de x_0.

Pour démontrer ce lemme, observons d'abord que l'ensemble E est ouvert. Soit $x_0 \in E$; alors il existe un ξ tel que $\xi > x_0$ avec $g(\xi) > G(x_0)$. Nous allons montrer que cette propriété reste valable pour tout x «proche» de x_0. Si x varie légèrement à gauche de x_0, alors $g(x_0)$ est proche de $g(x^{-})$; si x varie légèrement à droite de x_0, alors $g(x_0)$ est proche de $g(x^{+})$. Dans les deux cas, $G(x)$ ne peut excéder $G(x_0)$ que légèrement. Comme $g(\xi) > G(x_0)$, nous avons aussi $g(\xi) > G(x)$ tant que $G(x)$ n'est pas beaucoup plus grand que $G(x_0)$, ce qui, comme nous venons juste de voir, est le cas lorsque x est proche de x_0.

On démontre de la même façon dans le cas où $\xi < x_0$.

Prouvons maintenant la deuxième partie de la conclusion du lemme. Soit $]a_k, b_k[$ l'un des intervalles composant E ; alors le point b_k n'appartiendra pas à cet ensemble. Soit x un point intermédiaire entre a_k et b_k ; nous allons prouver que

$g(x) \leqq G(b_k)$; l'inégalité à démontrer s'ensuivra en faisant tendre x vers a_k par la droite.

Comme $x \in E$, il existe des points $\xi > x$ avec $g(\xi) > G(x)$, et ainsi avec $G(\xi) \geq g(x)$. Nous voulons montrer que b_k est l'un de ces points.

Supposons le contraire ; à savoir que $G(b_k) < g(x)$. Soit z_1 la plus petite borne supérieure des points z dans $[x, b_k]$ tels que $G(z) \geq g(x)$. Cet ensemble de points n'est pas vide car x en est un. Nous allons montrer que $z_1 = b_k$. Supposons que ce ne soit pas le cas, alors z_1 appartient à E et par conséquent il existe un $\xi > z_1$ avec $g(\xi) > G(z_1)$. Si ce ξ est dans $]z_1, b_k]$ on a : $G(\xi) < g(x) \leq G(z_1)$, ce qui contredit l'inégalité précédente. Donc $\xi > b_k$. Mais alors, puisque b_k n'est pas dans E, on a : $g(\xi) \leq G(b_k)$. Ainsi, si $G(b_k) < g(x)$, on a finalement :

$$g(x) \leq G(z_1) < g(\xi) \leq G(b_k) < g(x),$$

Ce qui aboutit à une contradiction. Donc $z_1 \geq b_k$: $z_1 = b_k$. CQFD

Bien évidemment la démonstration peut être faite de la même façon dans le cas de l'inégalité inversée ($\xi < x$) de l'hypothèse du lemme. $\square$

Lemmes (Conséquences du lemme de Riesz)

Lemme 3.0.8. *Soit f une fonction croissante sur $[a, b]$ et E_R l'ensemble des points x où f est continue et $\Lambda_d > R$ ($E_R = \{x \in [a, b]/f$ est continue en x et $\Lambda_d > R\}$). Alors E_R peut être recouvert par une suite dénombrable d'intervalles $]a_k, b_k[$ dont la longueur totale est au plus égale à :*

$$R^{-1} \sum [f(b_k^+) - f(a_k^+)] \leq [f(b^-) - f(a^+)]/R.$$

Preuve :

Si $x \in E_R$, il existe un $\xi > x$ tel que $\dfrac{f(\xi) - f(x)}{\xi - x} > R$, c'est-à-dire $f(\xi) - R\xi > f(x) - Rx$. On applique le lemme de Riesz à la fonction g telle que $g(x) = f(x) - Rx$. Comme f est continue en x, donc g l'est aussi, et $G(x) = g(x)$. On conclut que E_R est recouvert par une suite dénombrable d'intervalles ouverts $]a_k, b_k[$ tels que $G(b_k) \geq g(a_k^+)$, en d'autres termes

$$f(b_k^+) - Rb_k \geq f(a_k^+) - Ra_k \quad \text{(car f est croissante)}$$

c'est-à-dire

$$f(b_k^+) - f(a_k^+) \geq R(b_k - a_k)$$

En sommant cette inégalité sur tous les k, et en utilisant de nouveau le fait que f est croissante, on a :

$$R\sum(b_k - a_k) \leq \sum[f(b_k^+) - f(a_k^+)].$$

Lemme 3.0.9. *Soit f une fonction croissante sur $[a, b]$ et E_r l'ensemble des points x où f est continue et $\lambda_g < r$ ($E_r = \{x \in [a, b]/f \quad$ est continue en x et $\lambda_g < r\}$). Alors E_r peut-être recouvert par une suite dénombrable d'intervalles $]a_k, b_k[$ tels que :*

$$\sum[f(b_k^-) - f(a_k^+)] \leq r\sum(b_k - a_k) \leq r(b - a).$$

Preuve :

Si $x \in E$, il existe un $\xi < x$ tel que $\dfrac{f(\xi) - f(x)}{\xi - x} < r$, c'est-à-dire (puisque $\xi < x$)

$$f(\xi) - f(x) > (\xi - x)r$$
$$f(\xi) - r\xi > f(x) - rx$$

Le lemme de Riesz, dans le cas de l'inégalité inversée dans l'hypothèse, peut être appliqué à la fonction g définie par $g(x) = f(x) - rx$. On conclut que E_r est recouvert par un ensemble dénombrable d'intervalles $]a_k, b_k[$ tels que $g(b_k^-) \leq G(a_k)$; puisque f est croissante, l'existence de ce recouvrement implique que :

$$f(b_k^-) - rb_k \leq f(a_k^+) - ra_k$$

c'est-à-dire

$$f(b_k^-) - f(a_k^+) \leq r(b_k - a_k)$$

finalement

$$\sum[f(b_k^-) - f(a_k^+)] \leq r\sum(b_k - a_k).$$

Démonstration du théorème.

