CHAPITRE I.

DÉRIVATION.

THÉORÈME DE LEBESGUE SUR LA DÉRIVÉE D'UNE FONCTION MONOTONE.

1. Exemple d'une fonction continue non dérivable.

En Analyse classique, on suppose généralement que les fonctions qu'on envisage admettent des dérivées, même des dérivées continues jusqu'à tel ou tel ordre. Quelquefois on ne recule pas devant quelques points d'exception. Mais jusqu'au commencement du présent siècle on ne s'est posé que rarement la question de savoir si les fonctions appartenant à telle et telle catégorie, comme par exemple les fonctions continues ou les fonctions monotones, admettent nécessairement des dérivées. Ou bien si elles n'en admettent pas partout, au moins à l'exception d'un ensemble dont la nature peut être précisée. On n'est arrivé dans cette direction qu'à quelques résultats presque évidents, comme par exemple la dérivabilité de gauche et de droite des fonctions convexes, ce qui entraîne l'existence de leur dérivée pour toute valeur de x, sauf au plus pour une infinité dénombrable de valeurs exceptionnelles.

On convient généralement de dater ces questions de l'année 1806, d'un mémoire "sur la théorie des fonctions dérivées" d'AMPÈRE [1][1]) où ce grand savant essayait sans succès d'établir la dérivabilité d'une fonction "quelconque" à l'exception de certaines valeurs "particulières et isolées" de la variable. D'ailleurs, en tenant compte de l'évolution de l'idée de fonction, on a le sentiment — bien que le texte original ne dise rien de positif sur ce point — que les efforts d'Ampère ne pouvaient guère être dirigés au delà des fonctions composées de traits monotones.

Pendant tout le XIX[e] siècle, quelque fécond que ce siècle ait été pour le développement de l'Analyse, la solution du problème n'a pas avancé, il semble même, au premier abord, qu'on s'en éloignât. En effet, le premier résultat essentiel, après plus d'un demi-siècle, est fourni par la critique de WEIERSTRASS[2]),

---

[1]) Les nombres ou les astérisques entre chrochets renvoient à la bibliographie placée à la fin du livre.

[2]) Publié par DU BOIS-REYMOND [1].

qui mit fin aux tentatives répétées d'établir la dérivabilité des fonctions continues du type le plus général et cela en construisant *une fonction continue sans dérivée.* Puis ces exemples se sont multipliés, on en a inventé de plus en plus simples, on s'est employé à mettre en évidence leurs rapports, mutuels et avec d'autres problèmes, oeuvre à laquelle on pris part presque tous les grands maîtres de l'Analyse de la seconde moitié du siècle et qui s'est continuée jusqu'à nos jours. En voici un, peut-être le plus élémentaire, dû à VAN DER WAERDEN [1] et fondé sur le fait évident qu'une suite infinité formée de nombres entiers ne peut être convergente que si ses termes restent égaux à partir d'un certain rang.

Convenons de dire, comme d'habitude, que la fonction $f(x)$ admet une dérivée au point x lorsque le rapport

$$\frac{f(x+h)-f(x)}{h}$$

tend, pour $h \to 0$ et $x+h$ parcourant des valeurs pour lesquelles $f(x+h)$ a un sens, vers une limite finie et déterminée. Désignons par $\{x\}$ la distance de x à l'entier le plus proche. Cela étant, formons la fonction

$$f(x)=\sum_{n=0}^{\infty}\frac{\{10^n x\}}{10^n}; \tag{1}$$

les termes de cette série représentant des fonctions continues et, de plus, la série étant majorée par la série géométrique $\Sigma 10^{-n}$, la fonction $f(x)$ est continue. Alors, en essayant de calculer la dérivée au point x, nous allons nous heurter à une contradiction.

Pour atteindre ce but, observons que l'on se peut borner évidemment au cas où $0 \leqq x < 1$ et écrivons x sous la forme

$$x=0, a_1 a_2 \cdots a_n \cdots$$

et cela en convenant de préférer, le cas échéant, l'écriture sous forme de fraction décimale finie, complétée par des zéros. Nous distinguons deux cas suivant que

$$0, a_{n+1} a_{n+2} \cdots \leqq \frac{1}{2} \text{ ou } > \frac{1}{2}.$$

Dans le premier cas

$$\{10^n x\}=0, a_{n+1} a_{n+2} \cdots,$$

tandis que dans le second

$$\{10^n x\}=1-0, a_{n+1} a_{n+2} \cdots.$$

Posons $h_m=-10^{-m}$ lorsque a_m est égal à 4 ou à 9 et $h_m=10^{-m}$ au cas contraire. Envisageons le rapport

$$\frac{f(x+h_m)-f(x)}{h_m}; \tag{2}$$

par la formule (1), ce rapport s'exprime par une série de la forme

$$10^m \sum_{n=0}^{\infty} \pm \frac{\{10^n(x \pm 10^{-m})\}-\{10^n x\}}{10^n}.$$

Or, il est manifeste que les numérateurs s'annulent à partir de $n=m$ et que d'autre part, pour $n<m$, ils se réduisent à $\pm 10^{n-m}$; de cette façon, les termes correspondants de notre expression donnent ± 1, et par conséquent la valeur du rapport (2) est un entier, positif ou non, mais en tout cas pair ou impair suivant la parité de m. Donc la suite des rapports (2), formée d'entiers de parité variable, ne peut point converger.

2. Théorème de Lebesgue sur la dérivée d'une fonction monotone. Ensembles de mesure nulle.

Passons à la classe des fonctions monotones. On doit à LEBESGUE le théorème suivant, un des plus frappants et des plus importants de l'Analyse réelle:

Théorème. *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.*

Nous indiquerons le sens de ces expressions tout à l'heure. Ajoutons que LEBESGUE a établi son théorème, d'ailleurs dans *l'hypothèse additionnelle de la continuité de $f(x)$*, en 1904, dans la première édition de son livre [*] sur l'intégration, à la fin du dernier chapitre, comme dernière conséquence de la théorie entière. Cependant, ni l'idée de l'intégrale, ni celle de la mesure n'interviennent 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 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 $\varepsilon/2^n$; la longueur totale de tous ces intervalles, enfermant la réunion de nos ensembles, ne dépassera pas alors la quantité ε. En particulier, chaque ensemble fini ou dénombrable de valeurs x est de mesure nulle.

Quelquefois il sera avantageux de présenter notre définition sous la forme suivante. L'ensemble E est de mesure nulle, si l'on peut l'enfermer dans une suite d'intervalles de longueur totale finie de sorte que tout point de E

soit intérieur à une infinité de ces intervalles. Les deux définitions sont équivalentes. En effet, la seconde implique la prémière, puisque quand tous les points de E appartiennent à une infinité d'intervalles de longueur totale finie, on pourra abaisser cette longueur totale à volonté en supprimant un nombre fini d'intervalles. Inversement, quand E est de mesure nulle d'après la première définition, on n'aura qu'à l'enfermer successivement en des systèmes d'intervalles dont la longueur totale reste respectivement inférieure à $1/2^n$, et s'il était nécessaire, d'élargir les intervalles à gauche et à droite par exemple en redoublant leurs longueurs; la réunion de tous ces systèmes satisfera alors aux exigences de la seconde définition.

Le terme "presque partout" (en abrévation: p. p.) est utilisé pour dire que le fait en question subsiste partout, sauf peut-être en tous les points d'un ensemble de mesure nulle.

Avant de démontrer le théorème fondamental de Lebesgue, montrons d'abord que, dans une certaine direction, il donne le maximum et ne pourra être perfectionné. En effet, étant donné un ensemble E de mesure nulle, nous allons construire une fonction croissante qui n'admet pas de dérivée finie en les points de E (à vrai dire, notre fonction admet en ces points une dérivée infinie). À cet effet, on n'aura qu'à enfermer E en des intervalles au sens de la seconde définition et à poser $f(x)$ égale à la somme des longueurs des intervalles ou des segments d'intervalles situés à gauche du point x; la fonction ainsi définie jouira évidemment de la propriété exigée.

3. Démonstration du théorème de Lebesgue.

Nous allons démontrer la dérivabilité presque partout des fonctions monotones et cela sans nous reporter à la théorie de l'intégration. Les premières démonstrations qui tiennent compte de pareille indépendance, sont dues à FABER [1] et G. C. YOUNG et W. H. YOUNG [1].

Pour la commodité du langage nous supposerons d'abord qu'il s'agisse d'une fonction continue et monotone et nous indiquerons seulement à la fin les modifications, d'ailleurs presque évidentes, qu'on aura à faire pour lever l'hypothèse de la continuité.

La démonstration sera fondée sur un théorème auxiliaire que voici:

Lemme.[3]) *Soit $g(x)$ une fonction continue, définie dans l'intervalle $a \leqq x \leqq b$, et soit E l'ensemble des points x intérieurs à cet intervalle et tels qu'il existe un ξ situé à droite de x, de sorte que $g(\xi) > g(x)$. L'ensemble E est alors ou bien vide ou bien c'est un ensemble ouvert, c'est-à-dire qu'il se décompose en un nombre fini ou une infinité dénombrable d'intervalles ouverts et disjoints (a_k, b_k), et l'on a pour tous ces intervalles*

$$g(a_k) \leqq g(b_k).$$

[3]) F. RIESZ [17].

