has no zeros if $\alpha = 2$, and an infinity of real zeros, but no complex zeros, if $\alpha = 4, 6, 8, \ldots$.

[We have, by § 8.47,

$$F_\alpha(z) = \frac{1}{\alpha}\sum_{n=0}^{\infty}(-1)^n \frac{\Gamma\left(\frac{2n+1}{\alpha}\right)}{\Gamma(2n+1)} z^{2n}.$$

If $\alpha = 2$, we have, as in § 8.47,

$$F_2(z) = \tfrac{1}{2}\sqrt{\pi}e^{-\frac{1}{4}z^2},$$

which has no zeros.

If $\alpha = 2k$, where k is a positive integer, let

$$\phi(w) = \Gamma\{(2w+1)/2k\}\Gamma(w+1)/\Gamma(2w+1).$$

Then $\phi(w)$ is an integral function satisfying the conditions of Laguerre's theorems of § 8.6. Hence the zeros of

$$\sum_{n=0}^{\infty}\frac{\phi(n)}{n!}z^n = 2k\,F_{2k}(i\sqrt{z})$$

are all real and negative, so that the zeros of $F_{2k}(z)$ are all real. Also (§ 8.47) $\rho = 2k/2k-1$, so that $1 < \rho < 2$, and there must be an infinity of zeros.

If α is not an even integer, it can be proved that there are an infinity of complex zeros, and a finite number of real zeros.]

8.64. Functions with real negative zeros. If all the zeros of a function are real and negative, the modulus of the function is related to the distribution of its zeros in a specially simple way.

Suppose that $f(z)$ is such a function, and that its order ρ is less than 1. Then

$$f(z) = \prod_{n=1}^{\infty}\left(1+\frac{z}{z_n}\right).$$

Hence, if z is real,

$$\log f(z) = \sum_{n=1}^{\infty}\log\left(1+\frac{z}{z_n}\right) = \sum_{n=1}^{\infty} n\left\{\log\left(1+\frac{z}{z_n}\right)-\log\left(1+\frac{z}{z_{n+1}}\right)\right\}^*$$

$$= \sum_{n=1}^{\infty} n\int_{z_n}^{z_{n+1}}\frac{z\,dt}{t(z+t)} = z\int_0^{\infty}\frac{n(t)\,dt}{t(z+t)},$$

where $n(t)$ has its usual meaning.

Suppose now that as $t \to \infty$, $n(t) \sim \lambda t^\rho$. Then

$$\log f(x) \sim \pi\lambda \operatorname{cosec}\pi\rho\, x^\rho.$$

For we have $\qquad (\lambda-\epsilon)t^\rho < n(t) < (\lambda+\epsilon)t^\rho$

* The reader should justify this step, which is a simple example of partia summation.

for $t > t_0(\epsilon)$. Hence

$$\log f(x) < x \int_0^{t_0} \frac{n(t)\,dt}{t(x+t)} + x \int_{t_0}^{\infty} \frac{(\lambda+\epsilon)t^\rho}{t(x+t)}\,dt$$

$$= x \int_0^{t_0} \frac{n(t)-(\lambda+\epsilon)t^\rho}{t(x+t)}\,dt + x \int_0^{\infty} \frac{(\lambda+\epsilon)t^\rho}{t(x+t)}\,dt.$$

The first term is plainly $O(1)$, and, putting $t = xu$ in the second integral, we obtain

$$x^\rho(\lambda+\epsilon) \int_0^{\infty} \frac{u^{\rho-1}\,du}{1+u} = x^\rho(\lambda+\epsilon)\pi \operatorname{cosec} \pi\rho$$

by § 3.123. A similar result holds with $\lambda-\epsilon$, and the theorem follows.

*More generally,**

$$\log f(re^{i\theta}) \sim e^{i\rho\theta}\pi\lambda \operatorname{cosec} \pi\rho\, r^\rho$$

for any fixed θ in $(-\pi, \pi)$, $\log f(z)$ denoting the branch which is real on the positive real axis.

In fact the above expression for $\log f(z)$ as an integral, obtained for real z, holds by analytic continuation for $-\pi < \arg z < \pi$. Hence we obtain as before

$$\log f(re^{i\theta}) \sim re^{i\theta} \int_0^{\infty} \frac{\lambda t^\rho\,dt}{t(re^{i\theta}+t)}.$$

Turning the line of integration to $t = ue^{i\theta}$, we obtain

$$\lambda re^{\rho i\theta} \int_0^{\infty} \frac{u^\rho\,du}{u(r+u)} = \lambda r^\rho e^{\rho i\theta}\pi \operatorname{cosec} \pi\rho$$

as before.

It is also possible to prove theorems of the converse type, viz. to deduce the asymptotic behaviour of $n(r)$ from that of $\log|f(z)|$. The most interesting is that *if, as $x \to \infty$ by real values, $\log f(x) \sim \pi\lambda \operatorname{cosec} \pi\rho\, x^\rho$, then $n(r) \sim \lambda r^\rho$.* This theorem† is closely connected with the Tauberian results of §§ 7.41–7.44, but the proof is too complicated to give here.

* Pólya and Szegö, *Aufgaben*, IV Abschn., no. 61.

† See Valiron (1), Titchmarsh (5), (6).

8.7. The minimum modulus. Let $m(r)$ denote the minimum of $|f(z)|$ on the circle $|z|=r$.

The function $m(r)$ cannot be expected to behave in as simple a way as $M(r)$, since it vanishes whenever r is the modulus of a zero of $f(z)$. But we shall see that, if we exclude the immediate neighbourhood of these exceptional points, we can set a lower limit to $m(r)$; and, in general, $m(r)$ tends to zero in somewhat the same way as $1/M(r)$.

8.71. Consider first a canonical product $P(z)$ of order ρ, with zeros $z_1, z_2, \ldots, z_n, \ldots$.

If about each zero z_n $(|z_n|>1)$ we describe a circle of radius $|z_n|^{-h}$, where $h>\rho$, then in the region excluded from these circles

$$|P(z)|>e^{-r^{\rho+\epsilon}} \qquad (r>r_0(\epsilon)).$$

Following the method of § 8.25, it is clear that

$$\log|P(z)| \geqslant \sum_{r_n\leqslant kr} \log\left|1-\frac{z}{z_n}\right| - \sum_{r_n\leqslant kr} O\left\{\left(\frac{r}{r_n}\right)^p\right\} - \sum_{r_n>kr} O\left\{\left(\frac{r}{r_n}\right)^{p+1}\right\}$$

$$= \sum_{r_n\leqslant kr} \log\left|1-\frac{z}{z_n}\right| - O(r^{\rho+\epsilon}).$$

Since $\sum r_n^{-h}$ is convergent, the sum of the radii of the circles is finite, and so there are circles with centre the origin and arbitrarily large radius which lie entirely in the excluded region. Now if z lies outside every circle $|z-z_n|=r_n^{-h}$, and $r_n\leqslant kr$,

$$\left|1-\frac{z}{z_n}\right| > r_n^{-1-h} > (kr)^{-1-h}.$$

Hence
$$\sum_{1<r_n\leqslant kr} \log\left|1-\frac{z}{z_n}\right| > -(1+h)\log kr\,.\,n(kr)$$
$$> -K\log kr\,.\,r^{\rho+\epsilon} > -r^{\rho+2\epsilon}.$$

Finally
$$\sum_{r_n\leqslant 1} \log\left|1-\frac{z}{z_n}\right| > 0 \qquad (r>2),$$

and the result follows.

8.711. *If $f(z)$ is a function of order ρ, then*

$$m(r)>e^{-r^{\rho+\epsilon}}$$

on circles of arbitrarily large radius.

For
$$f(z)=P(z)e^{Q(z)},$$

where $Q(z)$ is a polynomial of degree $q \leqslant \rho$; hence

$$|e^{Q(z)}| > e^{-Ar^q} > e^{-Ar^\rho},$$

for sufficiently large values of r; and by the previous result

$$|P(z)| > e^{-r^{\rho_1+\epsilon}} > e^{-r^{\rho+\epsilon}}$$

on circles of arbitrarily large radius. Hence the result.

8.72. Another proof of Hadamard's factorization theorem. The theorem of § 8.71 leads to an alternative proof of Hadamard's factorization theorem. Let

$$f(z) = P(z)e^{Q(z)}$$

where $P(z)$ is the canonical product formed with the zeros of $f(z)$. Then $Q(z)$ is an integral function. Let ρ be the order of $f(z)$, ρ_1 the exponent of convergence. Then $P(z)$ is of order ρ_1, and $\rho_1 \leqslant \rho$. Hence

$$|P(z)| > e^{-r^{\rho_1+\epsilon}} > e^{-r^{\rho+\epsilon}}$$

on circles of arbitrarily large radius. Also

$$f(z) = O(e^{r^{\rho+\epsilon}}).$$

Hence $$\mathbf{R}\{Q(z)\} = \log\left|\frac{f(z)}{P(z)}\right| = \log\{O(e^{r^{\rho+\epsilon}})\} < Kr^{\rho+\epsilon}$$

on circles of arbitrarily large radius. Hence, by the theorem of § 2.54, $Q(z)$ is a polynomial of degree $\leqslant \rho$.

8.73. In special cases it is possible to prove much more precise results than the theorem of § 8.711.

If $\rho < \frac{1}{2}$, there is a sequence of values of r tending to infinity through which $m(r) \to \infty$.

