ON DERIVATES AND THEIR PRIMITIVE FUNCTIONS

By W. H. YOUNG, Sc.D., F.R.S.

[Received November 28th, 1912.—Read December 12th, 1912.]

1. The most general of Lebesgue's already classical results in connexion with the fundamental theorem of the Integral Calculus * is that which states that, if one of the derivates of a continuous function is finite and summable, its Lebesgue integral is the function itself. It was pointed out by Beppo Levi † that, in the case of an analogous result, it followed by simple Cantor induction that that result is still true if the value $+\infty$, or $-\infty$, or both, are assumed by the derivate at a reducible set of points only. By employing the notions of positive and negative variation rather more in the form originally given to them by Jordan, I showed ‡ that the theorem, equally in the Lebesgue or in the extended form, could be split up into two components, relating respectively to functions upper semi-continuous on the right and lower semi-continuous on the left, and functions upper semi-continuous on the left and lower semi-continuous on the right respectively. In a slightly more recent paper § I showed by an entirely different method that another result of Lebesgue's, almost as general as the one just referred to, namely, that a function of bounded variation is an integral if one of its derivates is finite, could be generalised in this sense that the assuming of the value infinity with determinate sign by the derivate was permissible at any countable set of points or, even more generally, at any set of points which contained no perfect sub-set.

Since then a careful examination of the proof given by myself of Lebesgue's former theorem has led me to the conclusion that that result,

---

* The first statement of this theorem is on p. 123 of Lebesgue's *Leçons sur l'intégration*, Paris, 1904. The first wholly satisfactory proof is to be found in Lebesgue's paper, "Sur les fonctions dérivées," 1906, *Rend. dei Lincei*, Ser. 2, Vol. xv, p. 5. The only correct proof in English, as far as I am aware, is that in my paper "On Functions of Bounded Variation," 1910, *Quarterly Journal of Mathematics*, Vol. XLII, p. 78.

† "Ancora alcune osservazioni sulle funzioni derivate," 1906, *Rend. dei Lincei*, Ser. 2, Vol. xv, pp. 359 *seq.*

‡ *Loc. cit., supra*, first footnote.

§ "Note on the Fundamental Theorem of Integration," 1910, *Proc. Camb. Phil. Soc.*, Vol. XVI, pp. 35–38.

like the one just referred to, is capable of a similar extension, that we are, in fact, in the extended enunciation, at liberty to substitute for the word "reducible" the word "countable."

Combining this result with those corresponding to the two components for the two classes of non-continuous functions above described, I find that we get the following general statements, which include all others so far made :—

If a lower derivate of a function which is upper semi-continuous on the right and lower semi-continuous on the left assumes the value $+\infty$ *at a countable set of points only, and is summable over the set of points at which it is positive, the function is a lower semi-integral and is accordingly expressible as the sum of an integral and a monotone nowhere increasing function of* x. *Moreover, its positive variation is the integral of any one of its derivates over the set of points where the lower derivate in question is positive.*

If an upper derivate of a function which is lower semi-continuous on the right and upper semi-continuous on the left assumes the value $-\infty$ *at a countable set of points only, and is summable over the set of points where it is negative, the function is an upper semi-integral and is accordingly expressible as the sum of an integral and a nowhere decreasing function. Moreover, its negative variation is the integral of any one of its derivates over the set of points where the upper derivate in question is negative.*

As I think it would be difficult to exaggerate the importance in the theory of derivates of Lebesgue's original theorem, I propose to show in the present communication how these generalisations may be obtained. In the method of proof I have adopted I have kept in view the principle which actuated my wife and myself in a recent joint paper on the Existence of a Differential Coefficient, published in the *Proceedings* of this Society.* In that paper the fact that we had not succeeded to our own satisfaction in proving Lebesgue's general theorem without the use of Cantor's numbers was mentioned.

From one point of view we might, indeed, have claimed that we had been successful in doing so; for it is only in the assumption of the truth of his lemma on the existence † of a chain of intervals that Lebesgue's

* Ser. 2, Vol. 9, pp. 325–335.