Pour démontrer ce lemme, observons d'abord que l'ensemble E est ouvert puisque si $\xi > x_0$ et $g(\xi) > g(x_0)$, alors, à cause de la continuité, les relations $\xi > x$, $g(\xi) > g(x)$ restent valables lorsque x varie dans le voisinage du point x_0. Cela étant soit (a_k, b_k) l'un quelconque des intervalles ouverts dont se compose 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. A cet effet, soit, entre x et b_k, x_1 le point le plus proche de ce dernier pour lequel $g(x_1) \geqq g(x)$, nous avons à montrer que x_1 coïncide avec b_k. Or, s'il n'en était pas ainsi, les points ξ_1 qui correspondent à x_1 par l'hypothèse du théorème devraient etre situés au delà de b_k et, comme de plus b_k n'appartient pas à l'ensemble E, on aurait $g(x_1) < g(\xi_1) \leqq g(b_k) < g(x_1)$, ce qui implique contradiction.

On peut d'ailleurs montrer et le lecteur s'en rendra aisément compte qu'on a précisément $g(a_k) = g(b_k)$, sauf peut-être si $a_k = a$. Cependant ce fait est sans conséquence pour l'application qui suit.

Cela étant, soit $f(x)$ une fonction continue et monotone pour $a \leqq x \leqq b$; pour fixer les idées, nous la supposons non décroissante. Pour examiner la dérivabilité de $f(x)$, nous allons comparer ses nombres dérivés. Comme on sait, on appelle *nombres dérivés* supérieur et inférieur à droite et on désigne respectivement par Λ_d et λ_d la plus grande et la plus petite des limites du rapport $\frac{1}{h}(f(x+h)-f(x))$ pour $h > 0$, $h \to 0$, et on définit d'une manière analogue les nombres dérivés à gauche Λ_g et λ_g. Des valeurs infinies sont admises. Une derivée finie et déterminée existe en tout point x où les quatre nombres dérivés ont la même valeur finie.

Pour démontrer le théorème de Lebesgue, on n'aura qu'à prouver que l'on a presque partout

$$1^0\ \Lambda_d < \infty; \qquad 2^0\ \Lambda_d \leqq \lambda_g.$$

En effet, en appliquant 2° à la fonction $-f(-x)$, il résulte que l'on a aussi p. p.

$$\Lambda_g \leqq \lambda_d,$$

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

$$\Lambda_d \leqq \lambda_g \leqq \Lambda_g \leqq \lambda_d \leqq \Lambda_d < \infty;$$

donc ce sont précisément les signes d'égalité qui doivent être valables, ce qu'il fallait démontrer.

Pour vérifier l'assertion 1°, c'est-à-dire que l'ensemble E_∞ des points x pour lesquels $\Lambda_d = \infty$, est de mesure nulle, observons que cet ensemble est compris dans l'ensemble E_C pour lequel $\Lambda_d > C$, C désignant une quantité choisie aussi grande qu'on voudra. Or la relation $\Lambda_d > C$ implique l'existence d'un $\xi > x$ tel que

$$\frac{f(\xi)-f(x)}{\xi - x} > C,$$

c'est-à-dire que $g(\xi) > g(x)$, où l'on a posé $g(x) = f(x) - Cx$. Donc, l'ensemble E_C est emboîté dans les intervalles (a_k, b_k) de notre théorème auxiliaire, et d'après ce théorème on a

$$f(b_k) - Cb_k \geqq f(a_k) - Ca_k,$$

c'est-à-dire que

$$C(b_k - a_k) \leqq f(b_k) - f(a_k).$$

Cela donne, par addition,

$$C \sum (b_k - a_k) \leqq \sum [f(b_k) - f(a_k)] \leqq f(b) - f(a),$$

ce qui montre que, pour C suffisamment grand, la longueur totale des intervalles (a_k, b_k) sera aussi petite que l'on voudra. C'est-à-dire que l'ensemble E_∞ est de mesure nulle.

La seconde assertion se vérifie par un raisonnement analogue mais répété alternativement sous deux formes différentes. Soient $c < C$ deux quantités positives. Formons d'abord la fonction $g(x) = f(-x) + cx$ et soit Σ_1 le système des intervalles qui y correspondent par notre théorème auxiliaire ou plutôt celui de leurs symétriques par rapport à l'origine; alors, pour des raisons analogues à celles de tout à l'heure, Σ_1 renfermera tous les x pour lesquels $\lambda_g < c$. Soit de plus Σ_2 le système formé des intervalles (a_{kl}, b_{kl}) qui correspondent à la fonction $g(x) = f(x) - Cx$, mais considérée séparément à l'intérieur de chaque intervalle (a_k, b_k). On aura alors pour ces intervalles

$$f(b_k) - f(a_k) \leqq c(b_k - a_k), \quad C(b_{kl} - a_{kl}) \leqq f(b_{kl}) - f(a_{kl}),$$

et il s'ensuit que

$$C\Sigma_2 \leqq V_2 \leqq V_1 \leqq c\Sigma_1$$

c'est-à-dire que

$$\Sigma_2 \leqq \frac{c}{C} \Sigma_1$$

où l'on a désigné par Σ_1, Σ_2 la longueur totale des deux systèmes d'intervalles et par V_1, V_2 les sommes des variations respectives de la fonction $f(x)$.

En répétant les deux procédés alternativement, on obtiendra une suite $\Sigma_1, \Sigma_2, \ldots$ de systèmes d'intervalles, chacun emboîté dans les précédents et l'on aura d'une façon générale

$$\Sigma_{2n} \leqq \frac{c}{C} \Sigma_{2n-1}.$$

Il s'ensuit que

$$\Sigma_{2n} \leqq \left(\frac{c}{C}\right)^n \Sigma_1 \to 0.$$

Or, les points x pour lesquels on a simultanément $\Lambda_d > C$ et $\lambda_g < c$, sont évidemment compris dans tous les systèmes Σ_n; c'est-à-dire qu'ils forment un ensemble E_{cC} de mesure nulle. Enfin, chaque point x tel que $\Lambda_d > \lambda_g$ appartient

à de tels ensembles, et l'on peut même supposer que c et C sont des nombres rationnels, pour la seule raison que, entre deux nombres réels différents, on peut toujours intercaler deux nombres rationnels. C'est-à-dire que, en formant les ensembles E_{cC} 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 et par conséquent E^* et, à plus forte raison, l'ensemble envisagé qui y est compris, seront eux-mêmes de mesure nulle.

Ainsi le théorème est démontré dans le cas où la fonction monotone $f(x)$ est *continue*. Pour l'étendre au cas des fonctions *discontinues*, observons que notre théorème auxiliaire reste valable, après des modifications presque évidentes, pour des fonctions discontinues. En effet, pour notre fin, il suffit d'envisager le cas où les limites $g(x-0)$ et $g(x+0)$ existent, tel étant le cas évidemment pour les fonctions monotones $f(x)$, donc aussi pour $f(x) - Cx$ et $f(-x) + cx$. Désignons, pour $a < x < b$, par $G(x)$ la plus grande des valeurs $g(x-0)$, $g(x)$, $g(x+0)$, et convenons de poser $G(a) = g(a+0)$ et $G(b) = g(b-0)$. Cela étant, les points x, s'il y en a, intérieurs à (a, b) et tels qu'il existe un $\xi > x$ de sorte que $g(\xi) > G(x)$, forment un ensemble ouvert, et pour les intervalles (a_k, b_k) dont se compose cet ensemble, on aura $g(a_k + 0) \leqq G(b_k)$.

Les modifications que l'on aura à faire dans les raisonnements qui précèdent, pour démontrer le théorème auxiliaire sous cette forme élargie et pour l'appliquer au cas des fonctions monotones discontinues, sont tellement évidentes qu'il nous est permis de supprimer les détails. Le seul point sur lequel nous ayons à insister, c'est que l'introduction de la fonction $G(x)$ ne touche en rien les points de continuité et quant aux points de discontinuité, formant un ensemble dénombrable [1] et à plus forte raison de mesure nulle, on pourra à chaque instant les ajouter à l'ensemble actuel ou les en exclure, suivant les exigences.

---

[1]) Il suffit de voir que, k étant un entier positif quelconque, les points x où

$$|g(x+0) - g(x-0)| > \frac{1}{k}$$

sont en nombre fini tout au plus. En effet, s'il y en avait une infinité, on en pourrait tirer une suite partielle convergente monotone, ou bien croissante ou bien décroissante. La limite étant désignée par ξ, l'une au moins des limites $g(\xi-0)$, $g(\xi+0)$ ne peut pas exister, contrairement à l'hypothèse faite.