In the first place, there is no line $\arg z = \text{constant}$ on which $f(z)$ is bounded; for the whole plane, bounded by this line, forms an angle 2π, and $2\pi < \pi/\rho$ if $\rho < \frac{1}{2}$. Hence, by § 5.61, if $f(z)$ is bounded on this line it is bounded everywhere, and so reduces to a constant.

Suppose now that

$$f(z) = cz^k \prod_{n=1}^{\infty}\left(1 - \frac{z}{z_n}\right),$$

and let $$\phi(z) = cz^k \prod_{n=1}^{\infty}\left(1 + \frac{z}{r_n}\right),$$

where $r_n = |z_n|$. Then

$$\min_{|z|=r} |f(z)| \geqslant |\phi(-r)|,$$

since

$$\left|1-\frac{z}{z_n}\right| \geqslant \left|1-\frac{r}{r_n}\right|$$

for every n. Also $\phi(-r)$ is unbounded, since $\phi(z)$ is an integral function of the same order as $f(z)$. This proves the theorem.

8.74. The following result is still more precise.

If $0 < \rho < 1$, there are arbitrarily large values of r for which

$$m(r) > \{M(r)\}^{\cos \pi\rho - \epsilon}.$$

The following proof is due to Pólya (3). Define $f(z)$ and $\phi(z)$ as before, and we may plainly take $c = 1$, $k = 0$. It is sufficient to prove the theorem for $\phi(z)$. If $0 < \rho < \frac{1}{2}$, i.e. $\cos \pi\rho > 0$, this follows at once from the relations $m(r) \geqslant |\phi(-r)|$, $M(r) \leqslant \phi(r)$. In any case, if z' is a point where $|f(z')| = m(r)$, we have

$$|\phi(r)\phi(-r)| = \left|\prod_{n=1}^{\infty}\left(1-\frac{r^2}{r_n^2}\right)\right| \leqslant |f(z')f(-z')| \leqslant m(r)M(r).$$

Hence, if the theorem is true for $\phi(z)$, then

$$m(r)M(r) \geqslant |\phi(r)|^{1+\cos \pi\rho - \epsilon} \geqslant \{M(r)\}^{1+\cos \pi\rho - \epsilon}$$

for arbitrarily large r, and the result for $f(z)$ follows.

If the theorem is false for $\phi(z)$, there are positive constants ϵ and a such that

$$\log|\phi(-x)| < (\cos \pi\rho - \epsilon)\log \phi(x) \qquad (x > a).$$

By § 8.4, ex. (xii), for $\rho < s < 1$, and so also for $\rho < \mathbf{R}(s) < 1$,

$$\int_0^\infty \{\cos \pi s \log \phi(x) - \log|\phi(-x)|\} x^{-s-1}\, dx = 0.$$

Since the integral over $(0, a)$ is regular for $0 < \mathbf{R}(s) < 1$, so is the integral over (a, ∞). Hence

$$F(s) = \int_\alpha^\infty \{\phi_1(e^\xi) + \psi(s)\phi_2(e^\xi)\} e^{-s\xi}\, d\xi,$$

where

$$\phi_1(x) = (\cos \pi\rho - \epsilon)\log \phi(x) - \log|\phi(-x)|, \qquad \phi_2(x) = \log \phi(x),$$

$$\psi(s) = \cos \pi s - \cos \pi\rho + \epsilon, \qquad \alpha = \log a,$$

is regular for $0 < \mathbf{R}(s) < 1$, and in particular at $s = \rho$. Here

ϕ_1 and ϕ_2 are positive for $x > a$, and $\psi(s)$ is positive for s real and sufficiently near to ρ.

Let $h > 0$, $D = d/ds$. Then

$$\left|\left(1-\frac{hD}{m}\right)^m F(s)\right| = \left|\sum_{\mu=0}^{m} \frac{(-h)^\mu}{\mu!} \frac{m(m-1)\ldots(m-\mu+1)}{m^\mu} F^\mu(s)\right|$$

$$\leqslant |F(s)| + \frac{h}{1!}|F'(s)| + \frac{h^2}{2!}|F''(s)| + \ldots = M,$$

say, the series being convergent for sufficiently small positive h and $s-\rho$. Also

$$\left(1-\frac{hD}{m}\right)^m \psi(s)e^{-s\xi} = e^{-s\xi}\left(1+\frac{h\xi-hD}{m}\right)^m \psi(s)$$

$$= e^{-s\xi}\left(1+\frac{h\xi}{m}\right)^m \psi(s) + e^{-s\xi} \sum_{\mu=1}^{m} \binom{m}{\mu}\left(1+\frac{h\xi}{m}\right)^{m-\mu}\left(\frac{-h}{m}\right)^\mu \psi^{(\mu)}(s)$$

$$\geqslant e^{-s\xi}\left(1+\frac{h\xi}{m}\right)^m \psi(s) - e^{-s\xi}\left(1+\frac{h\xi}{m}\right)^m \sum_{\mu=1}^{m} \frac{h^\mu}{\mu!}|\psi^\mu(s)|.$$

Since $|\psi^\mu(s)| \leqslant \pi^\mu$ for real s, this plainly exceeds

$$\tfrac{1}{2}e^{-s\xi}\left(1+\frac{h\xi}{m}\right)^m \psi(s)$$

if h is small enough. Hence

$$\left(1-\frac{hD}{m}\right)^m F(s) \geqslant \frac{1}{2}\int_\alpha^\infty \{\phi_1(e^\xi)+\psi(s)\phi_2(e^\xi)\}\left(1+\frac{h\xi}{m}\right)^m e^{-s\xi}\, d\xi.$$

In particular for any $\omega > \alpha$

$$\int_\alpha^\omega \phi_2(e^\xi)\left(1+\frac{h\xi}{m}\right)^m e^{-s\xi}\, d\xi \leqslant 2M/\psi(s).$$

Making $m \to \infty$, then $\omega \to \infty$, it follows that

$$\int_\alpha^\infty \phi_2(e^\xi)e^{(h-s)\xi}\, d\xi = \int_a^\infty \frac{\log \phi(x)}{x^{s-h+1}}\, dx$$

is convergent for a value of $s-h$ less than ρ. Hence $\sum r_n^{-s+h}$ is convergent, contrary to § 8.26. This proves the theorem.

8.75. A similar result holds for functions of order 1 and exponential type, i.e. such that $f(z) = O(e^{K|z|})$.

If $f(z) = O(e^{k|z|})$, *then* $m(r) > e^{-(k+\epsilon)r}$ *for some arbitrarily large* r. If $r_1, r_2, \ldots$ are the moduli of the zeros of $f(z)$, we find as in § 8.21 that $n(r) = O(r)$, $1/r_n \leqslant K/n$. Hence

$$\phi(z) = \prod_{n=1}^{\infty}\left(1+\frac{z}{r_n^2}\right) = O\prod_{n=1}^{\infty}\left(1+\frac{K^2r}{n^2}\right) = \frac{\sinh(\pi K\sqrt{r})}{\pi K\sqrt{r}}.$$

Define $h(\theta)$ (§ 5.7) for $\phi(z)$ with $V(r) = \sqrt{r}$. Then $h(\theta) \leqslant \pi K$ for all θ. Since $|\phi(z)| \geqslant 1$ if $\mathbf{R}(z) \geqslant 0$, $h(\theta)$ is finite for $-\frac{1}{2}\pi \leqslant \theta \leqslant \frac{1}{2}\pi$, and so everywhere (§ 5.712). Also $h(-\theta) = h(\theta)$; and § 5.713, with $\theta_1 = -\pi$, $\theta_2 = 0$, $\theta_3 = \pi$, $\rho = \frac{1}{2}$, gives $h(\pi) \geqslant 0$. Hence

$$|f(z)f(-z)| \geqslant |f(0)|^2|\phi(-r^2)| > e^{-\epsilon r}$$

for some arbitrarily large values of r, and the result follows.

8.8. The a-points of an integral function. Our discussion of integral functions has so far centred round the distribution of the zeros of the function. A more general question is that of the distribution of the points where the function takes any given value a—the 'a-points', as we may call them.

There is one case in which we have already obtained fairly precise results, namely, that of functions of finite non-integral order. If $f(z)$ is of order ρ, where ρ is not an integer, then it has an infinity of zeros, and the exponent of convergence of the zeros is ρ. But clearly $f(z)-a$ is also of order ρ, where a is any constant. Hence $f(z)$ has an infinity of a-points, and their exponent of convergence is ρ; i.e. their density is roughly the same for all values of a.

A similar argument may be applied to functions of zero order. Such a function has an infinity of zeros unless it reduces to a polynomial; and $f(z)-a$ is a polynomial for every value of a or for none.

If $f(z)$ is of positive integral order, and $\neq a$, then $f(z)-a = e^{Q(z)}$, where $Q(z)$ is a polynomial. If $b \neq a$, then $Q(z) = \log(b-a)$ for some z, i.e. $f(z) = b$ for some z. Hence $f(z)$ takes every value with one possible exception.

8.81. Picard's theorem. The main theorem of the subject is due to Picard; it is independent of any considerations of order.

An integral function which is not a polynomial takes every value, with one possible exception, an infinity of times.