† Lebesgue also used Cantor's numbers to justify the use of such a chain in calculating the variation, "Encore une observation sur les fonctions dérivées," 1907, *Rend. dei Lincei*, Ser. 5, Vol. xv, p. 96. Beppo Levi has given an argument independent of Cantor's numbers, which serves the same purpose, "Ancora alcune osservazioni sulle funzioni derivate," 1906, *Rend. dei Lincei*, Ser. 2, Vol. xv, p. 361.

proof involves Cantor's numbers implicitly. Now one of the things we did in the paper quoted was to show how that lemma could be proved without Cantor's numbers. Our object was, however, to avoid not only the use of the numbers themselves, but also the use of what might be objected to as a transfinite construction. A proof is even more convincing if we are able to employ finite sets of intervals only. The contention, indeed, that it is only with a finite number of intervals that we ought to operate is not wholly ungrounded.

In the present paper, therefore, I have been at pains to avoid all trace of the transfinite construction.

In my process of analysis of Lebesgue's theorem into its component parts I have been able to take a further step. In the enunciations above given, the hypothesis of the summability of the derivate considered over either the whole interval or a set of points in that interval has to be made. Now I had succeeded in proving* without transfinite numbers the theorem that the integral of the derivates of an integral is the integral itself, another of the important results due to Lebesgue. Subtracting, so to speak, this theorem from the two above stated as to be proved, we find that the residual parts take the following form, from which all reference to the concept of summability has disappeared :

If a function is upper semi-continuous on the right and lower semi-continuous on the left, and one of its lower derivates is negative or zero except at a set of points of zero content, then either

(1) *that derivate assumes the value* $+\infty$ *at a more than countable set of points,*

Or (2) *the function is a monotone nowhere increasing function, so that all its derivates are everywhere negative or zero.*

Again,

If a function is lower semi-continuous on the right and upper semi-continuous on the left, and one of its upper derivates is positive or zero except at a set of points of zero content, then either

(1) *that derivate assumes the value* $-\infty$ *at a more than countable set of points, or*

* "Functions of Bounded Variation," *loc. cit.*, §§ 23 and 24, pp. 81 and 82.

(2) *the function is a monotone nowhere decreasing function, so that all its derivates are everywhere positive or zero.*

From these results those mentioned above follow.

2. I shall first prove the following extension of the lemma in the joint paper already cited.

LEMMA.—*If with each point of a closed interval* (a, b) *as left-hand end-point we are given an interval or several intervals, and* $e_1 > e_2 > \ldots$ *be any monotone descending sequence of constants with zero as limit, we can find a finite number of the given intervals, nowhere overlapping, such that the complementary intervals have lengths which are respectively less than different terms of the e-sequence.**

By the lemma proved in a preceding number of these *Proceedings* † we saw that we could choose out a finite number, say m of the given intervals, numbered from left to right, $d_1, d_2, \ldots, d_m$, nowhere overlapping, and such that the sum of the lengths of the complementary intervals or gaps is less than a quantity as small as we please, which we may take to be e_1. This may be said to constitute the first stage in the process of finding the required intervals.

Now the intervals were so chosen that there is a definite integer i, different for each gap, so that the chosen interval on the right of the gap in question is d_i, and a definite integer r_i, *not necessarily different for each gap,* such that

$$d_i \geqslant e_{r_i}, \tag{1}$$

while, if X is any point belonging to the gap, whether as internal or end-point, the length of each of the given intervals corresponding to it is less than e_{r_i-1}, say

$$H < e_{r_i-1}. \tag{2}$$

To render this quite clear it would be well, perhaps, to recall the mode in which the intervals $d_1, d_2, \ldots, d_m$ were chosen. We begin by elongating all the given intervals on the left by one n-th of their length. Starting

* More generally we have by similar reasoning the following lemma:—

LEMMA.—*If with each point of a closed set S as left-hand end-point, we are given an interval, or several intervals, and* $e_1 > e_2 > \ldots$ *be any monotone descending sequence of constants with zero as limit, we can find a finite number of the given intervals, nowhere overlapping, such that in the complementary intervals the points of S can be enclosed in a finite number of intervals whose lengths are respectively less than different terms of the e-sequence.*

† *Loc. cit., supra,* § 1, footnote §.