4. Fonctions à variation bornée.

Nous allons énoncer notre résultat pour une classe encore plus étendue de fonctions, celle des fonctions à variation bornée. Cette classe de fonctions joue un rôle capital dans plusieurs branches de l'Analyse, entre autres dans la théorie des séries de Fourier et dans la rectification des courbes, et bien entendu aussi dans la théorie de l'intégration. On y arrive entre autres, en observant que la dérivabilité presque partout, quand elle est vraie pour deux fonctions, disons $f_1(x)$ et $f_2(x)$, reste valide pour leur différence $f(x)=f_1(x)-f_2(x)$ et que, d'autre part, lorsque $f_1(x)$ et $f_2(x)$ sont non décroissantes, on a évidemment, pour toute décomposition de l'intervalle (a, b) en des intervalles partiels (x_{k-1}, x_k) $(k=1, 2, \ldots, n;\ x_0=a, x_n=b)$,

$$\sum_1^n |f(x_k)-f(x_{k-1})| \leqq f_1(b)-f_1(a)+f_2(b)-f_2(a).$$

On appelle *fonctions à variation bornée* les fonctions $f(x)$, continues ou non, pour lesquelles la somme envisagée

$$\Sigma_{ab}=\sum_1^n |f(x_k)-f(x_{k-1})| \tag{3}$$

ne dépasse pas une borne finie, indépendante du choix particulier de la décomposition. La plus petite des bornes admissibles s'appelle *la variation totale de $f(x)$ dans l'intervalle* (a, b); on la désignera par $T(a, b)$.

La variation totale est une "fonction additive de l'intervalle". Cela veut dire que, c étant un point entre a et b, la fonction $f(x)$ est à variation bornée dans (a, b) si et seulement si elle l'est dans (a, c) et dans (c, b), et qu'alors

$$T(a, b)=T(a, c)+T(c, b).$$

Il n'y a qu'à observer à cet effet que, les sommes Σ_{ab} ne pouvant qu' augmenter lorsqu'on intercale un nouveau point de décomposition, il suffit d'envisager les décompositions de (a, b) qui proviennent d'une décomposition de (a, c) et d'une décomposition de (c, b); alors $\Sigma_{ab}=\Sigma_{ac}+\Sigma_{cb}$ et la proposition se vérifie en passant aux plus petites bornes supérieures.

Nous venons de voir que la différence de deux fonctions non-décroissantes est à variation bornée. On doit à CAMILLE JORDAN la réciproque, savoir le

Théorème. *Toute fonction à variation bornée est la différence de deux fonctions non-décroissantes.*

La démonstration de ce fait est très simple; il n'y a qu'à introduire la fonction $T(x)=T(a, x)$, égale à la variation totale de $f(x)$ calculée pour l'intervalle (a, x), et que nous proposons appeler aussi, par analogie aux intégrales indéfinies, *la variation totale indéfinie de* $f(x)$; $T(x)$ ainsi que $T(x)-f(x)$ sont alors non-décroissantes et fournissent la décomposition exigée

$$f(x)=T(x)-[T(x)-f(x)].$$

Cela est évident pour $T(x)$; en effet, si $x<\xi$, on a

$$T(a, \xi)=T(a, x)+T(x, \xi),$$

donc

$$T(\xi)-T(x)=T(x, \xi)\geqq 0.$$

Pour montrer que

$$T(x)-f(x)\leqq T(\xi)-f(\xi),$$

ou, ce qui revient au même, que

$$f(\xi)-f(x)\leqq T(\xi)-T(x)=T(x, \xi),$$

il n'y a qu'à observer que $|f(\xi)-f(x)|$ est une somme particulière de type Σ (où il n'y a pas de points de décomposition intérieurs), et que par conséquent

$$|f(\xi)-f(x)|\leqq T(x, \xi).$$

D'ailleurs, la dernière inégalité suggère encore une seconde décomposition de $f(x)$ en fonctions monotones:

$$f(x)=P(x)-N(x)$$

où

$$P(x)=\frac{1}{2}[T(x)+f(x)] \quad \text{et} \quad N(x)=\frac{1}{2}[T(x)-f(x)]$$

sont, à une constante additive près, ce qu'on appelle respectivement *les variations positive et négative* de $f(x)$ dans l'intervalle (a, x) et que nous appellerons aussi *les variations positive et négative indéfinies* de $f(x)$.

Enfin, comme la dérivation d'une différence se fait terme à terme, et comme de plus la réunion des deux ensembles d'exception, s'il y en a, dont chacun de mesure nulle, est elle-même de mesure nulle, nous sommes à même d'énoncer notre théorème sous sa forme finale:

Théorème de Lebesgue. *Toute fonction à variation bornée admet presque partout une dérivée finie.*

QUELQUES CONSÉQUENCES IMMÉDIATES DU THÉORÈME DE LEBESGUE.

5. Théorème de Fubini sur la dérivation des séries à termes monotones.

Dans ce qui suit, nous allons tirer quelques conséquences plus ou moins immédiates du théorème fondamental que nous venons d'établir. Commençons par le théorème de FUBINI concernant la dérivation des séries à termes monotones.

Théorème de Fubini.[5]) *Soit*

$$f_1(x)+f_2(x)+\cdots=s(x) \tag{4}$$

une série convergente dont les termes sont des fonctions monotones de même type, définies sur l'intervalle $a\leqq x\leqq b$. *On aura*

$$f_1'(x)+f_2'(x)+\cdots=s'(x), \tag{5}$$

sauf peut-être dans un ensemble de mesure nulle, c'est-à-dire que la dérivation terme à terme est permise presque partout.

Voici la démonstration. Sans restreindre la généralité, on peut manifestement supposer que $f_n(a)=0$. Pour fixer les idées, nous supposons aussi que les f_n sont des fonctions non-décroissantes. Cela étant, posons

$$s_n(x)=f_1(x)+f_2(x)+\cdots+f_n(x); \quad s_n(x)\to s(x).$$

Sauf un ensemble E_0 de mesure nulle, réunion de l'infinité dénombrable de tous les ensembles d'exception, toutes ces fonctions admettent des dérivées finies et déterminées. Comme on a évidemment

$$s_n'(x)\leqq s_{n+1}'(x)\leqq s'(x),$$

la série au premier membre de (5) a un sens et est convergente sauf au plus dans E_0. De plus, les s_n' allant en croissant, il suffira de vérifier la relation (5), c'est-à-dire la relation $s'(x)-s_n'(x)\to 0$, pour une suite partielle $n_1, n_2, \ldots$ convenablement choisie. On fera ce choix de sorte que $s(b)-s_{n_k}(b)$ diminue assez rapidement pour que la série formée de ces différences et, avec elle, celle formée des différences

$$s(x)-s_{n_k}(x)\leqq s(b)-s_{n_k}(b)$$

soit elle-même convergente; on pourra par exemple choisir les n_k de sorte que

$$s(b)-s_{n_k}(b)<2^{-k}.$$

Alors, la série formée des $s-s_{n_k}$ étant du même type que (4), la série qui en résulte par dérivation terme à terme convergera presque partout et, à plus forte raison, on aura presque partout

$$s'(x)-s_{n_k}'(x)\to 0,$$

ce qu'il fallait démontrer.

6. Points de densité des ensembles linéaires.

Le théorème que nous venons d'établir fournit, comme application immédiate, un théorème sur la *densité* des ensembles linéaires. Pour ce but, disons tout d'abord ce qu'on entend par points de densité d'un tel ensemble. Leur définition est fondée sur la notion de *mesure extérieure*.

[5]) Fubini [2], cf. encore Tonelli [1], Rajchman—Saks [1].

La mesure extérieure $m_e(E)$ d'un ensemble E de points x se définit comme la borne inférieure des longueurs totales de tous les systèmes d'intervalles qui renferment E. Il s'ensuit immédiatement de la définition que $m_e(E_1\cup E_2)=m_e(E_1)+m_e(E_2)$ dans tous les cas où E_1 et E_2 sont compris dans deux intervalles disjoints, fait que l'on peut aussi exprimer en disant que la mesure extérieure de la partie d'un ensemble E comprise dans un intervalle variable est une fonction additive d'intervalle.

Convenons de dire que le point x, appartenant ou non à l'ensemble E, en est un *point de densité* lorsqu'on a

$$\frac{m_e(E; x-h, x+k)}{h+k}\to 1 \qquad (0<h, k\to 0),$$

l'expression au numérateur indiquant la mesure extérieure de la partie de E comprise entre $x-h$ et $x+k$. Avec cette dénomination, le theorème dont il s'agit s'énonce de la facon suivante:

Théorème.[6]) *Presque tous les points (c'est-à-dire tous, sauf au plus ceux d'un ensemble de mesure nulle) d'un ensemble linéaire quelconque sont des points de densité de cet ensemble.*

Enonçons le théorème sous forme analytique en introduisant la fonction $f(x)$ égale dans l'intervalle (a, b) où est situé l'ensemble E, à la mesure extérieure de la partie de E qui est comprise entre a et x. Cela posé, le théorème affirme que $f'(x)=1$ en presque tout point de E. Pour le démontrer, on n'aura qu'à envisager des ensembles ouverts (systèmes d'intervalles) $\Sigma_1, \Sigma_2, \ldots$ auxquelles E est intérieur et dont les longueurs totales tendent rapidement vers $m_e(E)$. En désignant par $f_n(x)$ la fonction analogue à $f(x)$ formée avec Σ_n au lieu de E, c'est-à-dire la longueur totale des intervalles ou de leurs segments appartenant à Σ_n et situés à gauche de x, on n'aura qu'à appliquer le theorème de Fubini à la série composée des différences $f_n(x)-f(x)$; il s'ensuivra que $f_n'(x)-f'(x)$ tend vers zéro presque partout dans (a, b) donc dans E, et comme de plus $f_n'(x)=1$ sur E pour tous les n, le théorème est démontré.