Picard's proof of the theorem depends on the properties of the elliptic modular function. This function, which we shall

denote by $\varpi(z)$, has the following properties: *it is regular everywhere except at* $z = 0$, 1, *and* ∞; *and its imaginary part is never negative.*

By means of this function we can easily prove that *an integral function which is not a constant takes every value, with one possible exception, at least once.*

Suppose that $f(z)$ is an integral function which does not take either of the values a or b, where $a \neq b$. Then

$$g(z) = \frac{f(z)-a}{b-a}$$

is an integral function, which does not take either of the values 0 or 1. Consider the function $\varpi\{g(z)\}$. It is regular except at infinity, since $g(z)$ does not take either of the finite values for which ϖ is singular. Also its imaginary part is positive. Hence, by § 2.54, it is a constant. But ϖ is not constant, and so $g(z)$ must be constant. This proves the theorem.

As we have not discussed the construction of the modular function, we shall not complete this proof, but shall give a more direct proof, depending on a theorem of Schottky.*

8.82. The characteristic feature of Picard's theorem is that it admits the possibility of there being an exceptional value. This exceptional value may actually exist; for example, the function e^z never takes the value 0. A value with this property is said to be 'exceptional P'.

There is another sense in which a value may be exceptional. A function may take the value a, but only at points which have a convergence-exponent less than ρ. For example, the function $e^z \cos \sqrt{z}$, of order 1, has zeros, but their convergence exponent is $\frac{1}{2}$. A value with this property is said to be 'exceptional B', i.e. exceptional in the sense of Borel. It is clear that a value which is exceptional P is *a fortiori* exceptional B.

8.83. For functions of positive integral order, Picard's theorem is a consequence of the following theorem of Borel, which shows not merely that there is at most one value 'exceptional P', but at most one 'exceptional B'.

* Another direct proof, depending on a theorem of Bloch, is given by Landau, *Ergebnisse* . . ., ed. 2 (1929), Ch. VII, and by Dienes, *The Taylor Series*, Ch. VIII.

Borel's theorem. *If the order of $f(z)$ is a positive integer, then the exponent of convergence of the a-points of $f(z)$ is equal to the order, except possibly for one value of a.*

Suppose that there are two exceptional values, a and b. Then

$$f(z)-a = z^{k_1}e^{Q_1(z)}P_1(z) \tag{1}$$

and

$$f(z)-b = z^{k_2}e^{Q_2(z)}P_2(z), \tag{2}$$

where $Q_1(z)$ and $Q_2(z)$ are polynomials of degree ρ, and $P_1(z)$ and $P_2(z)$ are canonical products of order less than ρ.

Subtracting, we have

$$b-a = z^{k_1}e^{Q_1(z)}P_1(z)-z^{k_2}e^{Q_2(z)}P_2(z), \tag{3}$$

or

$$z^{k_1}P_1(z)e^{Q_1(z)-Q_2(z)} = z^{k_2}P_2(z)+(b-a)e^{-Q_2(z)}.$$

Since $Q_2(z)$ is of degree ρ, the right-hand side is of order ρ. Hence so is the left-hand side, and so $Q_1(z)-Q_2(z)$ is of degree ρ, since $P_1(z)$ is of order less than ρ.

Differentiating (3), we have

$$\begin{aligned}(z^{k_1}P_1Q_1'+k_1z^{k_1-1}P_1+z^{k_1}P_1')e^{Q_1}\\ = (z^{k_2}P_2Q_2'+k_2z^{k_2-1}P_2+z^{k_2}P_2')e^{Q_2}.\end{aligned} \tag{4}$$

Now the order of P_1' is the same as that of P_1, and so is less than ρ. Hence the coefficient of e^{Q_1} is of order less than ρ, and, similarly, so is that of e^{Q_2}. Hence we may write (4) in the form

$$z^{k_3}P_3e^{Q_1+Q_3} = z^{k_4}P_4e^{Q_2+Q_4},$$

where Q_3 and Q_4 are polynomials of degree $\rho-1$ at most, and P_3 and P_4 are canonical products. The two sides must have the same zeros, so that $k_3 = k_4$, $P_3 = P_4$, and so $Q_1+Q_3 = Q_2+Q_4$, i.e. $Q_1-Q_2 = Q_4-Q_3$, which is of degree less than ρ. This contradicts the previous result, that Q_1-Q_2 is of degree ρ, and so proves the theorem.

8.84. For the proof of Schottky's theorem we require the following lemma:

Let $\phi(r)$ be a real function of r for $0 \leqslant r \leqslant R_1$, and let

$$0 \leqslant \phi(r) \leqslant M \qquad (0 < r \leqslant R_1), \tag{1}$$

and also

$$\phi(r) < \frac{C\sqrt{\phi(R)}}{(R-r)^2} \qquad (0 < r < R \leqslant R_1). \tag{2}$$

Then

$$\phi(r) < \frac{AC^2}{(R_1-r)^4} \qquad (0 < r < R_1). \tag{3}$$

The actual form of the result (3) is not particularly important. What is important is that it depends only on r, R_1, and C, and not on M.

From (1) and (2) we obtain

$$\phi(r) < \frac{C\sqrt{M}}{(R-r)^2} \qquad (0 < r < R \leqslant R_1), \tag{4}$$

so that the upper bound M given in (1) is reduced at once to a multiple of $\sqrt{M}$. If we repeat the process, using first (4) with r_1, r_2 for r, R, and then (2) with r_1 for R, we obtain

$$\phi(r) < \frac{C}{(r_1-r)^2}\left\{\frac{C}{(r_2-r_1)^2}\right\}^{\frac{1}{2}} M^{\frac{1}{4}} \qquad (0 < r < r_1 < r_2 \leqslant R_1).$$

So generally

$$\phi(r) < \frac{C}{(r_1-r)^2}\left\{\frac{C}{(r_2-r_1)^2}\right\}^{\frac{1}{2}} \cdots \left\{\frac{C}{(r_n-r_{n-1})^2}\right\}^{\frac{1}{2^{n-1}}} M^{\frac{1}{2^n}}.$$

Taking $r_1 = \frac{1}{2}(R_1+r)$, $r_2 = \frac{1}{2}(R_1+r_1), \ldots$, this gives

$$\phi(r) < 4^{1+1+\frac{3}{4}+\cdots+\frac{n}{2^{n-1}}}\left\{\frac{C}{(R_1-r)^2}\right\}^{1+\frac{1}{2}+\cdots+\frac{1}{2^{n-1}}} M^{\frac{1}{2^n}},$$

and, making $n \to \infty$, the result follows.

8.85. Schottky's theorem. *If $f(z)$ is regular and does not take either of the values 0 or 1 for $|z| \leqslant R_1$, then for $|z| \leqslant R < R_1$*

$$|f(z)| < \exp\left\{\frac{KR_1^4}{(R_1-R)^4}\right\},$$

where K depends on $f(0)$ only. For all functions which satisfy the given conditions and are such that $\delta < |f(0)| < 1/\delta$, $|1-f(0)| > \delta$, we can take K to depend on δ only.

We shall not require the actual form of the upper bound for $f(z)$, which could be considerably improved if necessary; what is important is that it depends only on $f(0)$ in the manner stated, and on R/R_1.

Let $\qquad g_1(z) = \log\{f(z)\}, \qquad g_2(z) = \log\{1-f(z)\},$

where each logarithm has its principal value at $z = 0$. Then $g_1(z)$ and $g_2(z)$ are regular for $|z| \leqslant R_1$. Let $M_1(r)$ and $M_2(r)$ be the maxima of $|g_1(z)|$ and $|g_2(z)|$ respectively on $|z| = r$, and let

$$M(r) = \max\{M_1(r), M_2(r)\}.$$

Let $\qquad B_1(r) = -\min\limits_{|z|=r} \mathrm{R}\{g_1(z)\} = \max\limits_{|z|=r} \log \dfrac{1}{|f(z)|}.$

Then Carathéodory's theorem (§ 5.5), applied to $g_1(z)$, gives

$$M_1(\rho) \leqslant \frac{2\rho}{r-\rho} B_1(r) + \frac{r+\rho}{r-\rho} |g_1(0)| \quad (0 < \rho < r). \tag{1}$$

There are now two possibilities. Either $B_1(r)$ is not large—say $B_1(r) \leqslant 1$, in which case (1) is a result of the required type; or $B_1(r)$ is large, in which case there is a point z' on $|z| = r$ where $|f(z')|$ is small. But if $|f(z')|$ is small, $g_2(z')$ is (apart from a term $2n\pi i$) approximately equal to $-f(z')$; and then Carathéodory's theorem, applied to $\log g_2$, gives on the left M_1 (not $\log M_1$ as we should in general expect), and on the right $\log M_2 = O(\sqrt{M_2})$. We thus obtain an inequality of the type considered in the elementary lemma. This is the general plan of the proof, and we now proceed to fill in the details.