Soit $f(x)$ une fonction monotone pour $a \leqq x \leqq b$; pour fixer les idées, nous la supposons croissante. Pour examiner la dérivabilité de $f(x)$, nous allons comparer ses nombres dérivés. Rappelons que pour les *nombres dérivés* supérieur et inférieur $\Lambda_d, \lambda_d, \Lambda_g, \lambda_g$, des valeurs infinies sont admises. Une dérivée finie et déterminée existe en tout point x où les quatre nombres dérivés ont la même valeur finie.

Pour démontrer le théorème de Lebesgue, nous utilisons le même raisonnement qu'à la démonstration «standard» figurant aux pages 17 et 18. C'est-à-dire que nous aurons à prouver que l'on a presque partout :

$$1^\circ \quad \Lambda_d < \infty \qquad ; \quad 2^\circ \quad \Lambda_d \leq \lambda_g$$

En effet, en appliquant 2° à la fonction $f_0(x) = -f(-x)$, il résulte que l'on a aussi presque partout :

$$\Lambda_g \leq \lambda_d$$

et en combinant cela avec 1° et 2°, on obtient :

$$\Lambda_d \leq \lambda_g \leq \Lambda_g \leq \lambda_d \leq \Lambda_d < \infty$$

Autrement dit $\Lambda_d = \lambda_g = \Lambda_g = \lambda_d = \Lambda_d < \infty$.

Pour y parvenir, il suffit de montrer que l'ensemble

$$E_{rR} = \{x \in [a, b] / \lambda_g(x) < r < R < \Lambda_d(x)\}$$

est de mesure nulle, quelque soient r et R. Et l'on peut même supposer que r et R sont des nombres rationnels, pour la simple raison que, entre deux nombres réels différents, on peut toujours intercaler deux nombres rationnels. C'est-à-dire que, en formant les ensembles E_{rR} pour tous les couples rationnels, leur réunion E^* contiendra tous les x pour lesquels $\Lambda_d > \lambda_g$. Mais d'autre part, il n'y a qu'une infinité dénombrable de couples rationnels ; donc l'ensemble E^* est la réunion d'une infinité dénombrable d'ensembles de mesure nulle

$$E^* = \{x / \Lambda_d(x) > \lambda_g(x)\} = \bigcup_n E_{r_n R_n}$$

et par conséquent E^* et, à plus forte raison, l'ensemble envisagé qui est compris, seront de mesure nulle.

Comme première étape, nous observons que le lemme 3.0.8 implique que $\Lambda_d < \infty$ presque partout. Car, si $\Lambda_d = \infty$ sur E les hypothèses du lemme 3.0.8 sont réunies pour tout R positif, de sorte que E est recouvert par des intervalles $]a_k, b_k[$ dont la longueur totale est au plus égale à $[f(b^-) - f(a^+)]/R$ pour tout R. C'est-à-dire, E est recouvert par des intervalles de longueur totale arbitrairement petite ; d'où E est de mesure nulle.

Ensuite considérons un ensemble E où f est continue et $\lambda_g < r < R < \Lambda_d$; les hypothèses du lemme 3.0.8 et du lemme 3.0.9 étant toutes satisfaites à la fois. Appliquons le lemme 3.0.8 à chaque intervalle $]a_k, b_k[$ obtenu par le lemme 3.0.9. On conclut que $E\cap]a_k, b_k[$ est recouvert par des intervalles dont la longueur totale L_k est au plus égale à $[f(b_k^-) - f(a_k^+)]/R$. En sommant cette inégalité sur tous les k, et en appliquant lemme 3.0.9, on obtient :

$$\sum L_k \leq (1/R) \sum [f(b_k^-) - f(a_k^+)] \leq (r/R)(b-a).$$

Le même raisonnement s'applique à tout sous-intervalle de $]a, b[$; c'est-à-dire que l'intersection $E\cap]p, q[$ est recouvert par des intervalles de longueur totale au plus égale à $(r/R)(q-p)$.

Maintenant soit E recouvert par des sous-intervalles $]p_k, q_k[$ imbriqués ou non. $E\cap]p_k, q_k[$ peut être recouvert par des intervalles non imbriqués de longueur totale au plus égale à $(r/R)(q_k - p_k)$, et donc E peut être recouvert par des intervalles dont la longueur totale est au plus égale à $(r/R)\sum(q_k - p_k)$. Par le lemme 3.0.6, l'existence de tels recouvrements implique que E est de mesure nulle.