with the point a, we then determine the first integer r_1 such that there is at least one of the given intervals of original length $\geqslant e_{r_1}$, which in its elongated form reaches at least to a. The integer r_1 is accordingly such that there is no such interval of length $\geqslant e_{r_1-1}$. That such an integer must exist follows from the fact that the given intervals include intervals with a as left-hand end-point. We then selected one of these intervals of length $\geqslant e_{r_1}$, and called it d_1. If a is not the left-hand end-point of d_1, there is a gap between a and d_1 of length $\leqslant d_1/n$, and therefore $\leqslant e_{r_1-1}/n$. If a is the left-hand end-point of d_1 there is no gap at a. In any case we then proceed on the right of d_1 as before. The gaps left are all precisely similar in character. Let us suppose, therefore, for definiteness, that there was a gap formed at a, and that X is a point internal to the gap, while H is the length of any one of the given intervals corresponding to X. Then, since the distance of X from a is less than the width of the gap, and therefore $< e_{r_1-1}/n$, the interval would, if its length were $\geqslant e_{r_1-1}$, be an interval of length $\geqslant e_{r_1-1}$ reaching, when elongated, at least as far as a, contrary to our choice of the integer r_1. Thus, the inequality (2) holds for every point X internal to the gap. The argument applies as it stands if X is the right-hand end-point of the gap. That it is true when X is the left-hand end-point a follows from the choice of r_1; for in the contrary case such an interval would be an interval $\geqslant e_{r_1-1}$ and would reach, of course, to a; thus, r_1 would not be the lowest integer of the type considered.

Thus, the inequality (2) is true, as stated. Moreover, the integer r_i in that inequality has been determined in accordance with the inequality (1).

We have thus shown that the gaps and intervals constructed at what we may call the first stage possess a certain property, as defined by the inequalities (1) and (2). At the end of this first stage the extreme gap on the right is of length less than e_1. It may also happen that the remaining gaps, taken in order from right to left, are respectively less than $e_2, e_3, \ldots, e_{m+1}$.

If so, our process is complete; if not, we stop at the first gap, say G_s, in order from right to left, whose length is too great. The gaps to the right of this we call *permanent gaps*, and the others *transitory gaps at the first stage.*

The gap G_s which has arrested us we treat in precisely the same way as we treated the whole interval, taking the quantity e_s instead of e_1 and replacing the whole e-sequence by that part of it which begins with e_s.

At the end of the second stage we shall then have at least one permanent gap added to those already obtained, and may have more.

In the third stage, if it be needed, we have to deal with gaps not made permanent lying in G_s, if there be any, and the transitory gaps at the first stage which remain transitory at the second stage. As we proceed steadily with our permanent gaps from right to left, the gaps other than G_s which were transitory at the first stage will necessarily remain transitory at the second stage, unless G_s has been replaced by permanent gaps only.

As our process proceeds we advance steadily from right to left, always adding at least one permanent gap to those already obtained and leaving no transitory gaps behind us. We must, then, in this way either obtain a finite number of permanent gaps which answer to our requirements or we must approach a limiting point X, say, of the permanent gaps, our approach being, of course, from the right.

This point X cannot be internal to any of the intervals chosen at any any one of the stages as we approach X; for if so that chosen interval would contain one of the permanent gaps of which X is the limit, and this is a contradiction in terms.

At each stage of our approach to X, therefore, X must belong to a transitory gap, either as internal or end-point. Now this is easily seen to be inconsistent with the inequalities (1) and (2).

In fact, the quantity H in inequality (2) may be supposed to be the length of an interval $(X, X+H)$ of the given set. This inequality shows that H can only be less than an infinite number of the quantities e_{r_i-1} if an infinite number of them coincide. In other words, there is an infinite succession of integers i for which r_i is the same. But to each such integer i we have, as the inequality (1) shows, a distinct interval d_i of length not less than the value assumed by an infinite number of the e_{r_i}'s. Thus, such d's must overlap. But it follows, from our mode of construction, that they do not overlap; thus, by a *reductio ad absurdum*, such a limiting point X cannot exist, and, therefore, at one stage all the gaps have been replaced by a finite number of permanent gaps answering to our requirements. Thus our lemma has been proved to be true.

3. In the preceding proof I have avoided the use of Cantor's numbers. The method of transfinite numbers is a powerful instrument of research. I propose to illustrate this fact by now deducing the lemma of the preceding article from the existence of a Cantor chain of intervals postulated by Lebesgue.

We suppose, then, a chain of intervals to stretch in the usual manner from the left-hand extremity of the interval of integration to the right-hand end, and we wish to show that we may omit all but a finite number of these intervals in such a manner as to obtain gaps of the lengths required by the lemma.