Suppose that $B_1(r) > 1$, and let z' be a point where

$$B_1(r) = \log 1/|f(z')|.$$

Then
$$|f(z')| = e^{-B_1(r)} < e^{-1} < \tfrac{1}{2}. \tag{2}$$

There is therefore an integer n such that

$$g_2(z') - 2n\pi i = -\sum_{m=1}^{\infty} \frac{\{f(z')\}^m}{m}.$$

Hence
$$|g_2(z') - 2n\pi i| < \sum_{m=1}^{\infty} 2^{-m} = 1,$$

and so
$$2|n|\pi < 1 + |g_2(z')| \leqslant 1 + M_2(r). \tag{3}$$

Let
$$h(z) = \log\{g_2(z) - 2n\pi i\},$$

where the logarithm has its principal value at $z = 0$. Then $h(z)$ is regular for $|z| \leqslant R_1$, since $f(z) \neq 0$ and so $g_2(z) \neq 2n\pi i$; and Carathéodory's theorem gives

$$\max_{|z|=r} |h(z)| \leqslant \frac{2r}{R-r} \max_{|z|=R} \log|g_2(z) - 2n\pi i| + \frac{R+r}{R-r} |h(0)|. \tag{4}$$

The left-hand side is not less than

$$\log\left|\frac{1}{g_2(z') - 2n\pi i}\right| \geqslant \log \frac{1}{|f(z')| + |f(z')|^2 + |f(z')|^3 + \ldots}$$
$$\geqslant \log \frac{1}{2|f(z')|} = B_1(r) - \log 2,$$

by (2). On the right we have

$$\max_{|z|=R} \log|g_2(z) - 2n\pi i| \leqslant \log\{M_2(R) + 2|n|\pi\} < \log\{2M_2(R) + 1\},$$

by (3); if $n \neq 0$, $|g_2(0)-2n\pi i| \geqslant \pi > 1$, and so

$$|h(0)| \leqslant \log|g_2(0)-2n\pi i|+\pi \leqslant \log\{|g_2(0)|+1+M_2(r)\}+\pi.$$

If $n = 0$, $|h(0)| \leqslant |\log|g_2(0)||+\pi$. Hence (4) gives

$$B_1(r) \leqslant \frac{2R}{R-r}[2\log\{2M_2(R)+|g_2(0)|+1\}+|\log|g_2(0)||+\pi]+\log 2$$

$$< \frac{4R}{R-r}[\log\{2M_2(R)+|g_2(0)|+1\}+|\log|g_2(0)||+\pi]. \qquad (5)$$

This inequality, proved for $B_1(r) > 1$, is obviously true for $B_1(r) \leqslant 1$. Hence (1) and (5) give in any case

$$M_1(\rho) < \frac{8Rr}{(R-r)(r-\rho)}[\log\{2M_2(R)+|g_2(0)|+1\}+$$
$$+|\log|g_2(0)||+|g_1(0)|+\pi].$$

Since we may interchange $g_1(z)$ and $g_2(z)$ in the whole argument, the inequality is still true if the suffixes 1 and 2 are interchanged. Combining the two results, we have

$$M(\rho) < \frac{8Rr}{(R-r)(r-\rho)}\{\log M(R)+K\},$$

where K depends on $|g_1(0)|$ and $|g_2(0)|$ only. Taking $r = \frac{1}{2}(R+\rho)$, we obtain

$$M(\rho) < \frac{32R_1^2}{(R-\rho)^2}\{\log M(R)+K\} < \frac{KR_1^2\sqrt{M(R)}}{(R-\rho)^2},$$

since $\log M(R) = O\{\sqrt{M(R)}\}$. Hence, by the lemma,

$$M(\rho) < \frac{KR_1^4}{(R-\rho)^4},$$

and
$$|f(z)| \leqslant e^{M(r)} < \exp\left\{\frac{KR_1^4}{(R-r)^4}\right\}.$$

Since K depends on $|g_1(0)|$ and $|g_2(0)|$ only, the last part of the theorem also is true.

8.86. Picard's first theorem. *An integral function which is not constant takes every value, with one possible exception, at least once.*

Suppose that $f(z)$ does not take either of the values a or b $(a \neq b)$. Then $g(z) = \{f(z)-a\}/(b-a)$ does not take either of the values 0 or 1. Hence, by Schottky's theorem,

$$|g(z)| < \exp\left\{\frac{KR_1^4}{(R_1-R)^4}\right\} \qquad (|z| \leqslant R < R_1).$$

Taking $R_1 = 2R$, $|g(z)| < K$. Hence $g(z)$ is a constant.

8.87. We can also prove the following generalization of Picard's theorem.

Landau's theorem.* *If α is any number, and β any number other than 0, there is a number $R = R(\alpha, \beta)$ such that every function*

$$f(z) = \alpha + \beta z + a_2 z^2 + a_3 z^3 + \ldots,$$

regular for $|z| \leqslant R$, takes in this circle one of the values 0 or 1.

We may suppose that $\alpha \neq 0$, $\alpha \neq 1$, for otherwise we have the result at once. If $f(z)$ does not take either of the values 0 or 1, then by Schottky's theorem $|f(z)| < K(\alpha)$ for $|z| \leqslant \frac{1}{2}R$. Hence

$$|\beta| = \left| \frac{1}{2\pi i} \int\limits_{|z| = \frac{1}{2}R} \frac{f(z)}{z^2}\, dz \right| \leqslant \frac{K(\alpha)}{\frac{1}{2}R},$$

$$R \leqslant 2K(\alpha)/|\beta|,$$

and the result follows.

8.88. We have so far stated Picard's theorem in terms of integral functions, i.e. functions with an essential singularity at infinity. But a corresponding theorem holds for any function with an isolated essential singularity.

Picard's second theorem. *In the neighbourhood of an isolated essential singularity, a one-valued function takes every value, with one possible exception, an infinity of times.*

In other words, *if $f(z)$ is regular for $0 < |z - z_0| < \rho$, and there are two unequal numbers a, b, such that $f(z) \neq a$, $f(z) \neq b$, for $|z - z_0| < \rho$, then z_0 is not an essential singularity.*

We may suppose that $z_0 = 0$, $\rho = 1$, $a = 0$, and $b = 1$. We prove that there is a sequence of circles $|z| = r_n$, where $r_n \to 0$, on which $f(z)$ is bounded. By § 2.71 this precludes the existence of a singularity at $z = 0$.

We start from Weierstrass's theorem that, in the neighbourhood of an essential singularity, a function approaches arbitrarily near to any given value an infinity of times. Thus there is a sequence of points z_1, $z_2, \ldots$ such that $|z_1| > |z_2| > \ldots$, $|z_n| \to 0$, and

$$|f(z_n) - 2| < \tfrac{1}{2}. \tag{1}$$

It is clear that Schottky's theorem would enable us to construct a sequence of circles, with these points as centres, in which $f(z)$ is bounded. These circles do not, of course, include

* Landau (1), and *Ergebnisse*, § 25.

the origin; but this is, so to speak, an accident arising from the fact that we have proved Schottky's theorem for a class of convex curves (viz. circles). We can remove the difficulty by making a conformal transformation, which has the effect of replacing a circle by an elongated curve which, though it excludes the origin, passes right round it and overlaps itself on the far side.

Let $z = e^w$ $(w = u+iv)$, and consider the half-strip of the w-plane $u < 0$, $-\pi \leqslant v \leqslant \pi$. This corresponds to the interior of the circle $|z| = 1$. Let $w_n = \log z_n$ $(-\pi < \mathbf{I}(w_n) \leqslant \pi)$, so that $\mathbf{R}(w_n) \to -\infty$; and let $f(z) = g(w)$.

We apply Schottky's theorem to the function

$$h(w') = g(w_n + w').$$

This is regular for $|w'| \leqslant 4\pi$ if n is large enough, and it does not take either of the values 0 or 1. Hence

$$|h(w')| < K = K\{h(0)\} \qquad (|w'| \leqslant 2\pi);$$

and, the numbers $h(0) = g(w_n) = f(z_n)$ satisfying (1), we can replace the right-hand side by an absolute constant. Hence $|g(w)| < A$ for $|w - w_n| \leqslant 2\pi$, and in particular for $u = \mathbf{R}(w_n)$, $-\pi \leqslant v \leqslant \pi$. Hence

$$|f(z)| < A \qquad (|z| = |z_n|),$$

and the result follows.

8.89. Asymptotic values. A number a is said to be an asymptotic value of an integral function $f(z)$ if there is a continuous curve from a given point to infinity, i.e. along which $|z| \to \infty$, and along which $f(z) \to a$ as $z \to \infty$. Thus 0 is an asymptotic value of e^z, since $e^z \to 0$ as $z \to \infty$ along the negative real axis. The function

$$\int_0^z e^{-t^q}\, dt,$$

where q is a positive integer, has the q asymptotic values

$$e^{\frac{2\pi ik}{q}} \int_0^\infty e^{-t^q}\, dt \qquad (k = 0, 1, \ldots, q-1),$$

as $z \to \infty$ along the lines $\arg z = e^{2\pi ik/q}$.

We may define the 'asymptotic value ∞' similarly.

We shall now prove the following theorems.

Every function with an isolated essential singularity at infinity, which is not a constant, has the asymptotic value infinity.