$\square$

Chapitre 4

LA DÉMONSTRATION DE M.W.BOTSKO

Le but dans cette partie est de fournir une démonstration alternative tout à fait élémentaire du théorème de Lebesgue. Elle utilise les dérivées inférieure et supérieure à la place des dérivées de Dini, mais n'exige pas que celles-ci soient mesurables. Pour notre preuve, nous aurons besoin du lemme suivant :

Lemme 4.0.10. *Soient $f : [a, b] \to \mathbb{R}$ une fonction, $P = \{x_0, x_1, x_2, \ldots, x_n\}$ une partition de $[a, b]$, S un sous-ensemble non vide de $\{1, 2, 3, \ldots, n\}$, et soit $A > 0$. Si $f(a) \leq f(b)$ et*

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} < -A$$

pour chaque k dans S, alors

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| > |f(b) - f(a)| + AL,$$

où $L = \sum_{k \in S}(x_k - x_{k-1})$.

Démonstration.

Puisque $f(a) \leq f(b)$, il s'ensuit que

$$\begin{aligned}
|f(b) - f(a)| = f(b) - f(a) &= \sum_{k=1}^{n}(f(x_k) - f(x_{k-1})) \\
&= \sum_{k \in S}(f(x_k) - f(x_{k-1})) + \sum_{k \notin S}(f(x_k) - f(x_{k-1})) \\
&< -A \sum_{k \in S}(x_k - x_{k-1}) + \sum_{k \notin S}(f(x_k) - f(x_{k-1})) \\
&\leq -AL + \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|
\end{aligned}$$

Par conséquent,

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| > |f(b) - f(a)| + AL,$$

Ce qui termine la preuve. □

Remarque 4.0.2.

En posant $g = -f$ il est facile de s'apercevoir que la conclusion du lemme tient si $f(a) \geq f(b)$ et

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} > A$$

pour tout k dans S.

Pour que notre preuve ne dépende pas de la théorie de la mesure des ensembles de mesure supérieure à zéro, nous aurons besoin du lemme de recouvrement élémentaire qui suivra. Sa preuve repose uniquement sur le théorème de Heine-Borel et le fait qu'un ouvert peut s'écrire comme réunion dénombrable d'intervalles ouverts disjoints. Bien entendu, ces deux résultats sont établis dans un cours d'analyse élémentaire.

Rappels-Lemmes sans démonstrations :

Théorème de Heine-Borel : *De tout recouvrement ouvert d'un intervalle fermé borné on peut extraire un sous-recouvrement fini.*

Tout sous-ensemble ouvert de $\mathbb{R}$ peut s'écrire comme réunion disjointe finie ou dénombrable d'intervalles ouverts.

Lemme 4.0.11. *Soit E un sous-ensemble de $]a, b[$ qui n'est pas de mesure nulle (c'est-à-dire qu'il existe un $\epsilon_0 > 0$ tel que $\sum_{n=1}^{\infty} \lambda(J_n) \geq \epsilon_0$ pour toute suite $\{J_n\}$ d'intervalles ouverts qui couvrent E). Si $\mathfrak{F}$ est une famille de sous-intervalles ouverts de $[a, b]$ qui couvrent E, alors il existe une sous-famille finie d'intervalles deux à deux disjoints $\{I_1, I_2, \ldots, I_N\}$ de $\mathfrak{F}$ tel que $\sum_{k=1}^{N} \lambda(I_k) > \frac{\epsilon_0}{3}$.*

Démonstration.

Puisque $\bigcup_{I \in \mathfrak{F}} I$ est un ouvert, il existe une suite d'intervalles ouverts deux à deux disjoints $\{]a_n, b_n[\}$ tels que $\bigcup_{I \in \mathfrak{F}} = \bigcup_{n=1}^{\infty}]a_n, b_n[$. Puisque $E \subseteq \bigcup_{n=1}^{\infty}]a_n, b_n[$,

$$\sum_{n=1}^{\infty} (b_n - a_n) \geq \epsilon_0 \tag{4.0.7}$$

Maintenant pour chaque $]a_n, b_n[$, on choisit un sous-intervalle fermé $[a'_n, b'_n]$, tel que $b'_n - a'_n = \dfrac{3(b_n - a_n)}{4}$. Après sommation et en utilisant (4.0.7) on obtient :

$$\sum_{n=0}^{\infty} b'_n - a'_n = \frac{3}{4}\sum_{n=0}^{\infty}(b_n - a_n) \geq \frac{3\epsilon_0}{4}$$

c'est-à-dire

$$\sum_{n=1}^{\infty}(b'_n - a'_n) \geq \frac{3\epsilon_0}{4}.$$

Soit n un entier positif fixé. Pour chaque x dans $[a'_n, b'_n]$, il existe un J_x dans $\mathfrak{F}$ tel que $x \in J_x \subseteq]a_n, b_n[$. Alors $\{J_x : x \in [a'_n, b'_n]\}$ est un recouvrement ouvert de $[a'_n, b'_n]$.

Grâce au théorème de Heine-Borel, il existe un nombre fini d'ensembles $J_1, J_2, \ldots, J_p$ de ce recouvrement ouvert tel que $[a'_n, b'_n] \subseteq \bigcup_{k=1}^{p} J_k$. Par ailleurs, on peut supposer en se débarrassant de quelques-uns de ces intervalles si nécessaire, qu'aucun intervalle de $\{J_k\}_{k=1}^{p}$ n'est un sous-ensemble de la réunion des autres intervalles de $\{J_k\}_{k=1}^{p}$.

