Brian S. Thomson
Mathematics Department,
Simon Fraser University,
B.C., Canada V5A 1S6
(e-mail: thomson@sfu.ca)

Characterization of an indefinite Riemann integral

Dedicated to Stefan Schwabik (1941–2009).

What are necessary and sufficient conditions in order that a function may be an indefinite integral in the Riemann sense? The problem has been explicitly posed in a short note [3] published by Erik Talvila in 2008 in This Exchange. Since neither Erik nor I have been able to find a solution in the literature I propose the following solution which is the sole subject of the paper.

The easiest route to a conjecture that might work for this problem is to compare it to a similar problem solved by Riesz [2] for functions of bounded variation. In order for a function $F : [a, b] \to \mathbb{R}$ to be represented in the form

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some function f that has bounded variation on $[a, b]$ it is necessary and sufficient that there is a constant K so that

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_{i-1})}{\xi_i - x_{i-1}} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| \le K \tag{1}$$

for every subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ and every choice of points $x_{i-1} < \xi_i \le \xi_i' < x_i$. This property has been labeled *bounded slope*

Key Words: Riemann integral, Lebesgue integral, indefinite integral, bounded slope variation.
Mathematical Reviews subject classification: 26A42, 26A16.

variation and has received some attention by later authors. This is more often expressed by placing a bound on the sums

$$\sum_{i=1}^{n-1}\left|\frac{F(x_{i+1})-F(x_i)}{x_{i+1}-x_i}-\frac{F(x_i)-F(x_{i-1})}{x_i-x_{i-1}}\right| \tag{2}$$

but the equivalent formulation here makes many computations more transparent. For details connecting the two expressions (1) and (2), see Ene [1, p. 719].

This characterization of Riesz, along with Riemann's own characterization of integrability, suggests a solution to the problem. Note that our condition (3) is easily implied by the stronger condition (1).

Theorem 1 *A necessary and sufficient condition in order for a function $F : [a,b] \to \mathbb{R}$ to be representable in the form*

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b)$$

for some constant C and for some Riemann integrable function f on $[a,b]$ is that, for all $\epsilon > 0$, a positive δ can be found so that

$$\sum_{i=1}^{n}\left|\frac{F(\xi_i)-F(x_{i-1})}{\xi_i-x_{i-1}}-\frac{F(x_i)-F(\xi_i')}{x_i-\xi_i'}\right|(x_i-x_{i-1}) < \epsilon \tag{3}$$

for every subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ that is finer than δ and every choice of associated points $x_{i-1} < \xi_i \le \xi_i' < x_i$.

Proof. For the proof that the condition is necessary let us suppose that F is the indefinite integral of a Riemann integrable function f. Let $\epsilon > 0$ and choose $\delta > 0$ so that

$$\sum_{i=1}^{n}\omega_f([x_{i-1},x_i])\,(x_i-x_{i-1}) < \epsilon$$

for every subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ that is finer than δ. Here

$$\omega_f([c,d]) = \sup\{|f(x)-f(y)| : x,\, y \in [c,d]\}$$

is used to denote the oscillation of the function f on a closed interval $[c,d]$. Since f is Riemann integrable this is possible (indeed it is one of Riemann's own characterizations of integrability).

Observe that, if $s \leq f(x) \leq t$ on an interval $[c,d]$, then

$$s - t \leq \frac{F(\xi) - F(c)}{\xi - c} - \frac{F(d) - F(\xi')}{d - \xi'} \leq t - s$$

for every $c < \xi \leq \xi' < d$. It follows that

$$\left| \frac{F(\xi) - F(c)}{\xi - c} - \frac{F(d) - F(\xi')}{d - \xi'} \right| \leq \omega_f([c,d]).$$

Consequently, using a subdivision $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ that is finer than δ,

$$\sum_{i=1}^{n} \left| \frac{F(\xi_i) - F(x_i)}{\xi_i - x_i} - \frac{F(x_i) - F(\xi_i')}{x_i - \xi_i'} \right| (x_i - x_{i-1})$$

$$\leq \sum_{i=1}^{n} \omega_f([x_{i-1}, x_i])\,(x_i - x_{i-1}) < \epsilon$$

proving (3) for any choice of associated points $x_{i-1} < \xi_i \leq \xi_i' < x_i$.

In the opposite direction we suppose $\epsilon > 0$ and that $\delta > 0$ has been chosen so that the condition (3) is satisfied for such subdivisions.

First we claim that F is Lipschitz. The argument that bounded slope variation implies Lipschitz is classical (cf. [1, p. 721]); this is closely related but requires some different details. We note that F must be bounded, even continuous, otherwise the condition (3) is easily violated. Suppose then that $|F(x)| < K$ for all $x \in [a,b]$.

Fix a number $0 < t < \delta$. We work in the interval $[a, b-t]$. For any $x \in [a, b-t]$ we use the interval $[x, x+t]$ and observe, for any $0 < h < t/2$, that

$$\left| \frac{F(x+h) - F(x)}{h} - \frac{F(x+t) - F(x+t/2)}{t/2} \right| (x + t - x) < \epsilon$$

because of the condition (3). Consequently

$$\left| \frac{F(x+h) - F(x)}{h} \right| < \frac{4K + \epsilon}{t}.$$

This imposes a bound on all the right-hand derived numbers of the continuous function F in the interval $[a, b-t]$. It follows that this bound also serves as a Lipschitz constant for F in $[a, b-t]$. By identical arguments, working on the left side, we can show that this same bound is a Lipschitz constant for F on the interval $[a+t, b]$. It follows that F is Lipschitz on $[a,b]$.