By Laurent's theorem, such a function is of the form $f(z)+g(z)$, where $f(z)$ is an integral function, and $g(z)$ tends uniformly to a limit as $|z| \to \infty$. Hence it is sufficient to consider integral functions. For an integral function not a constant, the maximum modulus $M(r)$ tends steadily to infinity. Consider an indefinitely increasing sequence of numbers $X_1 = M(r_1)$, $X_2, \ldots$. It follows from Liouville's theorem that there is a point outside the circle $|z| = r_1$ at which $|f(z)| > X_1$. The set of points where $|f(z)| > X_1$ constitutes the interior of one or more regions bounded by curves on which $|f(z)| = X_1$; and these regions must be exterior to the circle $|z| = r_1$. Let one such region be D_1. Now D_1 must extend to infinity; for otherwise we should have a finite region with $|f(z)| = X_1$ on the boundary and $|f(z)| > X_1$ inside, contrary to the maximum-modulus theorem. Further, $f(z)$ is unbounded in D_1. For otherwise the principle of Phragmén and Lindelöf would also show that $|f(z)| \leqslant X_1$ at all points inside D_1. In fact, the argument of § 5.6 applies with P at infinity and $\omega(z) = r_1/z$. Hence there is a point of D_1 at which $|f(z)| > X_2$, and consequently a domain D_2, interior to D_1, such that $|f(z)| > X_2$ at all points of D_2. We can now repeat the argument with $X_3, \ldots$. Hence there is a sequence of infinite regions $D_1, D_2, \ldots$, each interior to the preceding one and such that $|f(z)| > X_m$ in D_m, and $|f(z)| = X_m$ on its boundary. Now, take a point on the boundary of D_1, and join it to a point on the boundary of D_2 by a continuous curve lying in D_1; then this point to a point of D_3 by a continuous curve lying in D_2, and so on. We clearly obtain a continuous curve along which $f(z) \to \infty$.

If an integral function does not iake the value a, then a is an asymptotic value.

For $1/\{f(z)-a\}$ is an integral function, and so has the asymptotic value ∞.

The argument of § 5.64 shows that if an integral function has asymptotic values on two curves, and is bounded between the curves, then these asymptotic values must be the same. Asymptotic values not so connected we should consider as distinct, whether they are equal or not.

It was conjectured by Denjoy that *an integral function of finite order ρ can have at most 2ρ asymptotic values.* The theorem with 5ρ instead of 2ρ was proved by Carleman; and Denjoy's conjecture was finally proved by Ahlfors. The general proof is not easy. It is, however, easy to see that there can be at most 2ρ straight lines from 0 to ∞ along which a function of order ρ has distinct asymptotic values. For by § 5.61 the angle between two such lines must be at least equal to π/ρ.

8.9. Meromorphic functions. We shall now give a short introduction to the theory of meromorphic functions, i.e. functions whose only singularities, except at infinity, are poles.

The theory depends largely on the general Jensen formula (§ 3.61 (4)). Let $f(z)$ be a meromorphic function, with zeros $a_1, a_2, \ldots$ and poles $b_1, b_2, \ldots$ (other than 0) arranged with non-decreasing moduli. Suppose that in the neighbourhood of the origin it is of the form $cz^k + \ldots$, where k may be any integer. Then Jensen's formula for $z^{-k}f(z)$ is

$$\log\left|\frac{b_1\ldots b_n}{a_1\ldots a_m}\right| r^{m-n} + \log|c| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta - k\log r.$$

As in § 3.61

$$\log\frac{r^m}{|a_1\ldots a_m|} = \sum_{\nu=1}^{m-1} \nu \int_{|a_\nu|}^{|a_{\nu+1}|} \frac{dx}{x} + m\int_{|a_m|}^{r} \frac{dx}{x}.$$

Let $n(r, 0)$ be the number of zeros of $f(z)$ in $|z| \leqslant r$. If $k > 0$, $\nu = n(x, 0) - k$ for $|a_\nu| \leqslant x < |a_{\nu+1}|$; hence

$$\log\frac{r^m}{|a_1\ldots a_m|} = \int_0^r \frac{n(x,0)-k}{x}\,dx.$$

If $n(r, \infty)$ is the number of poles of $f(z)$ in $|z| \leqslant r$, we obtain similarly

$$\log\frac{r_n}{|b_1\ldots b_n|} = \int_0^r \frac{n(x,\infty)}{x}\,dx.$$

If k is negative, it appears in the second integral instead of the first. Writing

$$N(r, a) = \int_0^r \frac{n(x,a)-n(0,a)}{x}\,dx + n(0,a)\log r$$

we obtain in any case

$$N(r,0)-N(r,\infty) = \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\,d\theta - \log|c|. \qquad (1)$$

Let us write $\log^+\alpha = \max(\log\alpha, 0)$

for any positive α. Thus

$$\log\alpha = \log^+\alpha - \log^+ 1/\alpha.$$

Let $$m(r,a) = m\left(r, \frac{1}{f-a}\right) = \frac{1}{2\pi}\int_0^{2\pi} \log^+\left|\frac{1}{f(re^{i\theta})-a}\right| d\theta$$

and $$m(r,\infty) = m(r,f) = \frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|\,d\theta.$$

Then (1) may also be written

$$m(r,0)+N(r,0) = m(r,\infty)+N(r,\infty)-\log|c|. \qquad (2)$$

Now apply this formula to $f(z)-a$, where a is any number. If $f(z)-a = c_a z^k + \ldots$ in the neighbourhood of the origin, we obtain

$$m(r,a)+N(r,a) = m(r,f-a)+N(r,\infty)-\log|c_a|,$$

the term $N(r,\infty)$ being unaltered since the poles of $f(z)-a$ are the same for each a.

Now we have

$$|f|+|a| \leqslant 2|fa|, \quad 2|f|, \quad 2|a| \quad \text{or} \quad 2$$

according as $|f| \geqslant 1$ and $|a| \geqslant 1$, $|f| \geqslant 1$ and $|a| < 1$, $|f| < 1$ and $|a| \geqslant 1$, or $|f| < 1$ and $|a| < 1$. Hence

$$\log(|f|+|a|) \leqslant \log^+|f|+\log^+|a|+\log 2.$$

Hence $$\log^+|f-a| \leqslant \log^+|f|+\log^+|a|+\log 2,$$

and similarly

$$\log^+|f| \leqslant \log^+|f-a|+\log^+|a|+\log 2.$$

Hence $$|m(r,f-a)-m(r,f)| \leqslant \log^+|a|+\log 2.$$

It follows that

$$m(r,a)+N(r,a) = m(r,\infty)+N(r,\infty)+\phi(r,a),$$

where $$|\phi(r,a)| \leqslant |\log|c_a||+\log^+|a|+\log 2.$$

Hence if $f(z)$ is a meromorphic function and not a constant, the

values of the sum
$$m(r,a)+N(r,a)$$
for two given values of a differ by a bounded function of r.

All the sums being to this extent equivalent, we can represent them all, e.g. by the one with $a = \infty$. Thus *if we put*
$$T(r) = m(r,\infty)+N(r,\infty), \tag{3}$$
we have for all values of a
$$m(r,a)+N(r,a) = T(r)+\phi(r,a),$$
where $\phi(r,a)$ *is (for each* a*) bounded as* $r \to \infty$.

$T(r)$ is called the *characteristic function* of $f(z)$.

We shall next show that $T(r)$ *is an increasing convex function of* $\log r$.

Jensen's formula for $f(z)-e^{i\lambda}$ (λ real) is
$$N(r,e^{i\lambda})-N(r,\infty) = \frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})-e^{i\lambda}|\,d\theta - \log|f(0)-e^{i\lambda}|, \tag{4}$$
if $f(0) \neq e^{i\lambda}$. Also, for any a,
$$\frac{1}{2\pi}\int_0^{2\pi} \log|e^{i\theta}-a|\,d\theta = \log^+|a|,$$
e.g. by Jensen's theorem with $f(z) = z-a$, $r = 1$. Hence, multiplying (4) by $1/(2\pi)$, and integrating with respect to λ over $(0, 2\pi)$, we obtain
$$\frac{1}{2\pi}\int_0^{2\pi} N(r,e^{i\lambda})\,d\lambda - N(r,\infty) = \frac{1}{2\pi}\int_0^{2\pi} \log^+|f(re^{i\theta})|\,d\theta - \log^+|f(0)|,$$
i.e.
$$T(r) = \frac{1}{2\pi}\int_0^{2\pi} N(r,e^{i\lambda})\,d\lambda + \log^+|f(0)|.$$

Now, for any a, $N(r,a)$ is an increasing convex function of r, since
$$\frac{dN(r,a)}{d\log r} = n(r,a),$$
which is non-negative and non-decreasing. Hence $T(r)$ has the same property.

In the above formulae $N(r,a)$ measures the number of times the function $f(z)$ takes the value a. Since the largest contribu-

tions to $m(r, a)$ come from arcs where $f(z)$ is nearly equal to a, $m(r, a)$ measures in a sense the intensity of the approximation of $f(z)$ to a. We could describe $m(r, a)+N(r, a)$ as the total affinity of the function $f(z)$ for the value a.

For a given function, certain values may be exceptional, e.g. in the sense that the function does not take these values. The above theorem shows that there can be no exceptional values in the sense that the total affinity of the function for every value is the same, apart from bounded functions of r.