Ainsi chaque J_i contient un point x_i qui n'appartient pas à $\bigcup_{k \neq i} J_k$ et, en renumérotant les J_k si nécessaire, on peut supposer que $x_1 < x_2 < x_3 < \ldots < x_p$.

Par conséquent $\{J_1, J_3, J_5, \ldots\}$ et $\{J_2, J_4, J_6, \ldots\}$ sont toutes deux des sous-familles finies d'éléments disjoints deux à deux de $\mathfrak{F}$.

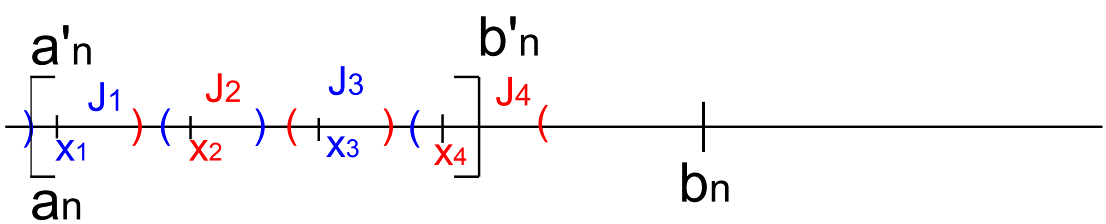

Fig. 4.1. Deux sous-familles finies d'intervalles disjoints deux à deux.

De toute évidence, soit $\sum_k \lambda(J_{2k-1}) \geq \dfrac{(\sum_{k=1}^{p}\lambda(J_k))}{2}$ ou $\sum_k \lambda(J_{2k}) \geq \dfrac{(\sum_{k=1}^{p}\lambda(J_k))}{2}$. Donc, indépendamment de laquelle des deux inégalités est vraie, nous avons trouvé une sous-famille finie d'intervalles disjoints deux à deux $\mathfrak{F}_n$ de $\mathfrak{F}$ telle que

$$\sum_{I\in\mathfrak{F}_n} \lambda(I) \geq \frac{\left(\sum_{k=1}^{p} \lambda(J_k)\right)}{2} \geq \frac{(b'_n - a'_n)}{2}$$

Il s'ensuit que pour chaque entier positif n il existe une sous-famille finie d'intervalles disjoints deux à deux $\mathfrak{F}_n$ de $\mathfrak{F}$ dont chacun des intervalles ouverts est un sous-ensemble de $]a_n, b_n[$ tel que $\sum_{I\in\mathfrak{F}_n} \lambda(I) \geq \dfrac{(b'_n - a'_n)}{2}$. En sommant les deux côtés de cette inégalité, on obtient

$$\sum_{n=1}^{\infty} \sum_{I\in\mathfrak{F}_n} \lambda(I) \geq \frac{\left(\sum_{n=1}^{\infty} (b'_n - a'_n)\right)}{2} \geq \frac{3\epsilon_0}{8}.$$

Par suite $\bigcup_{n=1}^{\infty} \mathfrak{F}_n = \{I_1, I_2, I_3, \ldots\}$ est une sous-famille dénombrable d'intervalles deux à deux disjoints de $\mathfrak{F}$ pour laquelle

$$\sum_{n=1}^{\infty} \lambda(I_n) \geq \frac{3\epsilon_0}{8}.$$

Pour terminer la preuve, on choisit N assez grand pour que $\displaystyle\sum_{k=1}^{N} \lambda(I_k) > \frac{\epsilon_0}{3}$ $\square$

De plus, nous aurons besoin de la variante facile du lemme 2 suivante :

Remarque 4.0.3.

Supposons E et ϵ_0 exactement comme dans le lemme 2. Soit P un sous-ensemble fini de $[a, b]$, et soit $\mathfrak{F}$ une famille de sous-intervalles ouverts de $[a, b]$ qui couvrent $E - P$. Alors en ajoutant à $\mathfrak{F}$ des intervalles ouverts suffisamment petits centrés en chaque point de P et en utilisant le lemme 2, il est facile de montrer qu'il existe une sous-famille finie d'intervalles deux à deux disjoints $\{I_1, I_2, \ldots, I_N\}$ de $\mathfrak{F}$ tel que $\displaystyle\sum_{k=1}^{N} \lambda(I_k) > \frac{\epsilon_0}{4}$.

Lemme 4.0.12. *L'ensemble des points de discontinuité d'une fonction monotone f est au plus dénombrable.*

Preuve :

Pour fixer les idées, on va supposer f croissante.

Soit $\bar{x}$ un point de discontinuité de f.

Soit $\{x_n\}_n$ une suite croissante qui converge vers $\bar{x}$, cela implique que $\lim f(x_n)$ existe ; limite que nous notons A . Car $\{x_n\}$ étant convergente, elle est donc bornée. Donc, $\{f(x_n)\}_n$ l'est aussi. Cette suite étant croissante et bornée, elle est donc convergente.

Soit $\{y_n\}_n$ une autre suite croissante qui converge vers $\bar{x}$, cela implique que $\lim f(y_n)$ existe ; limite que nous notons B

À chaque x_n correspond un y_{k_n} tel que $y_{k_n} > x_n$, donc $B \geq A$.