Since F is Lipschitz the derivative $F'(x)$ is a bounded function that exists at all points x in a set D having full measure in $[a,b]$ and F is an indefinite integral for F' in the Lebesgue sense. We define $f(x) = F(x)$ for $x \in D$ and, at points x not in D, we write

$$f(x) = \inf_{t>0} \sup\{F'(y) : y \in D,\ |x-y| < t\}.$$

Certainly

$$F(x) = C + \int_a^x f(t)\,dt \quad (a \le x \le b) \tag{4}$$

for some constant C, f is bounded and Lebesgue integrable. It remains only for us to prove that f is in fact a Riemann integrable function. To prove this we shall show that f is continuous at almost every point of $[a,b]$. It is enough to check that f is continuous at almost every point of the set D since the remaining points form a set of measure zero.

Let $\omega_f(x)$ denote the oscillation of the function f at a point x, i.e.,

$$\omega_f(x) = \inf_{t>0} \sup\{|f(x+h) - f(x)| : x+h \in [a,b],\ |h| < t\}.$$

The function f is continuous at a point x if and only if $\omega_f(x) = 0$. Thus the collection of discontinuity points of f can be expressed as the union of an increasing sequence of sets $\{E_m\}$ where

$$E_m = \{x \in [a,b] : \omega_f(x) > 1/m\} \quad (m = 1,2,3,\dots).$$

We show that each $|E_m| = 0$, i.e., that each is a set of Lebesgue measure zero.

For each $x \in D \cap E_m$ we may choose a sequence of nonzero numbers $h_n \to 0$ so that

$$|f(x+h_n) - f(x)| \ge 1/(2m).$$

By the way in which f was defined we may select these points so that $x + h_n$ are in D.

Thus for each point x that is in $D \cap E_m$ we may collect all the intervals of the form $[x,y]$ or $[y,x]$ with length smaller than δ and for which $y \in D$ and

$$|f(y) - f(x)| \ge 1/(2m).$$

This must form a Vitali cover of $D \cap E_m$.

By Vitali's theorem there is a disjoint collection $[x_1,y_1], [x_2,y_2], \dots, [x_p,y_p]$ chosen from the cover with the property that

$$|D \cap E_m| \le \sum_{k=1}^{p} (y_k - x_k) + \epsilon$$

and $0 < y_k - x_k < \delta$ and

$$|f(y_k) - f(x_k)| \geq 1/(2m) \quad (k = 1, 2, \dots, p).$$

For each $k = 1, 2, \dots, p$ select points ξ_k, ξ'_k with $x_k < \xi_k \leq \xi'_k < y_k$ in such a way that

$$\left| \frac{F(\xi_k) - F(x_k)}{\xi_k - x_k} - F'(x_k) \right| < \epsilon$$

and

$$\left| \frac{F(y_k) - F(\xi'_k)}{y_k - \xi'_k} - F'(y_k) \right| < \epsilon.$$

Now observe that

$$\frac{1}{2m}(y_k - x_k) \leq |f(y_k) - f(x_k)|(y_k - x_k) \leq$$

$$\left| \frac{F(\xi_k) - F(x_k)}{\xi_k - x_k} - F'(x_k) \right| (y_k - x_k) + \left| \frac{F(y_k) - F(\xi_k)}{y_k - \xi'_k} - F'(y_k) \right| (y_k - x_k)$$

$$+ \left| \frac{F(\xi_k) - F(x_k)}{\xi_k - x_k} - \frac{F(y_k) - F(\xi_k)}{y_k - \xi'_k} \right| (y_k - x_k).$$

But

$$\sum_{k=1}^{p} \left| \frac{F(\xi_k) - F(x_k)}{\xi_k - x_k} - \frac{F(y_k) - F(\xi_k)}{y_k - \xi'_k} \right| (y_k - x_k) < \epsilon$$

by the assumed condition (3). (This isn't a full subdivision of $[a, b]$ but the sum remains smaller than ϵ.)

The other inequalities we have imposed then show that

$$|D \cap E_m| \leq \sum_{k=1}^{p} (y_k - x_k) + \epsilon \leq (2m)\epsilon[2 + 2(b - a)].$$

As this argument works for any $\epsilon > 0$ it verifies the claim that $|D \cap E_m| = 0$ for each m. Thus the set of discontinuities of f in D have been expressed as the union of a sequence of sets of measure zero.

In particular we now know that f is continuous at almost every point of D and hence at almost every point of $[a, b]$. It is certainly bounded since F' is bounded by the Lipschitz constant for F. It follows that f is Riemann integrable and the representation in (4) can be interpreted in the Riemann sense. ∎

References

[1] V. Ene, Riesz type theorems for general integrals, Real Anal. Exchange 22, no. 2 (1997), 714-733.

[2] F. Riesz, Sur certain systèmes singulier d'equations intégrales, Annales de L'École Norm. Sup., Paris (3) 28 (1911), 33–62.

[3] E. Talvila, Characterizing integrals of Riemann integrable functions, Real Anal. Exchange 33, no. 2 (2007), 487-488.