Examples. (i) Let $f(z)$ be a rational function, say $= P(z)/Q(z)$, where $P(z)$ is of degree μ, $Q(z)$ of degree ν, P and Q having no common factor. If $\mu > \nu$, then

$$m(r, a) = O(1), \qquad N(r, a) = \mu \log r + O(1)$$

for every finite a, while

$$m(r, \infty) = (\mu-\nu)\log r + O(1), \qquad N(r, \infty) = \nu \log r + O(1).$$

If $\mu < \nu$, then

$$m(r, a) = O(1), \qquad N(r, a) = \nu \log r + O(1)$$

for $a \neq 0$, while

$$m(r, 0) = (\nu-\mu)\log r + O(1), \qquad N(r, 0) = \mu \log r + O(1).$$

If $\mu = \nu$, let a_0 and b_0 be the coefficients of x^μ in P and Q. Then if $a \neq a_0/b_0$,

$$m(r, a) = O(1), \qquad N(r, a) = \mu \log r + O(1),$$

while, if $a_0 Q - b_0 P$ is of degree α,

$$m\left(r, \frac{a_0}{b_0}\right) = (\mu-\alpha)\log r + O(1), \qquad N\left(r, \frac{a_0}{b_0}\right) = \alpha \log r + O(1).$$

In any case $T(r) = O(\log r)$.

(ii) The function e^z does not take the values 0 or ∞; on the other hand these are limiting values of the function as $z \to \infty$. Here

$$N(r, 0) = N(r, \infty) = 0, \qquad m(r, 0) = m(r, \infty) = \frac{r}{\pi},$$

while for $a \neq 0, \infty$,

$$m(r, a) = O(1), \qquad N(r, a) = \frac{r}{\pi} + O(1).$$

Here $T(r) = r/\pi$.

(iii) Consider similarly $\tan z$ ($\pm i$ are exceptional values).

8.91. Order of a meromorphic function. The meromorphic function $f(z)$ is said to be of order ρ if

$$\overline{\lim_{r\to\infty}} \frac{\log T(r)}{\log r} = \rho,$$

so that

$$T(r) = O(r^{\rho+\epsilon})$$

for every positive ϵ, but not for $\epsilon < 0$.

To show that this agrees with the definition of order in the case of an integral function, we shall prove:

If $f(z)$ is an integral function,

$$T(r) \leqslant \log^+ M(r) \leqslant \frac{R+r}{R-r} T(R)$$

for $0 < r < R$.

For an integral function, $N(r, \infty) = 0$, and $T(r) = m(r, \infty)$. The left-hand inequality is thus

$$\frac{1}{2\pi} \int_0^{2\pi} \log^+ |f(re^{i\theta})| \, d\theta \leqslant \log^+ \max |f(re^{i\theta})|,$$

which is plainly true. Also by the Poisson-Jensen formula

$$\log|f(re^{i\theta})| = \frac{1}{2\pi} \int_0^{2\pi} \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta-\phi) + r^2} \log|f(Re^{i\phi})| \, d\phi - \\ - \sum_{\mu=1}^{m} \log \left| \frac{R^2 - \overline{a_\mu} re^{i\theta}}{R(Re^{i\theta} - a_\mu)} \right|.$$

Each term in $\sum$ is negative, and

$$R^2 - 2Rr\cos(\theta-\phi) + r^2 \geqslant (R-r)^2.$$

Hence, taking θ so that the left-hand side is a maximum,

$$\log|M(r)| \leqslant \frac{R+r}{R-r} \frac{1}{2\pi} \int_0^{2\pi} \log|f(Re^{i\phi})| \, d\phi \\ \leqslant \frac{R+r}{R-r} T(R)$$

and the second inequality follows.

Taking $R = 2r$, the identity of the two definitions of the order of an integral function is clear.

Now let $r_n(a)$ be the moduli of the zeros of $f(z) - a$, $r_n(\infty)$ the moduli of the poles of $f(z)$. Then we have the following results.

If $f(z)$ is of order ρ, then for every a

$$m(r, a) = O(r^{\rho+\epsilon}), \qquad N(r, a) = O(r^{\rho+\epsilon}), \qquad n(r, a) = O(r^{\rho+\epsilon})$$

and

$$\sum \left(\frac{1}{r_n(a)} \right)^{\rho+\epsilon}$$

is convergent.

The first two results are immediate since

$$m(r,a) \leqslant T(r)+O(1), \qquad N(r,a) \leqslant T(r)+O(1).$$

The remaining results then follow from that on $N(r,a)$ as in the case of an integral function.

More precise results of the same kind are given by Nevanlinna, *Fonctions Méromorphes*, Ch. II.

8.92. Factorization of meromorphic functions. Let $f(z)$ be a meromorphic function of order ρ, with zeros a_n and poles b_n ($f(0) \neq 0$). Then it follows from the above results that there are integers p_1 and p_2 not exceeding ρ such that

$$P_1(z) = \prod_{n=1}^{\infty} E\left(\frac{z}{a_n}, p_1\right), \qquad P_2(z) = \prod_{n=1}^{\infty} E\left(\frac{z}{b_n}, p_2\right)$$

are convergent for all values of z. Hence $P_1(z)$ and $P_2(z)$ are integral functions of order not exceeding ρ. Also

$$f_1(z) = f(z)P_2(z)$$

is an integral function. Now

$$\begin{aligned} T(r,f_1) = m(r,\infty,f_1) &\leqslant m(r,\infty,f)+m(r,\infty,P_2) \\ &\leqslant T(r,f)+T(r,P_2) = O(r^{\rho+\epsilon})+O(r^{\rho+\epsilon}). \end{aligned}$$

Hence $f_1(z)$ is of order ρ at most, and hence

$$f_1(z) = e^{Q(z)}P_1(z),$$

where $Q(z)$ is a polynomial of degree not exceeding ρ.

We have thus proved that

$$f(z) = e^{Q(z)}P_1(z)/P_2(z),$$

an extension of Hadamard's factorization theorem to meromorphic functions.

A slightly deeper theorem, in which the numerator and denominator do not necessarily converge separately, is proved by Nevanlinna, *Fonctions Méromorphes*, Ch. III.

Further developments of the theory of meromorphic functions are largely concerned with extensions of the theorems of Picard and Borel. For these we must refer to the books of Nevanlinna.

MISCELLANEOUS EXAMPLES

1. Prove that, if a is not a multiple of π,

$$\sin(a-z) = \sin a\, e^{-z\cot a} \prod_{n=-\infty}^{\infty} \left(1-\frac{z}{a+n\pi}\right) e^{\frac{z}{a+n\pi}}.$$

2. Show that the equations

$$\sin z = z^2, \qquad \log z = z^3, \qquad \tan z = az+b,$$

where a and b are any complex numbers, each have an infinity of roots.

3. Find all the zeros of the function

$$e^{e^z}-1,$$

and show that they have no finite exponent of convergence.

4. If $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is a function of non-integral order, show that the coefficients in the polynomial $Q(z) = b_1 z + \ldots + b_q z^q$ of Hadamard's theorem can be expressed in terms of $a_1, a_2, \ldots, a_q$.

[If ρ is not an integer, $q < p+1$, and $P(z) = 1+O(z^{p+1})$ as $z \to 0$. Hence on equating coefficients $P(z)$ is not involved.]

5. If $f(z) = \sum a_n z^n$ is of order ρ, what is the order of $F(z) = \sum |a_n|^p z^n$?

6. The generalized hypergeometric function is defined by the formula

$${}_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;z) = \sum_{n=0}^{\infty} \frac{(\alpha_1)_n\cdots(\alpha_p)_n}{(\beta_1)_n\cdots(\beta_q)_n}\frac{z^n}{n!},$$

where $(\alpha)_n = \alpha(\alpha+1)\ldots(\alpha+n)$, $(\alpha)_0 = 1$. Show that it is an integral function if $q \geqslant p$, and find its order.

7. Show that

$$f(z) = \sum_{n=-\infty}^{\infty} q^{n^k} e^{inz}$$

is an integral function if $|q| < 1$, $k > 1$, and find its order.

8. If $\sigma > 1$, the function

$$P_\sigma(z) = \prod_{n=1}^{\infty} \left(1+\frac{z}{n^\sigma}\right)$$

is an integral function of order $1/\sigma$. [For further properties of the function see Hardy (4).]

9. The function

$$\prod_{n=1}^{\infty} \left(1+\frac{z}{e^n}\right)$$

is an integral function of zero order.

10. Show that, if $\alpha > 0$,

$$f_\alpha(z) = \prod_{n=1}^{\infty} \left(1+\frac{e^z}{e^{n^\alpha}}\right)$$

is an integral function of order $1+1/\alpha$; and express it in the standard factor form in the cases $\alpha = 1$, $\alpha = 2$.

11. If $a > 0$, the function

$$E_a(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(1+an)}$$

is an integral function of order $1/a$. [Several memoirs on this function are to be found in *Acta Math.* 29.]

12. If a is real, all the roots of the equation $\Gamma'(z) = a\Gamma(z)$ are real.

13. Show that

$$\sum_{n=0}^{\infty} \frac{\cosh \sqrt{n}}{n!} z^n$$

is an integral function of order 1, and that it has an infinity of zeros, all of them being real and negative.

14. A function $f(z)$ of order $\frac{1}{2}$ has all its zeros real and negative, and such that $n(r) \sim k\sqrt{r} \log r$. Determine the asymptotic behaviour of $M(r)$.

[Use the method of § 8.64.]

15. Show that, if $f(z)$ is a canonical product with zeros z_n such that $\sum 1/|z_n|$ is convergent, then $f(z) = O(e^{\epsilon|z|})$, and $|f(z)| > e^{-\epsilon|z|}$ on circles of arbitrarily large radius.