Starting from the right-hand end b, we can clearly take a gap of length less than e_1, and having its left-hand extremity coinciding with one of the right-hand end-points of the intervals of the chain. Since the semi-external points of the chain form a closed set, there is one of them which is nearest to the gap just formed of those on the left of the gap; this point is, moreover, not the left-hand extremity of that gap. We treat it in precisely the same way as we treated b, taking e_2 instead of e_1. Between the two gaps there is only a finite number of the intervals, because between them there is no limiting point of the original chain. Moreover, if we continue the process thus started, it is bound, after a finite number of steps, to come to an end. For if not the gaps so formed would necessarily have a limiting point to the left, which they would approach indefinitely near. This point would be a limiting point of the left-hand end-points of the gaps; that is, of end-points of intervals of the chain lying on its right. But the chain has only limiting points which are approached from the left, so that this is impossible. Thus, after a finite number of steps, we obtain the system of gaps and intervals required.*

4. We now pass to the following theorem on the nature of a derivate.

THEOREM.—*If $F(x)$ is a function which is upper semi-continuous on the right and lower semi-continuous on the left and has a derivate which is $\leqslant 0$ at a set whose complementary set has zero content, then either*

(i) *the derivate has the value $+\infty$ at a more than countable set of points, or*

(ii) *its values at all the points of content zero are also $\leqslant 0$, so that the function is a monotone nowhere increasing function, all of whose derivates are $\leqslant 0$.*

Suppose for definiteness that it is the lower right-hand derivate $F_+(x)$ which is known to be $\leqslant 0$, except at a set G of zero content, about which we only know that at all but a countable set of points $P_1, P_2, \ldots$, of G $F_+(x)$ is finite. At the points P_n the value is quite unknown, and may *a priori* be $+\infty$.

Let us attach to each point P_n an interval $(x, x+h)$ of length less than d,†

---

* The argument is a modification of that of Heine, used by Beppo Levi in dealing with the variation of a function, *loc. cit.*

† We may of course also secure that the *sum* of all these intervals is less than d. This is, however, not required in the present argument.

with P_n as left-hand end-point, and so small that

$$F(x+h)-F(x) \leqslant 2^{-n}e.$$

This is possible since $F(x)$ is upper semi-continuous on the right. Then the sum of the positive increments of $F(x)$ over any non-overlapping set of these intervals is less than e.

Next let us take any small positive quantity k, and let x be a point at which

$$(i-1)k < F_+(x) \leqslant ik,$$

where i is any positive integer.

Since these points form a sub-set of the exceptional set of zero content, we can enclose them all in a set of intervals of content less than $2^{-i}e/(i+1)k$. To each such point x let us attach an interval $(x, x+h)$ of length less than d, with x as left-hand end-point, contained in an interval of the enclosing set just constructed, and such that

$$\frac{F(x+h)-F(x)}{h} < (i+1)k.$$

Then the sum of the positive increments of $F(x)$ over any non-overlapping set of these intervals is less than $(i+1)k$ times the sum of their lengths; that is, it is less than $2^{-i}e$, since these intervals all lie inside the enclosing intervals of content less than $2^{-i}e/(i+1)k$.

Let us do this for every positive integer i.

Finally, let x be a point at which $F_+(x)$ is not positive, and let us attach to every such point an interval $(x, x+h)$ of length less than d and such that

$$\frac{F(x+h)-F(x)}{h} < k.$$

Then the sum of the positive increments of $F(x)$ over any non-overlapping set of these last intervals is less than $k(b-a)$, if (a, b) is the whole segment in which we are working.

We have now attached to each point of (a, b) an interval on the right of it of length less than d. Hence, by the lemma, we can choose out a finite number of these intervals, nowhere overlapping, and such that the complementary intervals are respectively less than different terms of any convenient monotone descending sequence $q_1, q_2, \ldots$, with zero as limit.

Let us choose the q-sequence so that $q_1 < d$, and in any interval $(y, y+k)$ of length less than q_n, $(n = 1, 2, \ldots)$,

$$F(y+k)-F(y) < 2^{-n}e,$$

which is possible, since $F(x)$ is upper semi-continuous on the right. Then the sum of the positive increments of $F(x)$ over these complementary intervals is less than e.

Also, by what has already been pointed out, the sum of the positive increments of $F(x)$ over the chosen intervals is less than

$$e+\sum_{i=1} 2^{-i}e+k(b-a) < 2e+k(b-a).$$

Now, since $F(x)$ is upper semi-continuous on the right and lower semi-continuous on the left, we can choose d so small that the positive increment of $F(x)$ over the finite set of intervals consisting of the chosen intervals and the complementary intervals has the positive variation P, finite or infinite, of $F(x)$ for limit as $d \to 0$.* Hence $P \leqslant 2e+k(b-a)$.