Il convient d'observer que Lebesgue n'a énoncé le théorème de densité que pour les ensembles dits mesurables dont nous parlerons plus loin; d'ailleurs, les deux énoncés ne diffèrent qu'en apparence; en réalité, ils sont des corollaires l'un de l'autre.

[6]) Lebesgue [*] (1. éd., p. 123—124, et 2. éd., p. 185—187).

7. Fonctions des sauts.

Voici un autre corrollaire du théorème de Fubini.

Soit $\{x_n\}$ un ensemble fini ou dénombrable dans l'intervalle (a, b), et envisageons la série

$$\sum f_n(x)$$

où

$$f_n(x)=0 \text{ pour } x<x_n, \quad f_n(x_n)=u_n, \quad f_n(x)=u_n+v_n \text{ pour } x>x_n.$$

Supposons que les séries

$$\sum u_n, \quad \sum v_n$$

soient absolument convergentes. En ce cas, la somme $s(x)$ de le série envisagée existe et est ce qu'on appelle une *fonction des sauts;* elle peut être écrite sous la forme explicite

$$(6) \qquad s(x)=\sum_{x_n\leqq x} u_n+\sum_{x_n<x} v_n.$$

Dans le cas où $u_n\geqq 0$ et $v_n\geqq 0$, les fonctions $f_n(x)$ sont croissantes et $f_n'(x)=0$ sauf au point $x=x_n$, donc, en vertu du théorème de Fubini, *on a aussi $s'(x)=0$ presque partout.*

Le même fait subsiste même dans le cas général où les quantités u_n, v_n peuvent être des nombres réels de signe quelconque. Tout revient à montrer que *toute fonction des sauts $s(x)$ est à variation bornée* et que *sa variation totale indéfinie $T(x)$ est aussi une fonction des sauts, notamment celle correspondant aux mêmes points x_n et aux quantités $|u_n|, |v_n|$ au lieu de u_n, v_n. Les variations positive et négative indéfinies sont alors engendrées respectivement par les parties positives et négatives de u_n, v_n.*

Il suffit de montrer à cet effet que la variation totale $T=T(b)$ est égale à $U=\Sigma(|u_n|+|v_n|)$; l'assertion relative à $T(x)$ en résultant si l'on envisage, au lieu de (a, b), l'intervalle partiel (a, x).

Etant donné $\varepsilon>0$, choisissons l'entier M de façon que la somme des M premiers termes de la série définissant U, approche de U à ε près. Envisageons une décomposition de l'intervalle (a, b) telle que chaque intervalle partiel fermé $\alpha\leqq x\leqq\beta$ contienne un et un seul des points $x_1, x_2, \ldots, x_M$ et cela à son extrémité gauche ou droite. Pour un intervalle partiel de type $\alpha\leqq x\leqq x_r$ où $r\leqq M$, on a

$$|s(x_r)-s(\alpha)|=\Big|u_r+\sum_{\alpha<x_n<x_r} u_n+\sum_{\alpha\leqq x_n<x_r} v_n\Big|\geqq|u_r|-\Big(\sum_{\alpha<x_n<x_r}|u_n|+\sum_{\alpha\leqq x_n<x_r}|v_n|\Big)$$

et tous les u_n et v_n qui figurent dans cette formule sont, sauf u_r, d'indice $>M$. La situation est analogue pour les intervalles partiels de type $x_r\leqq x\leqq\beta$. Il en résulte que la somme $\Sigma|s(\beta)-s(\alpha)|$, étendue à tous les intervalles partiels de la décomposition envisagée, est $\geqq \sum_1^M(|u_n|+|v_n|)-\varepsilon\geqq U-2\varepsilon$. Comme, d'autre part, il est manifeste par la représentation (6) de la fonction $s(x)$ que la somme analogue correspondant à une décomposition quelconque de (a, b) ne peut dépasser la quantité U, on conclut que la variation totale T est égale à U, ce qui achève la démonstration des propositions ci-dessus.

Etant à variation bornée, la fonction $s(x)$ admet en tout point des limites de gauche et de droite. *Montrons qu'aux points donnés x_n, $s(x)$ fait des sauts de gauche et de droite respectivement égaux à u_n et à v_n, et qu'elle est continue aux autres points.*

Servons-nous de nouveau à cet effet des décompositions du type particulier envisagé plus haut; on aura pour un intervalle de type (α, x_r):

$$|s(x_r)-s(\alpha)-u_r|<\varepsilon;$$

en faisant tendre α vers x_r il en résulte que

$$|s(x_r)-s(x_r-0)-u_r|\leqq\varepsilon;$$

et cela pour $r=1, 2, \ldots, M$. Or M n'était limité par ε qu'inférieurement, ce résultat est donc valable pour $r=1, 2, \ldots$. Comme ε était arbitraire, on obtient finalement que

$$s(x_r)-s(x_r-0)=u_r \qquad (r=1, 2, \ldots).$$

Un raisonnement analogue fournit la relation

$$s(x_r+0)-s(x_r)=v_r \qquad (r=1, 2, \ldots).$$

Le même raisonnement peut aussi être appliqué à un point x qui est différent de tous les x_n donnés; il n'y a qu'à l'ajouter à la suite $\{x_n\}$ en lui faisant correspondre les quantités u et v égales à 0, ce qui ne change rien à la définition de $s(x)$. Il en résulte que $s(x)-s(x-0)=0$, $s(x+0)-s(x)=0$, c'est-à-dire que x est un point de continuité.

L'intérêt des fonctions des sauts consiste en ce que *toute fonction à variation bornée $f(x)$ peut être décomposée en somme d'une fonction continue à variation bornée $g(x)$ et d'une fonction des sauts $s(x)$*; on les appelle la *partie continue* et la *partie des sauts de $f(x)$*. En effet, il n'y a qu'à définir $s(x)$ comme la fonction des sauts dont les points de discontinuité et dont les sauts correspondants sont égaux à ceux de $f(x)$; $g(x)=f(x)-s(x)$ est alors partout continue et, étant la différence de deux fonctions à variation bornée, elle est aussi à variation bornée.

On voit immédiatement que, *dans le cas où $f(x)$ est monotone, sa partie continue $g(x)$ et sa partie des sauts $s(x)$ sont aussi monotones et de même sens.*

8. Fonctions à variation bornée quelconques.

Nous venons de voir que pour une fonction des sauts $s(x)$ on a presque partout $s'(x)=0$. La variation totale indéfinie $T(x)$ de $s(x)$ étant aussi une fonction des sauts, on a presque partout $T'(x)=0$.

L'égalité de ces deux dérivées se généralise au cas d'une fonction à variation bornée quelconque $f(x)$ et de sa variation totale indéfinie $T(x)$.

Théorème. *On a presque partout*

$$T'(x) = |f'(x)|.$$

Commençons la démonstration en choisissant une suite $\{\Delta_n\}$ de décompositions de l'intervalle (a, b) telle que la somme (3) correspondant à la décomposition Δ_n approche de la variation totale $T = T(b)$ à 2^{-n} près. Faisons correspondre à la décomposition Δ_n la fonction $f_n(x)$ suivante. Dans chacun des segments $x_{k-1} \leqq x \leqq x_k$ de cette décomposition, soit $f_n(x)$ égale à

$$f(x) + \text{constante} \quad \text{ou à} \quad -f(x) + \text{constante}$$

selon que

$$f(x_k) - f(x_{k-1}) \geqq 0 \quad \text{ou} \quad < 0;$$

les constantes doivent être déterminées de sorte que $f_n(a) = 0$ et que les valeurs prises aux points x_k s'ajustent.

On a alors

$$f_n(x_k) - f_n(x_{k-1}) = |f(x_k) - f(x_{k-1})|$$

et par conséquent

$$T(b) - f_n(b) = T(b) - \sum_k [f_n(x_k) - f_n(x_{k-1})] \leqq 2^{-n}.$$

D'autre part, la fonction $T(x) - f_n(x)$ est *croissante*, ou, ce qui revient au même, on a

$$T(\xi) - T(x) \geqq f_n(\xi) - f_n(x) \quad \text{pour} \quad x < \xi.$$

Cela découle immédiatement de l'inégalité

$$T(\xi) - T(x) \geqq |f(\xi) - f(x)|$$

lorsque x et ξ appartiennent au même segment (x_{k-1}, x_k). De là on passe au cas où

$$x < x_k < \cdots < x_p < \xi$$

en ajoutant les inégalités relatives aux segments

$$(x, x_k), (x_k, x_{k+1}), \ldots, (x_p, \xi).$$

La série

$$\sum [T(x) - f_n(x)]$$

étant majorée par la série convergente $\Sigma 2^{-n}$, est aussi convergente. En vertu du théorème de Fubini, la série dérivée converge presque partout, et par conséquent

$$T'(x) - f_n'(x) \to 0$$

presque partout. Or on a évidemment

$$f_n'(x) = \pm f'(x);$$

vu que $T'(x) \geqq 0$ puisque $T(x)$ est croissante, cela prouve que $T'(x) = |f'(x)|$ presque partout.