[We have

$$|f(z)| \leqslant \prod_{n=1}^{N} \left(1 + \left|\frac{z}{z_n}\right|\right) \exp\left(\sum_{N+1}^{\infty} \left|\frac{z}{z_n}\right|\right),$$

whence the first result easily follows. The second part then follows from § 8.75.]

16. In Laguerre's theorem of § 8.52, show that $f'(z)$ has the same genus as $f(z)$.

[The only case in which there is anything to prove is the case $\rho = 1$, when the genus may be 0 or 1. Then we use the fact that the series $\sum 1/|z_n|$ and $\sum 1/|z_n'|$ converge or diverge together, compare $M(r)$ and $M'(r)$ by § 8.51, and apply the previous example.]

17. Show that the genus of a function of exponential type (§ 8.75) is 1.

18. Show that, if $f(z) = \sum_0^{\infty} a_n z^n$ is of exponential type, $f^{(n)}(0) = O(e^{An})$, and hence that $\phi(z) = \sum n! a_n z^n$ has a finite radius of convergence.

19. In order that $f(z)$ should be of exponential type, it is necessary and sufficient that it should be expressible in the form

$$f(z) = \frac{1}{2\pi i} \int_C e^{zw} \chi(w)\, dw,$$

where $\chi(w)$ is regular for sufficiently large values of w (including infinity), and C is a circle with centre the origin and sufficiently large radius.

[We have $\chi(w) = 1/w\phi(1/w)$, where ϕ is the function defined in the previous example.]

20. Let $f(z)$ be of exponential type, and let $h(\theta)$, supposed $\geqslant 0$, be the Phragmén-Lindelöf function associated with $f(z)$, with $V(r) = r$. Consider the radii vectores of length $h(\theta)$ making angles $-\theta$ with the real axis, and the perpendiculars to these radii vectores at their ends (cf. § 5.72). Then $\chi(w)$ is regular if w lies on the side of one of these perpendiculars opposite to the origin.

[We have

$$\phi(z) = \int_0^\infty e^{-t} f(zt)\, dt$$

by term-by-term integration, if $|z|$ is small enough; turning the contour through an angle λ,

$$\phi(z) = \int_0^\infty e^{-te^{i\lambda}} f(zte^{i\lambda}) e^{i\lambda}\, dt.$$

Here the integrand is $\quad O(e^{-t\cos\lambda + rt\{h(\theta+\lambda)+\epsilon\}})$,

and the integral is convergent if

$$rh(\theta+\lambda) < \cos\lambda. \tag{1}$$

Hence $\phi(z)$ is regular at $z = re^{i\theta}$ if (1) holds for some value of λ. If $w = r'e^{i\theta'}$, then $\chi(w)$ is regular if $r' > h(\lambda-\theta')\sec\lambda$ for some λ. This is equivalent to the above statement.

For a detailed discussion see Pólya (4).]

21. The function

$$\sum_{n=0}^{\infty} \frac{z^n}{(n+a)^s n!}$$

is of exponential type. [For further properties of the function see Hardy (2).]

22. Show that the function

$$f(z) = \int_a^b e^{izt} g(t)\, dt,$$

where $g(t)$ is continuous, is of exponential type, and that the corresponding function $\chi(w)$ is regular except in the interval (ia, ib) of the imaginary axis.

23. Show that the function $f(z)$ of the above example tends to zero as $z \to \infty$ in either direction along the real axis, and deduce that $f(z)$ has an infinity of zeros.

24. A function $f(z)$ is said to be of zero type if $f(z) = O(e^{\epsilon r})$.

In order that $f(z)$ should be of zero type, it is necessary and sufficient that

$$f(z) = \frac{1}{2\pi i} \int_C e^{zw} \chi(w)\, dw,$$

where $\chi(w)$ is an integral function of $\dfrac{1}{w}$.

[The situation is similar to that of examples 18–19, except that here $f^{(n)}(0) = O(e^{\epsilon n})$.]

25. A necessary and sufficient condition that $f(z) = \sum_{n=0}^{\infty} a_n z^n$ should be an integral function of $1/(1-z)$ is that there should be an integral function $g(z)$ of zero type such that $a_n = g(n)$ for $n = 1, 2, \ldots$.

[Carlson (1), Wigert (1), Hardy (14). If there is such a function $g(z)$, let

$$g(z) = \frac{1}{2\pi i} \int_C e^{zw} \chi(w)\, dw.$$

Then
$$f(z) - a_0 = \sum_{n=1}^{\infty} a_n z^n = \frac{1}{2\pi i} \int_C \frac{ze^w}{1-ze^w} \chi(w)\, dw,$$

if C is a contour enclosing the origin, and on which $\mathbf{R}(w) < \log|1/z|$. This is an integral of the type used in § 4.6, and by deforming it we can show that any branch of $f(z)$ is regular except at $z = 1$ (where the contour is nipped between 0 and $\log 1/z$). Also

$$f(z) - a_0 = \chi\left(\log \frac{1}{z}\right) + \frac{1}{2\pi i} \int_{C'} \frac{ze^w}{1-ze^w} \chi(w)\, dw,$$

where C' is a contour enclosing $w = \log 1/z$. This shows that $f(z)$ is one-valued near $z = 1$, and so an integral function of $1/(1-z)$.

Conversely, if $f(z)$ is of the type prescribed, we have

$$a_n = \frac{1}{2\pi i} \int \frac{f(z)}{z^{n+1}}\, dz,$$

and we can put $z = e^{-w}$, and deform the resulting contour into any simple closed contour which encloses the origin but lies entirely inside the circle $|w| = 2\pi$. Finally $f(e^{-w})$ is regular except for $w = 0$ and $w = \pm 2k\pi i$ $(k = 1, 2, \ldots)$, and so $f(e^{-w}) = g(w) + \psi(w)$, where $g(w)$ is an integral function of $1/w$, and $\psi(w)$ is regular for $|w| < 2\pi$. Hence the result.]

CHAPTER IX

DIRICHLET SERIES

9.1. Introduction. By a Dirichlet series we mean, in this chapter, a series of the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s}, \tag{1}$$

where the coefficients a_n are any given numbers, and s is a complex variable. The more general type of series

$$\sum a_n e^{-\lambda_n s}$$

is also known as a Dirichlet series. The special type is obtained by putting $\lambda_n = \log n$. For the theory of the general type we must refer to Hardy and Riesz's *General Theory of Dirichlet's Series.*

Throughout the chapter we shall write $s = \sigma + it$, where σ and t are real. If the Dirichlet series is convergent, we shall denote its sum by $f(s)$. We have already had one important example of a Dirichlet series, the zeta-function of Riemann,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}. \tag{2}$$

Dirichlet series are not of such importance in general analysis as power series because they only represent a very special class of analytic functions. They are, however, of great importance in applications of analysis to theory of numbers. In several ways their theory is more complicated than that of power series. For example, the circle of convergence, circle of absolute convergence, and circle of regularity of sum-function are all the same for a power series. In the theory of Dirichlet series, in which 'circle' must be replaced by 'half-plane', the three corresponding half-planes may be all different.

9.11. The association of half-planes with a Dirichlet series depends on the following theorem:

If the Dirichlet series is convergent for $s = s_0$, then it is uniformly convergent throughout the angular region in the s-plane defined by the inequality

$$|\arg(s - s_0)| \leqslant \tfrac{1}{2}\pi - \delta,$$

where δ is any positive number less than $\frac{1}{2}\pi$.

It is sufficient to consider the case where $s_0 = 0$; for

$$\sum \frac{a_n}{n^s} = \sum \frac{a'_n}{n^{s'}},$$

where $$a'_n = a_n n^{-s_0}, \qquad s' = s - s_0,$$

and the latter series is convergent for $s' = 0$.

We suppose, then, that $\sum a_n$ is convergent. Let

$$r_n = a_{n+1} + a_{n+2} + \dots$$

so that $r_n \to 0$. Then

$$\sum_{n=M}^{N} \frac{a_n}{n^s} = \sum_{M}^{N} \frac{r_{n-1} - r_n}{n^s} = \sum_{M}^{N} r_n \left\{ \frac{1}{(n+1)^s} - \frac{1}{n^s} \right\} + \frac{r_{M-1}}{M^s} - \frac{r_N}{(N+1)^s}. \quad (1)$$

Now

$$\left| \frac{1}{(n+1)^s} - \frac{1}{n^s} \right| = \left| s \int_n^{n+1} \frac{du}{u^{s+1}} \right| \leqslant |s| \int_n^{n+1} \frac{du}{u^{\sigma+1}} = \frac{|s|}{\sigma} \left\{ \frac{1}{n^\sigma} - \frac{1}{(n+1)^\sigma} \right\}, \quad (2)$$

and $|r_n| < \epsilon$ for $n \geqslant n_0 = n_0(\epsilon)$, n_0 being independent of s. Hence for $M > n_0$

$$\left| \sum_{n=M}^{N} \frac{a_n}{n^s} \right| < \frac{\epsilon |s|}{\sigma} \sum_{n=M}^{N} \left\{ \frac{1}{n^\sigma} - \frac{1}{(n+1)^\sigma} \right\} + \frac{\epsilon}{M^\sigma} + \frac{\epsilon}{(N+1)^\sigma}$$

$$= \frac{\epsilon |s|}{\sigma} \left\{ \frac{1}{M^\sigma} - \frac{1}{(N+1)^\sigma} \right\} + \frac{\epsilon}{M^\sigma} + \frac{\epsilon}{(N+1)^\sigma}$$

$$< 2\epsilon |s|/\sigma + 2\epsilon.$$

If $|\arg s| \leqslant \frac{1}{2}\pi - \delta$, i.e. $t/\sigma \leqslant \tan(\frac{1}{2}\pi - \delta) = \cot \delta$, we have

$$|s|/\sigma = \sqrt{(1 + t^2/\sigma^2)} \leqslant \operatorname{cosec} \delta.$$

Hence $$\sum_{n=M}^{N} \frac{a_n}{n^s} < 2\epsilon(\operatorname{cosec} \delta + 1).$$

The right-hand side is independent of s, and tends to 0 with ϵ; and this establishes uniform convergence.