Thus, since k and e are as small as we please, P must be zero. This proves that $F(x)$ is a monotone non-increasing function of x, and hence that, under the circumstances assumed, (ii) is true. The alternative circumstances are given in (i). Thus the theorem is true.

5. The corresponding theorem with regard to the derivate of a function which is lower semi-continuous on the right and upper semi-continuous on the left follows, of course, by a mere change of sign. It may be well also to point out, though this is scarcely less obvious, that the theorems are equally true if we replace the particular derivate referred to in the enunciation by any of the other derivates, upper, lower, or intermediate, provided it is always on the right hand or always on the left hand. In other words, if we regard the right-hand derivates, for example, as defining a many-valued function, defined for the whole interval, we may take any of the one-valued functions which may be formed by taking at each point of the interval, by any law whatever, some one of the values of the many-valued function in question, the statements of this and the preceding article are still true.

Similar remarks apply to the other theorems about to be proved. In fact, in all cases the case considered in our enunciations is the most unfavourable case; or, if we prefer it, instead of deducing the remaining cases from the unfavourable case as an immediate consequence, we may content ourselves with remarking that the proof is, word for word, the same in such remaining cases.

6. If we prefer to use instead of the lemma of § 2 the theorems of Lebesgue, as extended by myself,† which state (a) that we can find a chain of

* "Functions of Bounded Variation," *Quarterly Journal of Mathematics*, Vol. XLII, § 5, p. 59.
† *Ibid.*, § 19 and § 18, pp. 72–74.

the given intervals placed so that each interval has for left-hand end-point a right-hand end-point of another interval or a limiting point of intervals lying on its left, every point of (a, b) being either an internal point or a left-hand end-point of the intervals of the chain, and (β) that the positive variation may be calculated by means of such chains, instead of only by means of divisions of (a, b) into a finite number of parts, we can give an alternative ending to the proof of § 4, beginning after the sentence: "We have now attached to each point of (a, b) an interval on the right of it of length less than d." We proceed as follows.

By Lebesgue's lemma we can choose out a chain of these intervals stretching from a to b. The sum of the positive increments of $F(x)$ over the intervals of the chain is, by what has been pointed out, less than

$$e+\sum_{i=1} 2^{-i}e+k(b-a) < 2e+k(b-a).$$

Now, since $F(x)$ is upper semi-continuous on the right and lower semi-continuous on the left, the positive increment of $F(x)$ over the intervals of the chain has, as d decreases indefinitely the positive variation P of $F(x)$ as limit, since all the intervals of the chain are of length less than d. Hence

$$P \leqslant 2e+k(b-a),$$

and therefore since e and k are as small as we please,

$$P=0.$$

This proves that $F(x)$ is a monotone non-increasing function of x, and that, under the circumstances assumed, (ii) is true. The alternative circumstances are given in (i). Thus this proves the theorem.

7. From the theorems given in §§ 4 and 5 the generalisations of Lebesgue's theorem stated in the introduction at once follow. For definiteness, take that which refers to a function which is upper semi-continuous on the right and lower semi-continuous on the left, and denote this function by $f(x)$. Let f_1 denote the integral of, say, the lower right-hand derivate over the set of points where that derivate is positive. Then f_1 possesses a differential coefficient except at a set of content zero. Moreover, its derivates are everywhere $\geqslant 0$. Hence the lower right-hand derivate of $f-f_1$ is nowhere greater than the corresponding derivate of f. Hence it assumes the value $+\infty$ at a countable set of points at most. Again, since f_1 is an integral, its differential coefficient agrees with the integrand, except at a set of content zero; it therefore agrees with the

lower right-hand derivate of f at points where the latter is $\geqslant 0$, with the exception at most of a set of content zero. Hence the lower right-hand derivate of $f-f_1$ is $\leqslant 0$, except at a set of content zero. Moreover, it assumes the value $+\infty$ at a countable set of points at most. Hence, by the theorem just proved, $f-f_1$ is a monotone nowhere increasing function, and therefore, since f_1 is an integral, f has the form stated in the enunciation. Finally, the positive variation of f is clearly the positive variation of f_1; that is, f_1 itself.

Thus our theorem is proved. Similarly the corresponding theorem is proved.