La raisonnement ci-dessus permet aussi d'établir des relations entre les *discontinuités* de $f(x)$ et celles de sa variation totale indéfinie $T(x)$. On démontre notamment le

Théorème. *$f(x)$ et $T(x)$ ont les mêmes points de continuité et de discontinuité et leurs sauts sont égaux au signe près, c'est-à-dire qu'on a en tout point x*

$$T(x) - T(x-0) = |f(x) - f(x-0)|, \quad T(x+0) - T(x) = |f(x+0) - f(x)|.$$

En effet, on a pour $x < \xi$

$$|T(\xi) - T(x) - f_n(\xi) + f_n(x)| \leqq |T(\xi) - f_n(\xi)| + |T(x) - f_n(x)| \leqq 2 \cdot 2^{-n};$$

en faisant tendre ξ vers x, il résulte que

$$|T(x+0) - T(x) - f_n(x+0) + f_n(x)| \leqq 2^{1-n}.$$

Donc

$$f_n(x+0) - f_n(x) \to T(x+0) - T(x) \quad \text{lorsque} \quad n \to \infty.$$

Or, les sauts de f_n sont évidemment égaux, au signe près, à ceux de f, donc

$$|f(x+0) - f(x)| = T(x+0) - T(x).$$

L'assertion concernant les limites de gauche se vérifie de la même façon.

9. Théorème de Denjoy—Young—Saks sur les nombres dérivés des fonctions les plus générales.

Bien que nous n'en ayons pas besoin dans la suite, il ne manque certainement pas d'intérêt de faire connaître ici le théorème très général concernant la dérivation, ou plus précisément, *l'allure des quatre nombres dérivés des fonctions les plus générales.*

Le théorème est dû à Denjoy et à M^me^ Young[7]) qui l'ont établi indépendamment l'un de l'autre dans le cas des fonctions continues; puis, M^me^ Young l'a étendu aux fonctions mesurables[8]); enfin Saks a montré que le théorème subsiste aussi pour les fonctions les plus générales[9]). Conformément à ce que l'on attend toujours lorsqu'il s'agit d'un théorème de très grande généralité, la démonstration inventée par Saks est d'une extrême simplicité.

Considérons, avec Denjoy, deux nombres dérivés comme *associés* s'ils sont relatifs aux mêmes côtés, comme par exemple λ_g et Λ_g, et comme *opposés* s'ils correspondent à des côtés et à des rangs différents, comme par exemple λ_g et Λ_d. Le théorème en question peut être énoncé comme suit:

[7]) Denjoy [1] (en particulier p. 174—195); G. C. Young [1].
[8]) G. C. Young [2].
[9]) Saks [1]; cf. encore Hanson [1], Blumberg [1].

Théorème. *A part un ensemble de mesure nulle, il ne pourra se présenter que les cas suivants. Deux dérivés associés sont ou bien égaux et finis, ou bien inégaux et l'un au moins est infini; deux dérivés opposés sont simultanément finis et égaux ou infinis et inégaux, savoir celui de rang supérieur égal à $+\infty$ et l'autre égal à $-\infty$.*

D'ailleurs la première loi, celle sur les dérivés associés, n'est qu'une conséquence évidente de la seconde et nous n'aurons à nous occuper que de celle-ci.

Avant d'entrer dans les détails, nous attirons l'attention sur l'extrême simplicité de la loi que l'on obtient lorsqu'on cesse de distinguer entre droite et gauche et qu'on ne considère que les deux nombres dérivés *neutres*, inférieur et supérieur, définis par exemple comme la plus petite et la plus grande des limites du rapport

$$\frac{f(x+h)-f(x-k)}{h+k} \qquad (h, k \geqq 0,\ \ 0<h+k \to 0).$$

Alors, *à part un ensemble de mesure nulle, il ne se présente que les deux possibilités extrêmes: ou bien ces deux limites sont infinies et de sens contraire, l'une infinie négative et l'autre infinie positive, ou bien la fonction admet une dérivée déterminée et finie.*

Dans le cas particulier d'une fonction monotone, les infinis de sens contraire ne pouvant se présenter, il ne reste qu'une seule possibilité, celle de l'existence d'une dérivée finie et déterminée.

Esquissons la démonstration. Tout revient à prouver que, presque partout où le nombre dérivé λ_g n'est pas infini négatif, ce dérivé et son opposé Λ_d sont égaux et finis: la loi générale en ressort en remplaçant $f(x)$ successivement par $-f(x)$, $f(-x)$ et $-f(-x)$. Pour fixer les idées, nous supposerons que $f(x)$ est définie dans l'intervalle (a, b), que rien n'empêche d'ailleurs de remplacer par un ensemble quelconque. Soit E l'ensemble des points x pour lesquels λ_g diffère de $-\infty$; cet ensemble peut être envisagé comme la réunion d'une infinité dénombrable d'ensembles $E_{n,r}$ où $n=0, 1, 2, \ldots$ et où r parcourt les nombres rationnels compris dans l'intervalle (a, b), ces ensembles étant formés respectivement des points $x>r$ dans lesquels on a

$$\frac{f(x)-f(\xi)}{x-\xi} > -n$$

pour tous les ξ situés entre r et x. La somme d'une infinité dénombrable d'ensembles de mesure nulle étant elle-même de mesure nulle, il nous suffira de prouver que la loi dont il s'agit est valable presque partout dans chaque $E_{n,r}$, et comme de plus le cas général se ramène au cas $n=0, r=0$ et cela en remplaçant $f(x)$, par $f(x-r)+nx$, nous n'aurons à nous occuper que de l'ensemble $E_0 = E_{0,0}$. Faisons abstraction de ceux des points de cet ensemble qui n'en sont pas des points de densité (il suffirait aussi de n'exclure que ceux qui ne le sont pas par rapport à la fermeture $\bar{E}_0$ de E_0) et aussi de ceux où $f(x)$ n'admet pas de derivée finie et déterminée par rapport à l'ensemble E_0, c'est-à-dire calculée de sorte que l'on s'approche de x en ne quittant pas l'ensemble E_0. De la sorte, nous n'avons supprimé jusqu'ici qu'un ensemble de mesure nulle; en effet, grâce à la définition de l'ensemble E_0, la fonction $f(x)$, considérée à part en cet ensemble, y est monotone et par conséquent elle y admet, presque partout, une dérivée finie et déterminée par rapport à cet ensemble, ce dernier fait n'étant qu'un corollaire évident du théorème de Lebesgue.

Envisageons les points x de E_0 qui restent. Considérons le rapport d'accroissement

$$\frac{f(x')-f(x)}{x'-x}.$$

Lorsque x' tend vers x en ne quittant pas l'ensemble E_0, le rapport tend, d'après l'hypothèse faite, vers une limite qu'on pourra désigner par $f'_{E_0}(x)$. Dans les cas où x' n'appartient pas à l'ensemble E_0, mais est déjà suffisamment proche de x, x' pourra être remplacé, grâce à l'hypothèse de densité, par un $\xi>x'$ appartenant à E_0 et tel que la différence $\xi-x'$ soit infiniment petite par rapport à $x'-x$. Comme, par la définition de l'ensemble $E_0=E_{0,0}$, on aura $f(\xi)\geqq f(x')$, le numérateur du rapport envisagé ne diminue donc pas lorsqu'on remplace x' par ξ, quant au dénominateur, il n'en est pas sensiblement altéré. En tenant compte encore des deux signes que peut prendre ce dernier, on voit immédiatement que

$$\lambda_g \geqq f'_{E_0}(x) \geqq \Lambda_d,$$

et comme, d'autre part, la quantité $f'_{E_0}(x)$, d'après sa définition, représente elle-même une des limites, de gauche ainsi que de droite, du même rapport d'accroissement par lequel on définit les quantités λ_g et Λ_d, c'est-à-dire que

$$\lambda_g \leqq f'_{E_0}(x) \quad \text{et} \quad \Lambda_d \geqq f'_{E_0}(x),$$

ce n'est que le signe d'égalité qui pourra se présenter. Le théorème est donc démontré.

FONCTIONS D'INTERVALLE.

10. Préliminaires.

Nous aboutirons à des applications importantes en considérant d'autres fonctions d'intervalle, non nécessairement additives. D'une façon générale, on entend par fonction d'intervalle $f(I)=f(\alpha, \beta)$ une loi qui fait correspondre aux intervalles $I=(\alpha, \beta)$ appartenant à une certaine famille une quantité déterminée. D'ailleurs rien n'empêche de considérer des fonctions multivalentes d'intervalle, comme par exemple

$$f(\alpha,\beta)=(\beta-\alpha)f(\xi),$$

ξ désignant une valeur située entre α et β, d'ailleurs arbitraire, et $f(x)$ une fonction ordinaire; on sait comment la quantité $f(\alpha,\beta)$ intervient dans la théorie de l'intégration ou, plus précisément, de ce que l'on appelle l'intégrale de Riemann. Les enveloppes supérieure et inférieure de cette fonction, savoir

$$(7)\qquad (\beta-\alpha)\sup_{\alpha\leqq x\leqq\beta} f(x),\quad (\beta-\alpha)\inf_{\alpha\leqq x\leqq\beta} f(x)$$

qui sont uniquement déterminées, servent, somme on sait, à définir les intégrales par excès et par défaut, appelées aussi intégrales de Darboux. La variation absolue

$$|f(\beta)-f(\alpha)|$$

d'une fonction $f(x)$ intervient, nous l'avons vu plus haut, dans la notion de fonction à variation bornée. La longueur de la corde qui joint les points d'abscisses $x=\alpha$ et $x=\beta$ d'une courbe $y=f(x)$ ou, plus généralement, celle de la corde d'une courbe

$$x=x(t),\ y=y(t),\ z=z(t)$$

correspondant aux valeurs α et β du paramètre t, est à la base des notions de courbe rectifiable et de sa longueur. La fonction multivalente

$$(8)\qquad (g(\beta)-g(\alpha))f(\xi)$$

composée de deux fonctions ordinaires dont l'une, $g(x)$, à variation bornée, et l'autre, $f(x)$, continue, sert à définir l'intégrale de Stieltjes[10]). Nous verrons que tous ces exemples appartiennent, sous des hypothèses convenables, à la classe des *fonctions intégrables d'intervalle*. Citons enfin la fonction d'intervalle

$$(9)\qquad \frac{(f(\beta)-f(\alpha))^2}{g(\beta)-g(\alpha)}$$

sur laquelle se fonde l'intégrale de Hellinger[11]) dont cet auteur fait usage dans la théorie des formes quadratiques à une infinité de variables.