À chaque y_{k_n} correspond un $x_{n_{k_n}}$ tel que $x_{n_{k_n}} > y_{k_n}$, donc $A \geq B$

Donc A=B. C'est-à-dire que cette limite est unique ; c'est $\lim_{x \to \bar{x}^-} f(x) = f(\bar{x}^-)$. .

De même, pour toute suite $\{x_n\}_n$ décroissante qui converge vers $\bar{x}$, on montre que $\{f(x_n)\}_n$ converge vers $f(\bar{x}^+)$, ce qui veut dire que : $\lim_{x \to \bar{x}^+} f(\bar{x})$ existe$=f(\bar{x}^+)$

Comme f est croissante alors : $f(\bar{x}^-) < f(\bar{x}^+)$. C'est-à -dire qu'on a un intervalle ouvert sur l'axe des ordonnées correspondant à $\bar{x}$. On fait la même chose avec un autre point de discontinuité de f, $\bar{y}$, ainsi de suite (voir la figure 4.2 à la page suivante). À la fin, nous aurons une suite d'intervalles ouverts disjoints sur l'axe des ordonnées chacun contenant un nombre rationnel différent. Par conséquent, on aura une application bijective entre l'ensemble de ces intervalles (donc l'ensemble des points de discontinuité) et un sous-ensemble de $\mathbb{Q}$, qui est dénombrable. CQFD

Nous sommes maintenant préparés pour établir le résultat principal, énoncé comme suit :

Théorème 4.1. *Si f est une fonction croissante sur $[a, b]$, alors $f'(x)$ existe presque partout sur $[a, b]$.*

Démonstration du théorème.

Puisque f est continue sauf sur un ensemble dénombrable de points, il suffit de montrer que

$$F = \{x : x \in (a, b), f \text{ est continue en } x \text{ et } \overline{D}f(x) > \underline{D}f(x)\}$$

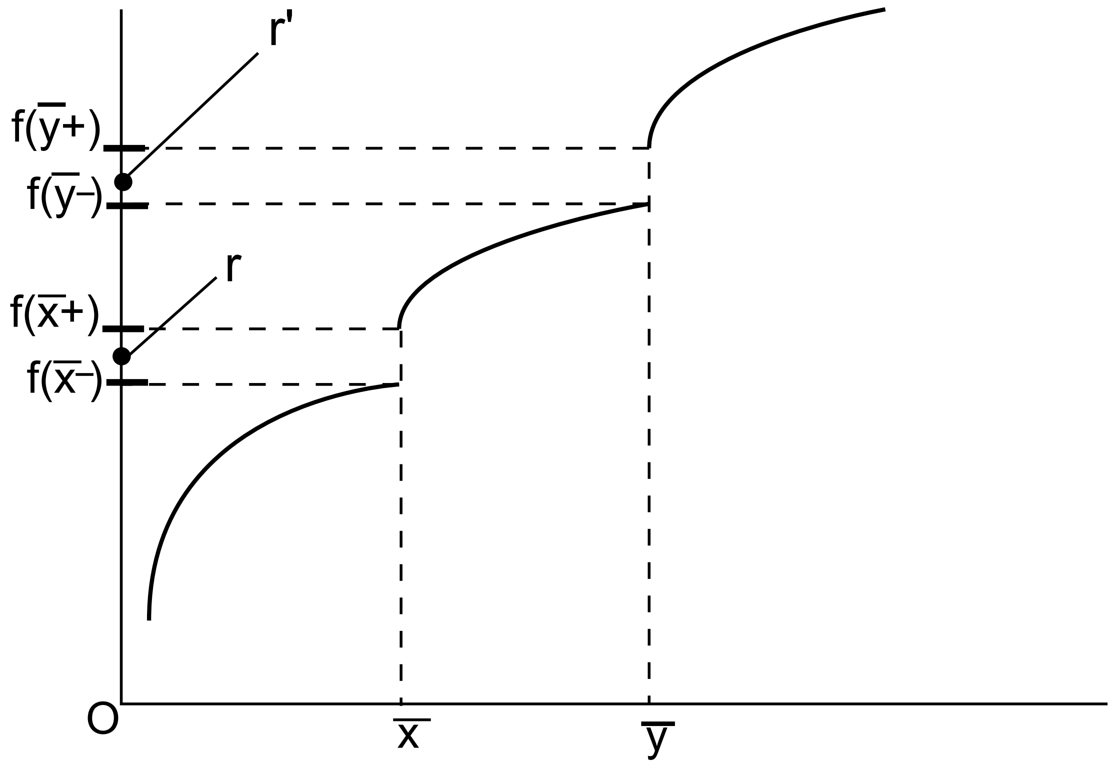

FIG. 4.2. Points de discontinuité d'une fonction monotone.

est de mesure nulle. Ici

$$\overline{D}f(x) = \limsup_{y\to x} \frac{f(y)-f(x)}{y-x},$$

$$\underline{D}f(x) = \liminf_{y\to x} \frac{f(y)-f(x)}{y-x}$$

sont respectivement les dérivées supérieure et inférieure de f en x. Clairement, F est une réunion dénombrable d'ensembles de la forme

$$E_{r,s} = \{x : x \in (a,b), f \text{ est continue en } x, \text{et } \overline{D}f(x) > r > s > \underline{D}f(x)\}$$

pour tous nombres rationnels r et s tels que $r > s$. Nous devons donc montrer simplement que chaque ensemble $E_{r,s}$ est de mesure nulle.