La définition de *fonction intégrable d'intervalle* et de son *intégrale* est bien simple; il ne s'agit là que d'une généralisation immédiate de l'intégrale de Riemann ou de celles de Darboux. On partage l'intervalle (a,b) en des intervalles partiels, on forme la somme des valeurs qui correspondent à ces intervalles et l'on regarde si ces sommes tendent vers une limite finie et déterminée lorsqu'on fait varier la décomposition de sorte que la longueur des intervalles partiels tende uniformément vers zéro, ou, en d'autres termes, si l'on peut faire

[10]) Les intégrales de Stieltjes seront analysées en détail au Chap. III.

[11]) Cf. Hellinger [1] (en particulier p. 25—51) et [2] (en particulier p. 234—240).

correspondre à tout $\varepsilon>0$ un $\delta=\delta(\varepsilon)>0$ de sorte que pour toute décomposition de (a,b) en des segments inférieurs à δ, les sommes correspondantes approchent à ε près une limite déterminée. Quand il en est ainsi, la *fonction d'intervalle* est dite *intégrable* et la limite de la somme est appelée son *intégrale* et elle est désignée par

$$\int_a^b f(\alpha,\beta).$$

Pour faire mieux ressortir la généralité de cette notion, qu'il nous suffise d'attirer l'attention du lecteur sur le fait évident que toutes les fonctions additives d'intervalle, ou, ce qui revient au même, toutes les fonctions $f(\alpha,\beta)$ de la forme $f(\beta)-f(\alpha)$ sont intégrables quelque singulière que soit la fonction $f(x)$.

Avant d'examiner les exemples particuliers que nous venons d'énumérer, nous allons établir deux théorèmes d'ordre général, concernant les relations qui existent entre l'intégration et la dérivation des fonctions d'intervalle.

On entend par *dérivée de la fonction d'intervalle* $f(\alpha,\beta)$ au point x la limite, lorsqu'elle existe, du rapport

$$\frac{f(\alpha,\beta)}{\beta-\alpha},$$

l'intervalle (α,β) se contractant au point x. Les nombres dérivés se définissent d'une manière analogue. Lorsque, en particulier, $f(\alpha,\beta)$ est additive, ces quantités ne sont autres que celles qui correspondent, au sens ordinaire, à la fonction $F(x)=f(a,x)$.

Le problème de la dérivation des fonctions d'intervalle a été étudié par plusieurs auteurs d'un point de vue très général[12]). Le premier théorème que nous allons démontrer n'est que la quintessence des principaux résultats obtenus.

11. Premier théorème fondamental.

Il s'agit du

Théorème. *Soit $f(\alpha,\beta)$ une fonction d'intervalle intégrable dans (a,b), de plus non-négative et d'intégrale nulle. Dans ces hypothèses, $f(\alpha,\beta)$ admet presque partout dans l'intervalle (a,b) une dérivée déterminée égale à zéro.*

La démonstration du théorème est presque immédiate. Soient $\delta_1,\delta_2,\ldots$ des quantités positives choisies de sorte que la somme des valeurs $f(\alpha,\beta)$ qui correspondent à une décomposition de l'intervalle entier (a,b) en des segments dont la longueur est égale au plus à δ_n, ne dépasse pas le terme général d'une série convergente, donnée d'avance, par exemple de la série $\Sigma 2^{-n}$. Cela étant, envisageons les fonctions $F(x)$ définies comme suit. La fonction $F_n(x)$ est

[12]) Cf., en particulier, Burkill [1]—[3]; R. C. Young [1]; Saks [*] (p. 102—107) et [*] p. 165—169).

égale à la borne supérieure des sommes des valeurs $f(\alpha, \beta)$ qui correspondent aux décompositions de l'intervalle (a, x) en des segments dont la longueur ne dépasse pas δ_n. Ces fonctions $F_n(x)$ seront évidemment des fonctions non-décroissantes, formant une série convergente, de sorte que, grâce au théorème de Fubini, $F'_n(x)$ tendra vers zéro presque partout. De plus, comme

$$f(\alpha, \beta) \leqq F_n(\beta) - F_n(\alpha) \qquad (\beta - \alpha \leqq \delta_n),$$

les dérivées des F_n seront, partout où elles existent, des majorantes des nombres dérivés de la fonction d'intervalle f. C'est-à-dire que ces nombres dérivés s'annulent presque partout, ce qu'il fallait démontrer.

12. Second théorème fondamental.

Avant de parler du second théorème fondamental, observons un fait que nous avons passé jusqu'ici sous silence. Ce fait consiste en ce que *l'intégrabilité de $f(\alpha, \beta)$ sur (a, b) entraîne celle sur tout intervalle partiel (c, d).*

A cet effet, envisageons un $\varepsilon > 0$ et le $\delta = \delta(\varepsilon)$ qui y correspond lorsqu'il s'agit de l'intervalle entier (a, b) (nº 10). Considérons deux décompositions de (c, d) en des segments inférieurs à δ et une décomposition de (a, c) et une de (d, b) du même type; ces décompositions définissent deux décompositions de l'intervalle entier (a, b). Or en formant la différence des sommes correspondantes, sommes qui approchent l'intégrale de $f(\alpha, \beta)$ sur (a, b) à ε près, la valeur numérique de cette différence ne dépasse pas 2ε, et d'autre part, les termes correspondant aux segments (a, c) et (d, b) se détruisent; par conséquent, la différence des deux sommes correspondant à (c, d) est au plus égale à 2ε; d'après le critère de convergence de Cauchy, cela garantit l'existence de la limite, c'est-à-dire de l'intégrale sur le segment (c, d).

De plus, la convergence des sommes vers les intégrales est uniforme par rapport à tous les segments. Enfin, l'intégrale sur les sous-intervalles (c, d) est évidemment une fonction *additive* d'intervalle et s'exprime sous la forme $F(d) - F(c)$ au moyen de l'intégrale indéfinie $F(x)$.

Cela étant, considérons, pour $\varepsilon > 0$ donné, une décomposition de l'intervalle (a, b) en des segments $I_1, I_2, \ldots, I_n$ ne dépassant pas en longueur la quantité $\delta = \delta(\varepsilon)$; pour cette décomposition ainsi que pour toute autre qui en découle en intercalant de nouveaux points de division, la différence entre les sommes correspondantes et l'intégrale sur (a, b) sera au plus égale à ε. En particulier, en maintenant inaltérés une partie des intervalles I_k, en divisant indéfiniment les autres et en passant à la limite, on arrivera à une somme à termes de type mixte; les uns seront de type $f(\alpha_k, \beta_k)$, les autres de type $F(\beta_k) - F(\alpha_k)$. Puis en échangeant le rôle des deux types et en formant la différence des deux sommes dont chacune approche l'intégrale de $f(\alpha, \beta)$ sur (a, b) à ε près, on arrive à l'inégalité

$$\sum_1^n \pm [f(\alpha_k, \beta_k) - (F(\beta_k) - F(\alpha_k))] \leqq 2\varepsilon,$$

et comme on dispose encore dans chaque terme des signes $\pm$, il résulte que

$$\sum_1^n |f(\alpha_k, \beta_k) - F(\beta_k) + F(\alpha_k)| \leqq 2\varepsilon.$$

C'est-à-dire que la fonction d'intervalle

$$g(I) = g(\alpha, \beta) = |f(\alpha, \beta) - F(\beta) + F(\alpha|,$$

évidemment non-négative, est intégrable et d'intégrale nulle; donc on peut y appliquer notre premier théorème fondamental et celui-ci nous assure que $g(I)$ admet presque partout une dérivée égale à zéro. On en conclut le

Théorème. *La fonction d'intervalle intégrable $f(I)$ et son intégrale indéfinie $F(x)$ admettent presque partout les mêmes nombres dérivés; en particulier, presque partout où l'une des deux admet une dérivée finie et déterminée, il en sera même de l'autre, et inversement.*

13. Les intégrales de Darboux et celle de Riemann.

Reprenons les deux fonctions d'intervalle (7) et formons leurs intégrales au sens que nous venons d'introduire; nous obtenons alors, à ce que l'on sait, les intégrales par défaut et par excès de la fonction $f(x)$, appelées aussi intégrales de Darboux. La première de ces intégrales est en général inférieure à la seconde; quand elles coïncident, on parle de l'intégrabilité et de l'intégrale au sens de Riemann. C'est-à-dire que la condition de l'intégrabilité de $f(x)$ au sens de Riemann dans l'intervalle (a, b) consiste précisément en ce que la fonction d'intervalle, évidemment non-négative,

$$(10) \qquad (\beta - \alpha)\,[\sup_{\alpha \leq x \leq \beta} f(x) - \inf_{\alpha \leq x \leq \beta} f(x)] = (\beta - \alpha)\,\omega(f; \alpha, \beta)$$

où $\omega(f; \alpha, \beta)$ désigne *l'oscillation de $f(x)$ dans l'intervalle (α, β)*, soit d'intégrale nulle. La dérivée de cette fonction d'intervalle, $\omega(x)$, *oscillation de $f(x)$ au point x*, existe et s'annule en tout point de continuité de $f(x)$, et inversement la relation $\omega(x) = 0$ au point x assure la continuité de $f(x)$ au point x.

En combinant ce fait avec le premier théorème fondamental, on aboutit à *une condition nécessaire de ce que $f(x)$ soit intégrable au sens de Riemann*, savoir que *la fonction $f(x)$ soit continue presque partout.*

Inversement, *quand $f(x)$ est supposé bornée, la même condition est aussi suffisante.* Généralement on demontre ce fait ainsi que sa nécessité sans se reporter à la théorie des fonctions d'intervalles, d'ailleurs, peut-être ne sera-t-il pas sans intérêt de le ranger dans notre ordre d'idées actuel. A cette fin, observons d'abord que $f(x)$, donc aussi $\omega(f; \alpha, \beta)$ étant bornées par hypothèse, l'intégrale indéfinie de (10), ou d'une façon précise, la fonction

$$\Omega(x)=\int_a^x (\beta-\alpha)\,\omega(f;\alpha,\beta)$$

est à rapport d'accroissement borné, c'est-à-dire qu'elle satisfait à une condition de Lipschitz,

$$\Omega(\beta)-\Omega(\alpha)\leqq C(\beta-\alpha) \qquad (\alpha<\beta); \tag{11}$$

de plus elle est non-décroissante. Enfin, d'après le second théorème fondamental, elle admet presque partout la même dérivée que la fonction d'intervalle (10), c'est-à-dire que, dans l'hypothèse faite, $\Omega'(x)=0$ presque partout. Tout revient donc à montrer qu'*une fonction* $\Omega(x)$ *non-décroissante, satisfaisant à la condition* (11) *et dont la dérivée s'annule presque partout, est nécessairement constante*, c'est-à-dire que l'image de l'intervalle (a, b) par la fonction $y=\Omega(x)$ se réduit à un seul point.

Envisageons, à cet effet, l'ensemble E des points x pour lesquels $\Omega'(x)$ n'existe ou ne s'annule pas, et son image par $y=\Omega(x)$ que nous désignons par $\Omega(E)$. E étant de mesure nulle, on peut le renfermer en un système d'intervalles de longueur totale arbitrairement petite, disons $<\varepsilon$; l'image de ces intervalles qui, grâce à (11), aura une longueur totale $<C\varepsilon$, renfermera l'ensemble $\Omega(E)$. Par conséquent, ce dernier ensemble est aussi de mesure nulle.

Il en sera de même quant à l'ensemble $\Omega(e)$, image de l'ensemble complémentaire $e=(a,b)-E$. En effet, comme $\Omega'(x)=0$ sur e, on pourra attacher à chaque point x de e des points $\xi>x$ pour lesquels $\Omega(\xi)-\Omega(x)<\varepsilon(\xi-x)$ où $\varepsilon>0$ est fixé arbitrairement petit. C'est-à-dire que, pour ε fixé, e est compris dans l'ensemble, formé par rapport à la fonction $g(x)=\varepsilon x-\Omega(x)$ figurant dans notre théorème auxiliaire, nº 3, et que nous désignerons par e_ε. D'après ce théorème, l'ensemble ouvert e_ε n'est qu'un système d'intervalles (a_k, b_k) pour lesquels $g(a_k)\leqq g(b_k)$; c'est-à-dire que $\Omega(b_k)-\Omega(a_k)\leqq\varepsilon(b_k-a_k)$; il s'ensuit que la longueur totale des intervalles $(\Omega(a_k), \Omega(b_k))$ dont se compose l'ensemble $\Omega(e_\varepsilon)$, ne surpasse pas $\varepsilon(b-a)$ et que, par conséquent, l'ensemble $\Omega(e)$, compris dans tous ces ensembles $\Omega(e_\varepsilon)$, est de mesure nulle, ce qu'il fallait démontrer.

Enfin, l'intervalle $(\Omega(a), \Omega(b))$ étant entièrement recouvert par deux ensembles de mesure nulle, devra être lui-même de mesure nulle.

A vrai dire, ce raisonnement ne sera complet qu'après avoir montré qu'un ensemble de mesure nulle ne peut épuiser l'intervalle entier. Jusqu'à présent nous n'avons pas eu à nous reporter à ce fait sans lequel cependant tous nos résultats du type "presque partout" ne seraient que des jeux mots. Pour le prouver, supposons le contraire, c'est-à-dire que l'intervalle (a, b) soit de mesure nulle; alors, quel que soit $\varepsilon>0$, l'intervalle fermé $[a, b]$ peut être recouvert par une suite d'intervalles de longueur totale $<\varepsilon$, et en prolongeant ces intervalles vers droite et gauche, par exemple en les redoublant, on arrive

à une suite d'intervalles telle que chaque point x de l'intervalle fermé $[a, b]$ est intérieur à au moins un de ces intervalles. D'après le théorème bien connu de Borel, notre suite peut être remplacée par un nombre fini de ses éléments sans cesser de recouvrir l'intervalle $[a, b]$; de là il decoule immédiatement que $2\varepsilon>b-a$, contrairement à l'hypothèse.

14. Théorème de Darboux.

Il nous reste une seconde lacune à combler. Au commencement du nº précédent nous nous sommes appuyés, sans la démontrer, sur l'existence des intégrales par défaut et par excès d'une fonction bornée d'ailleurs arbitraire, et nous en avons conclu à l'intégrabilité de la fonction d'intervalle (10). En effet, le théorème d'existence en question, le théorème de Darboux, se trouve démontré dans les Cours d'Analyse classiques, entre autres au premier chapitre du Cours de Camille Jordan. Cependant nous préférons le présenter sous la forme d'un principe général auquel nous aurons encore à nous reporter à plusieurs reprises.

Supposons que *la fonction d'intervalle* $f(\alpha,\beta)$ *soit non-décroissante par décomposition:* cela veut dire que, pour $\alpha<\beta<\gamma$,

$$f(\alpha,\gamma)\leqq f(\alpha,\beta)+f(\beta,\gamma)$$

ou, ce qui revient au même, que les sommes qui servent à définir l'intégrale augmentent ou ne diminuent pas quand on intercale un nouveau point de division. D'ailleurs, nous aurions aussi pu supposer $f(\alpha,\beta)$ décroissante au sens analogue; en effet, cela revient à considérer $-f(\alpha,\beta)$ au lieu de $f(\alpha,\beta)$.

Il nous faut encore faire une seconde hypothèse, à savoir que *la fonction* $f(\alpha,\beta)$ *soit continue partout*. Nous entendons par là que, pour tout point x, $f(\alpha,\beta)$ devient infiniement petite lorsque l'intervalle (α,β) se contracte au point x; cela veut dire que, pour chaque x, on peut attacher à chaque $\varepsilon>0$ une quantité $\delta=\delta(\varepsilon)>0$ de sorte que l'hypothèse $\alpha\leqq x\leqq\beta$, $\beta-\alpha<\delta$ entraîne que $|f(\alpha,\beta)|<\varepsilon$.

Supposons enfin que *les sommes* $\Sigma f(\alpha_{k-1},\alpha_k)$ correspondant à toutes les décompositions possibles de l'intervalle (a, b) *admettent une borne supérieure finie*.

Nous allons démontrer le

Théorème. *Sous les trois hypothèses faites, la fonction* $f(\alpha,\beta)$ *est intégrable et son intégrale est égale à la plus petite borne supérieure* L *des sommes* $\Sigma f(\alpha_{k-1},\alpha_k)$.

Envisageons à cet effet une décomposition Δ_ε pour laquelle la somme correspondante Δ_ε dépasse la quantité $L-\frac{\varepsilon}{2}$, soit ν le nombre des points de division dont consiste Σ_ε. Choisissons δ de sorte que, pour $\beta-\alpha<\delta$, on ait $|f(\alpha,\beta)|<\frac{\varepsilon}{6\nu}$ autour de tous ces points de division, et envisageons une décomposition Δ de (a,b) en des segments inférieurs à δ; il y correspond une somme Σ. En superposant les deux décompositions Δ_ε et Δ on obtient une décomposition Δ' et une somme Σ'. Or, notre fonction d'intervalle étant non-décroissante par décomposition, il s'ensuit que $\Sigma'\geqq\Sigma_\varepsilon>L-\frac{\varepsilon}{2}$. D'autre part, on pourra passer de Δ à Δ' en ν étapes, en intercalant toujours un des ν points de division dont se compose Δ_ε; à chaque pas on n'augmente la somme correspondante que de $\frac{\varepsilon}{2\nu}$ au plus; donc $\Sigma'\leqq\Sigma+\frac{\varepsilon}{2}$, et enfin, en combinant les deux inégalités, il résulte que $\Sigma\geqq L-\varepsilon$. Donc Σ approche L à ε près, ce qu'il fallait démontrer.