Supposons au contraire que pour un choix de r et s l'ensemble $E = E_{r,s}$ n'est pas de mesure nulle, et soit ϵ_0 un nombre positif comme dans le lemme 2. Si $A = \dfrac{(r-s)}{2}$, $B = \dfrac{r+s}{2}$ et $g = f - Bx$, alors il est clair que A et B sont tous les deux positifs et

$$E = \{x : x \in (a,b), g \text{ est continue en } x, \overline{D}g(x) > A, \text{et } \underline{D}g(x) < -A\}$$

Comme :

$$\begin{aligned}\sum_P |g(x_k) - g(x_{k-1}| &= \sum_P |f(x_k) - Bx_k - f(x_{k-1}) + Bx_{k-1}| \\ &= \sum_P |f(x_k) - f(x_{k-1}) + B(x_{k-1} - x_k)| \\ &\leq \sum_P |f(x_k) - f(x_{k-1})| + B\sum_P |x_k - x_{k-1}| \\ &= \sum_P (f(x_k) - f(x_{k-1}) + B\sum_P (x_k - x_{k-1}) \\ &\leq f(b) - f(a) + B(b-a) \quad \text{car } f \text{ est croissante.}\end{aligned}$$

Donc $\{\sum_P |g(x_k) - g(x_{k-1})| : P \quad \text{est une partition de } [a,b]\}$ est borné supérieurement par $f(b) - f(a) + B(b-a)$.

Nous pouvons appeler T la plus petite borne supérieure de cet ensemble. Puisque A et ϵ_0 sont tous deux positifs, il existe une partition $P = \{x_0, x_1, x_2, \ldots, x_n\}$ de $[a,b]$ pour laquelle

$$\sum_{k=1}^{n} |g(x_k) - g(x_{k-1})| > T - \frac{A\epsilon_0}{4} \tag{4.0.8}$$

Maintenant soit x appartenant à $E - P$, ce qui signifie que x appartient à $E \cap]x_{k-1}, x_k[$ pour un certain k. Comme $\underline{D}g(x) < -A, \overline{D}g(x) > A$, et g est continue en x, nous pouvons choisir a_x et b_x tel que $a_x < x < b_x,]a_x, b_x[\subseteq]x_{k-1}, x_k[$, et $\frac{g(x_k)-g(x_{k-1})}{x_k - x_{k-1}} < -A$ ou $> A$ selon que $g(x_{k-1}) \leq g(x_k)$ ou $g(x_{k-1}) \geq g(x_k)$. Alors $\mathfrak{F} = \{]a_x, b_x[: x \in E - P\}$ est une famille de sous-intervalles ouverts de $[a,b]$ qui couvrent $E - P$. Par la remarque 2, il existe une sous-famille finie de sous-intervalles disjoints deux à deux $\{I_1, I_2, \ldots, I_N\}$ de $\mathfrak{F}$ telle que

$$\sum_{k=1}^{N} \lambda(I_k) > \frac{\epsilon_0}{4} \tag{4.0.9}$$

Maintenant soit $Q = \{y_0, y_1, y_2, \ldots, y_p\}$ une partition de $[a,b]$ déterminée par les points de P et les extrémités des intervalles $I_1, I_2, \ldots, I_N$. Pour chaque $[x_{k-1}, x_k]$ contenant au moins un des intervalles de $\{I_1, I_2, \ldots, I_N\}$, nous déduisons du lemme 1 ou de la remarque 1 que

$$\sum_{[y_{i-1}, y_i] \subseteq [x_{k-1}, x_k]} |g(y_i) - g(y_{i-1})| > |g(x_k) - g(x_{k-1})| + AL_k \tag{4.0.10}$$

où la sommation est faite sur les intervalles fermés déterminés par les points de Q qui sont contenus dans $[x_{k-1}, x_k]$ et L_k est la somme des longueurs des intervalles $I_1, I_2, \ldots, I_N$ qui sont contenus dans $[x_{k-1}, x_k]$. En sommant l'inégalité (4.0.10) sur k et en utilisant (4.0.8) et (4.0.9), nous obtenons

$$\begin{aligned}\sum_{k=1}^{q} |g(y_k) - g(y_{k-1})| &> \sum_{k=1}^{n} |g(x_k) - g(x_{k-1})| + A \sum_{k=1}^{N} \lambda(I_k) \\ &> T - \frac{A\epsilon_0}{4} + A\Big(\frac{\epsilon_0}{4}\Big) = T\end{aligned}$$

Ce qui contredit la définition de T et termine la première partie de la preuve.

$\square$

Pour être tout à fait complet en rapport avec notre démonstration, il ne reste plus qu'à faire la preuve du théorème suivant :

Théorème 4.2. *Si f est une fonction croissante sur $[a, b]$, alors $\underline{D}f(x)$ est finie presque partout sur $[a, b]$.*

Démonstration.