L'intégrale par défaut d'une fonction bornée $f(x)$ entre évidemment dans le type considéré, celle par excès ainsi que la différence des deux intégrales, l'intégrale de la fonction d'intervalle (10), sont de type opposé.

15. Fonctions à variation bornée et rectification des courbes.

Deux autres notions importantes qui, considérées comme intégrales de fonctions d'intervalle, entrent dans le type que nous venons de considérer, sont la variation totale des fonctions à variation bornée et la longueur des courbes rectifiables. Bien que la première ne soit qu'un cas particulier de la seconde, il sera instructif de l'envisager d'abord séparément.

La variation totale d'une fonction continue $f(x)$ à variation bornée, formée pour un intervalle (a,b), n'est que l'intégrale sur (a,b) de la fonction d'intervalle

$$f(\alpha,\beta)=|f(\beta)-f(\alpha)|, \tag{12}$$

ce qui se reflète aussi dans la notation généralement adoptée

$$T(x)=\int_a^x |df(x)|.$$

Cette intégrale entre évidemment dans le type considéré et en y appliquant notre second théorème fondamental, il s'ensuit immédiatement que $T(x)$, variation totale indéfinie de $f(x)$, a sa dérivée $T'(x)$ égale presque partout, au signe près, à celle de $f(x)$, c'est-à-dire que $T'(x)=|f'(x)|$, comme nous l'avons déjà vu plus haut (n° 8), même pour les fonctions discontinues.

Nous reviendrons tout à l'heure au cas des fonctions discontinues. Parlons d'abord des *courbes rectifiables*. Soit donnée la courbe Γ (pour fixer les idées, nous nous plaçons dans l'espace ordinaire): $x=x(t)$, $y=y(t)$, $z=z(t)$ $(a\leqq t\leqq b)$, où $x(t)$, $y(t)$ et $z(t)$ sont des fonctions continues. Envisageons les lignes polygonales inscrites, c'est-à-dire les lignes formées en décomposant l'intervalle (a,b) en des segments (t_{k-1},t_k) où $k=1,2,\dots,n$ et $t_0=a, t_n=b$, et en menant les cordes $P_{k-1}P_k$ où P_k désigne le point de la courbe qui correspond à la valeur $t=t_k$. La courbe Γ est dite rectifiable si la longueur de la ligne polygonale $P_0\dots P_n$ ne dépasse pas une certaine borne finie, indépendante de la manière de décomposition de l'intervalle (a,b); la plus petite de ces bornes définit la longueur de la courbe. Les cordes $P_{k-1}P_k$ étant minorées par $|x(t_k)-x(t_{k-1})|$ ainsi que par les deux autres quantités analogues et majorées par leur somme, on voit immédiatement qu'une condition nécessaire et suffisante pour que la courbe soit rectifiable, consiste en ce que les fonctions continues $x(t), y(t), z(t)$ soient à variation bornée. D'autre part, nos théorèmes généraux nous assurent que cette condition est équivalente à celle que la fonction d'intervalle $f(\alpha,\beta)$, définie comme égale à la corde joignant les points $P(\alpha)$ et $P(\beta)$ qui correspondent aux valeurs α et β du paramètre t, soit intégrable, et que la longueur de Γ n'est que l'intégrale de cette fonction prise de a jusqu'à b. De plus, en désignant par $s(t)$ la longueur d'arc de notre courbe mesurée de a jusqu'à t, notre second théorème fondamental implique l'extension la plus générale de la formule classique

$$(s'(t))^2=(x'(t))^2+(y'(t))^2+(z'(t))^2, \tag{13}$$

donnée successivement par LEBESGUE[13]) et TONELLI[14]) et valable, d'après celui-ci, *pour toute courbe rectifiable et presque partout par rapport au paramètre t, d'ailleurs arbitraire.* En choisissant, en particulier, la longueur s comme paramètre, on obtient

$$(x'(s))^2+(y'(s))^2+(z'(s))^2=1, \tag{14}$$

relation qui assure *l'existence d'une tangente déterminée presque partout par rapport à s.*

On peut arriver aux mêmes résultats sans faire appel au second théorème fondamental et cela en appliquant le premier théorème à la fonction d'intervalle $f(\alpha,\beta)$ égale à la différence de l'arc et de la corde joignant les points $P(\alpha)$ et $P(\beta)$, fonction évidemment non-négative, non-croissante par décomposition et d'intégrale 0. La relation (14) se présente alors, avec des notations évidentes, sous la forme

$$\frac{\overline{PR}}{\widehat{PQR}}\to 1$$

presque partout.

Retournons au cas d'une seule fonction continue et à variation bornée, soit $x(t)$. On pourra y attacher la "courbe" $x=x(t)$, $y=0$, $z=0$, évidemment

[13]) LEBESGUE [*] (1. éd., p. 59—63, 125—129).

[14]) TONELLI [2]; LEBESGUE [*] (2. éd., p. 198—201).

rectifiable; et la variation totale indéfinie

$$T(t) = \int_a^t |dx(t)|$$

coïncide avec la fonction $s(t)$. Donc la relation (13) donne immédiatement que

$$T'(t) = |x'(t)|$$

presque partout, c'est-à-dire que ce fait constaté indépendamment tout à l'heure n'est qu'un cas particulier de la formule classique (13) envisagée au sens général que nous venons de préciser. Cela nous suggère aussi une des voies pour passer aux fonctions discontinues. Pour fixer les idées, supposons d'abord que $x(t)$ ne présente qu'une seule discontinuité, soit pour $t = c$ $(a < c < b)$, et considérons de nouveau la "courbe" $x = x(t), y = 0, z = 0$, mais complétée maintenant, pour ne pas être discontinue, par deux segments orientés, le premier allant de $x(c-0)$ à $x(c)$ et le second de $x(c)$ à $x(c+0)$. Le même procédé devra être répété, dans le cas général, un nombre fini ou une infinité dénombrable de fois de façon à supprimer tous les points de discontinuité. En appliquant nos résultats à la courbe ainsi complétée, il s'ensuivra aisément que la relation $T'(t) = |x'(t)|$ presque partout, reste valable aussi pour les fonctions discontinues à variation bornée.

La relation entre fonctions à variation bornée et courbes rectifiables devient encore plus formelle, du moins pour des courbes planes $x = x(t), y = y(t)$, si ce plan est regardé comme plan des nombres complexes $\zeta = x + iy$. En effet, la longueur de la courbe n'est autre que la *variation totale de la fonction à valeurs complexes* $\zeta(t) = x(t) + iy(t)$, définie, tout comme dans le cas réel, par la plus petite borne supérieure des sommes

$$\sum |\zeta(t_k) - \zeta(t_{k-1})|.$$

CHAPITRE II.

INTÉGRALE DE LEBESGUE.

DÉFINITION ET PROPRIÉTÉS FONDAMENTALES.

16. Intégrale des fonctions en escalier. Deux lemmes.

Après CAUCHY et RIEMANN, durant la seconde moitié du siècle dernier, de nombreuses définitions de l'intégrale ont été successivement proposées pour les fonctions bornées ainsi que pour celles non bornées. Mais ce n'est qu'au commencement du présent siècle, en 1902, que HENRI LEBESGUE a introduit, dans sa Thèse [1], une notion d'intégrale qui devait changer l'aspect de toute une foule de problèmes dépendant de l'intégration. Les raisons d'un tel changement et avec elles l'utilité et la beauté de la théorie de LEBESGUE se verront au cours des chapitres qui suivent; inutile d'en parler d'avance.

Dans la Thèse de LEBESGUE et dans ses Leçons sur l'intégration, professées au Collège de France, qui la suivirent, la route conduisant aux résultats est encore assez ardue; des années allaient passer avant que les contemporains s'habituaient aux méthodes nouvelles. Pendant ces années on s'est efforcé d'arriver au même but par des routes plus commodes et cela en remplaçant la définition primitive par d'autres, permettant, comme le dit DE LA VALLÉE POUSSIN, de faire entrer, si possible, la nouvelle théorie dans l'enseignement classique. Dans ce qui suit, nous allons exposer la théorie en partant de l'une de ces définitions, liée à l'idée d'opération fonctionnelle, tandis que la définition primitive s'attache à la notion de mesure. A la suite, nous allons présenter encore quelques autres définitions, avec en premier lieu celle de Lebesgue, et les comparer entre elles et à celle dont nous parlerons tout à l'heure.

Nous partons de la classe des fonctions en escalier définies dans un intervalle (a, b), fini ou infini, c'est-à-dire des fonctions à valeur constante c_k dans chacun d'un nombre fini d'intervalles partiels i_k de longueur finie $|i_k|$ et s'annulant hors de ces intervalles; quant aux extrémités de ces intervalles, nous pourrons y fixer les valeurs des fonctions à notre aise; en effet, il nous sera permis, dans ce qui suit, de négliger les ensembles de mesure nulle qui nous gênent. Pour ces fonctions, nous supposons l'intégrale définie, comme d'ordinaire, par la somme

$$\sum c_k |i_k|,$$