Comme $\underline{D}f(x) \geq 0$ pour tout x dans $]a, b[$, il suffit de prouver que $\underline{D}f(x) < \infty$ presque partout sur $]a, b[$. Pour y parvenir, nous montrerons que $E = \{x : x \in]a, b[\quad \text{et} \quad \underline{D}f(x) = \infty\}$ est de mesure nulle.

Supposons le contraire. Soit M un nombre positif arbitrairement grand, et soit $\epsilon_0 > 0$ tel que défini dans le lemme 2. Si x appartient à E, alors $\underline{D}f(x) > M$ et il existe a_x et b_x tel que $a_x < x < b_x$, $]a_x, b_x[\subseteq]a, b[$ et

$$\frac{f(b_x) - f(a_x)}{b_x - a_x} > M.$$

Donc l'ensemble $\mathfrak{F} = \{]a_x, b_x[: x \in E\}$ couvre E et par le lemme 2 il contient une sous-famille disjointe $\{I_1, I_2, \ldots, I_N\}$ tel que $\sum_{k=1}^{N} |I_k| > \frac{\epsilon_0}{3}$. Soit $]a_k, b_k[$ pour chaque k.

Puisque f est croissante, clairement

$$f(b) - f(a) \geq \sum_{k=1}^{N} (f(b_k) - f(a_k)) > \sum_{k=1}^{N} M(b_k - a_k) > M\frac{\epsilon_0}{3}.$$

Par conséquent $f(b)-f(a) > M\frac{\epsilon_0}{3}$. Vu que $\epsilon_0 > 0$ et M est arbitrairement grand, ceci contredit la finitude de $f(b) - f(a)$. $\square$

CONCLUSION

La démonstration «standard» et celle de F.Riesz démarrent sur le même principe, mais utilisent des moyens différents pour parvenir à la même conclusion. La démonstration de Botsko, par contre, se démarque complètement de celles qui la précèdent et, au regard de son élégance et de sa simplicité, peut être présentée dans un cours d'analyse élémentaire. D'autres variantes existent telle que celle géométrique de D.Austin . Et ceci ne fait justement qu'illustrer la richesse et la beauté de la science mathématique. Ainsi, selon les besoins et les objectifs de tout un chacun, on peut toujours puiser dans l'une ou l'autre des démonstrations de ce grand théorème de Lebesgue.

BIBLIOGRAPHIE

[RR] Robert Ryan, *L'intégrale de Lebesgue a cent ans* Pour la science n˚283, mai 2001.

[HL] Henri Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthiers-Villars, Paris, 1928.

[MA] Notes cours, *http ://www.dms.umontreal.ca/∼giroux/mesure.html*,

[RN] F. Riesz and B. Sz. Nagy, *Leçons d'analyse fonctionnelle*, Gauthier-Villars, Paris, 1968.

[MB] Michael W.Botsko, *An elementary Proof of Lebesgue's Differentiation Theorem*, Math.Monthly 110,2003.

[RP] Ralph Boas *A prime of real functions* -MAA

[CJ] Camille Jordan, *Cours d'analyse de l'École Polytechnique* Gauthiers-Villars, Paris, 1959

[KF] L. Kronecker and L. Fuchs, *G. Lejeune-Dirichlets's Werke*, Bronx, N.Y., Chelsea, 1969.

[DA] D. Austin, *A geometric proof of the Lebesgue differentiation theorem*, Proc.Amer.Math.Soc. 16, 1965.

[Leb] Wikipédia - Lebesgue, `http ://www.wikipedia.org/Lebesgue`

[Jor] Wikipédia-Jordan, `http ://www.wikipedia.org/Camille`

[Rie] Wikipédia - Riesz, `http ://www.wikipedia.org/Frigyes`

Annexe A

LEBESGUE

Fig. A.1. Henri-Léon Lebesgue (28 juin 1875 à Beauvais(France) - 26 juillet 1941 à Paris(France)).

Henri-Léon Lebesgue est un mathématicien français. Il est reconnu pour sa théorie de l'intégration publiée initialement dans sa dissertation Intégrale, longueur, aire à l'université de Nancy en 1902. Il fut l'un des grands mathématiciens français de la première moitié du XXe siècle. Son père qui était ouvrier dans une imprimerie et ses deux soeurs aînées moururent de tuberculose alors qu'il avait trois ans. Ensuite, sa mère a travaillé très dur pour qu'il puisse faire des études. Élève brillant dès l'école élémentaire, Lebesgue étudia plus tard à l'École normale supérieure.

Il a enseigné au lycée de Nancy et à celui de Rennes. Il se fera alors connaître par sa théorie de la mesure, laquelle prolonge les premiers travaux importants d'Émile Borel, l'un de ses professeurs et plus tard son ami.

Il mit au point une théorie des fonctions mesurables (1901) en se fondant sur les résultats d'Émile Borel : les tribus boréliennes.

Henri Léon Lebesgue a révolutionné et généralisé le calcul intégral. Sa théorie de l'intégration (1902-1904) est extrêmement commode d'emploi, et répond aux besoins des physiciens. En effet, elle permet de rechercher et de prouver l'existence de primitives pour des fonctions « irrégulières » et recouvre différentes théories antérieures qui en sont des cas particuliers :

- fonctions en escalier et fonctions continues de Riemann
- fonctions bornées de Darboux
- fonctions à variation bornée de Stieltjes

On lui doit aussi la transformée de Fourier établie dans la fin des années 30.

Il est nommé professeur à la Sorbonne en 1910, puis au Collège de France en 1921. Il donne également des cours à l'École supérieure de physique et de chimie industrielles de la ville de Paris de 1927 à 1937 et à l'École normale supérieure de Sèvres. Il sera élu à l'Académie des sciences en 1922.

Comme son père, Henri Léon Lebesgue a eu une santé déficiente tout au long de sa vie. Il se maria avec la soeur d'un de ses camarades de l'École normale supérieure, et eut deux enfants, Suzanne et Jacques.

Annexe B

JORDAN

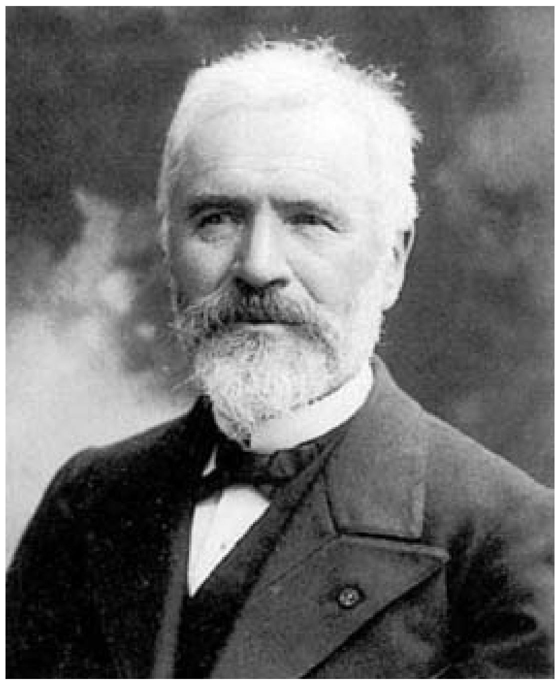

Fig. B.1. Camille Jordan(5 janvier 1838 à Lyon(France)-22 janvier 1922 à Paris(France)).

Son père Esprit-Alexandre Jordan (1800-1888) était polytechnicien (1818) et sa mère Joséphine était la soeur du peintre Pierre Puvis de Chavannes. Il étudia à l'École polytechnique (Promotion 1855). Il fut ingénieur au corps des mines puis plus tard, enseigna à l'École polytechnique et succéda à Joseph Liouville au Collège de France, où il avait une réputation de choix de notation excentriques.

Aujourd'hui on associe son nom à un certain nombre de résultats fondamentaux :

- le théorème de Jordan et la courbe de Jordan à laquelle ce théorème se réfère ;
- la forme normale de Jordan et la réduction de Jordan (parfois confondue avec les travaux de Wilhelm Jordan 1842- 1899 à qui l'on doit la méthode du pivot ou d'élimination de Gauss-Jordan) ;
- le théorème de Jordan-Hölder, qui est un résultat fondamental sur les groupes finis et les séries de compositions.
- la mesure de Jordan, qui préfigure la théorie de la mesure.

Camille Jordan a contribué à faire entrer la théorie de Galois dans le courant de pensée majoritaire. Il investiga aussi les groupes de Mathieu, premiers exemples de groupes sporadiques.

En 1919, il devient membre étranger de la Royal Society.

Officier de la Légion d'honneur.

L'université Lyon I a donné son nom à un institut de recherche en mathématiques.

Il était le petit-neveu de l'homme politique Camille Jordan.

Annexe C

RIESZ

Fig. C.1. Frédéric Riesz(22 juin 1880 à Györ(Hongrie)-28 février 1956 à Budapest(Hongrie)).

Frigyes Riesz (Friedrich en allemand et Frédéric en français), né le 22 juin 1880 à Györ et mort le 28 février 1956 à Budapest, est un mathématicien hongrois.

Il a étudié à Budapest, Göttingen et Zurich et fut diplômé en 1902 à Budapest. Il fut appelé en 1911 pour une chaire à l'université de Kolozsvár (en allemand Klausenburg, en Transylvanie). Comme Kolozsvár (aujourd'hui Cluj-Napoca, Roumanie) était devenue roumaine en 1920 avec la Paix du Trianon, l'université fut déplacée à Szeged. Riesz fonda avec Alfréd Haar en 1922 à Szeged l'Institut mathématique János-Bolyai . En 1945, il est rappelé à Budapest.

Riesz a publié en langues hongroise, allemande et française ; ses écrits étaient prisés pour leur extraordinaire clarté.

Riesz fut l'un des fondateurs de l'analyse fonctionnelle. Il prouva en 1907 le théorème aujourd'hui connu comme le théorème de Riesz-Fischer en analyse de Fourier dans les espaces de Hilbert, sur l'équivalence entre la mécanique matricielle et la mécanique ondulatoire.

Il est également à l'origine du théorème de compacité de Riesz qui fait le lien entre dimension et compacité des boules fermées dans un espace vectoriel normé. On lui doit aussi le théorème de représentation de Riesz qui établit une isométrie entre un espace de Hilbert et son dual

On donna son nom à une catégorie d'espaces vectoriels ordonnés en son honneur : les espaces de Riesz .

Le mathématicien Marcel Riesz était son frère cadet.