In particular, *if the series is convergent for $s_0 = \sigma_0 + it_0$, it is convergent for $s = \sigma + it$, provided that $\sigma > \sigma_0$.* For we can choose the δ of the above proof so small that $|\arg(s - s_0)| < \frac{1}{2}\pi - \delta$.

9.12. *The region of convergence of the series is a half-plane.* For we can divide values of σ' into two classes, those for which the series is convergent for $\sigma > \sigma'$, and other values of σ'. By the above theorem, every member of the first class lies to the

right of every member of the second class. Let σ_0 be the real number defined by this section. Then the series is convergent for $\sigma > \sigma_0$, divergent for $\sigma < \sigma_0$.

The number σ_0 is called *the abscissa of convergence* of the series.

The series may converge for all values of s (e.g. $a_n = 1/n!$), or for no values of s (e.g. $a_n = n!$).

The sum $f(s)$ of the series is an analytic function of s, regular for $\sigma > \sigma_0$. For every term of the series is analytic, and any point s with $\sigma > \sigma_0$ is included in a domain of uniform convergence.

The questions of the convergence of the series, and the regularity of the function, *on* the line $\sigma = \sigma_0$, remain open; and (as in the case of power series) various different cases are possible.

We have, however, the following analogue of Abel's theorem for power series:

If the series is convergent for $s = s_0$, and has the sum $f(s_0)$, then $f(s) \to f(s_0)$ when $s \to s_0$ along any path lying entirely inside the region $|\arg(s - s_0)| \leqslant \frac{1}{2}\pi - \delta$.

This follows at once from the theorem of uniform convergence.

9.13. Absolute convergence. *The region of absolute convergence of the Dirichlet series is a half-plane.*

For the series is absolutely convergent if the series

$$\sum_1^\infty \frac{|a_n|}{n^\sigma}$$

is convergent. If this is convergent for a particular value of σ, it is clearly convergent for any greater value. Hence, as in the case of convergence, there is a number $\bar{\sigma}$ such that it is convergent for $\sigma > \bar{\sigma}$, and divergent for $\sigma < \bar{\sigma}$.

Hence the original series is absolutely convergent for $\sigma > \bar{\sigma}$, and not absolutely convergent for $\sigma < \bar{\sigma}$.

The number $\bar{\sigma}$ is called the *abscissa of absolute convergence.*

The numbers σ_0 and $\bar{\sigma}$ are not necessarily equal, i.e. there may be a strip of the plane in which the series is convergent, but not absolutely convergent.

This is shown by the following example. If $\sigma > 1$, we have

$$\left(1 - \frac{1}{2^{s-1}}\right)\zeta(s) = \left(\frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \dots\right) - 2\left(\frac{1}{2^s} + \frac{1}{4^s} + \dots\right)$$
$$= \frac{1}{1^s} - \frac{1}{2^s} + \frac{1}{3^s} - \frac{1}{4^s} + \dots.$$

The last series, as it is arranged here, is convergent for $\sigma > 0$ (and uniformly convergent in any finite region to the right of $\sigma = 0$). For, by a well-known theorem (*P.M.* § 188), it is convergent if s is real and positive. Hence, by the theory of analytic continuation, the formula holds for $\sigma > 0$.

In this case $\sigma_0 = 0$, $\bar{\sigma} = 1$.

In any case $\bar{\sigma} - \sigma_0 \leqslant 1$.

For if $\sum a_n n^{-s}$ is convergent, $|a_n| n^{-\sigma}$ is bounded as $n \to \infty$, and hence

$$\sum \frac{a_n}{n^{s+1+\delta}}$$

is absolutely convergent if $\delta > 0$, which gives the result.

In the above example, $\bar{\sigma} - \sigma_0 = 1$, so that the strip of non-absolute convergence may be as wide as 1, though it can be no wider.

9.14. The abscissa of convergence. The formula for σ_0, analogous to the formula (§ 7.1) for the radius of convergence of a power series, takes slightly different forms according to whether $\sum a_n$ is convergent or not. Let

$$s_n = a_1 + a_2 + \ldots + a_n,$$

and, if $\sum a_n$ is convergent, let $r_n = a_{n+1} + a_{n+2} + \ldots$. Let

$$\alpha = \overline{\lim_{n\to\infty}} \frac{\log|s_n|}{\log n}, \qquad \beta = \overline{\lim_{n\to\infty}} \frac{\log|r_n|}{\log n},$$

β being defined only if $\sum a_n$ is convergent.

Then $\sigma_0 = \alpha$ *if* $\sum a_n$ *is divergent, and otherwise* $\sigma_0 = \beta$.

In the former case $\sigma_0 \geqslant 0$, and in the latter case $\sigma_0 \leqslant 0$; for $\sum a_n$ is convergent if $\sigma_0 < 0$.

(i) Let $\sum a_n$ be divergent, and let s have a real positive value for which the Dirichlet series is convergent. Let

$$b_n = a_n n^{-s}, \qquad B_n = b_1 + b_2 + \ldots + b_n, \qquad B_0 = 0,$$

so that B_n is bounded, say $|B_n| \leqslant B$. Then

$$s_N = \sum_{n=1}^{N} b_n n^s = \sum_{n=1}^{N} (B_n - B_{n-1}) n^s$$

$$= \sum_{n=1}^{N-1} B_n \{n^s - (n+1)^s\} + B_N N^s.$$

Hence $$|s_N| \leqslant B\sum_{n=1}^{N-1}\{(n+1)^s - n^s\} + BN^s < 2BN^s,$$
$$\log|s_N| < s\log N + \log 2B,$$
and so $\alpha \leqslant s$. Hence $\alpha \leqslant \sigma_0$.

A similar argument holds if $\sum a_n$ is convergent. If s has a real negative value for which the Dirichlet series is convergent,

$$r_N = \sum_{n=N+1}^{\infty} b_n n^s = \sum_{n=N+1}^{\infty}(B_n - B_{n-1})n^s$$
$$= \sum_{n=N}^{\infty} B_n\{n^s - (n+1)^s\} - B_N N^s,$$

so that $$|r_N| \leqslant B\sum_{n=N}^{\infty}\{n^s - (n+1)^s\} + BN^s = 2BN^s.$$

Hence, as in the other case, $\beta \leqslant \sigma_0$.

(ii) Since $s_n = O(n^{\alpha+\epsilon})$, and, if s is real,

$$\frac{1}{n^s} - \frac{1}{(n+1)^s} = s\int_n^{n+1}\frac{du}{u^{s+1}} = O(n^{-s-1}),$$

we have

$$\sum_{n=M+1}^{N}\frac{a_n}{n^s} = \sum_{n=M+1}^{N}\frac{s_n - s_{n-1}}{n^s}$$
$$= \sum_{n=M+1}^{N} s_n\left\{\frac{1}{n^s} - \frac{1}{(n+1)^s}\right\} + \frac{s_N}{(N+1)^s} - \frac{s_M}{(M+1)^s} \qquad (1)$$
$$= \sum_{M+1}^{N} O(n^{\alpha+\epsilon-s-1}) + O(N^{\alpha+\epsilon-s}) + O(M^{\alpha+\epsilon-s})$$
$$= O(M^{\alpha+\epsilon-s}) = o(1)$$

if $s > \alpha$ and ϵ is small enough. Hence the Dirichlet series is convergent if $s > \alpha$; hence $\sigma_0 \leqslant \alpha$. Since $a_n = r_{n-1} - r_n$ if $\sum a_n$ is convergent, we find similarly that $\sigma_0 \leqslant \beta$. This proves the theorem.

If $\alpha = \infty$, the series is nowhere convergent, and if $\beta = -\infty$ it is everywhere convergent. This easily follows from the above argument.

9.15. The abscissa of absolute convergence. We have

$$\bar{\sigma} = \overline{\lim_{n\to\infty}}\frac{\log(|a_1| + |a_2| + \ldots + |a_n|)}{\log n},$$

or $$\bar{\sigma} = \overline{\lim_{n\to\infty}}\frac{\log(|a_{n+1}| + |a_{n+2}| + \ldots)}{\log n},$$

according to whether $\sum |a_n|$ is divergent or convergent. This is a particular case of the previous result.

Example. Determine σ_0 and $\bar{\sigma}$ for the series in which $a_n = 1$, $(-1)^n$, $n^{-\frac{1}{2}}$, $(-1)^n n^{-\frac{1}{2}}$, $a^n (0 < a < 1)$, $\log n$, $1/\log n$, respectively; and for the series in which $a_n = 1$ (n a perfect square), $a_n = 0$ otherwise.

9.2. Convergence of the series and regularity of the function. The region of convergence of a power series is determined in a perfectly simple way by the analytic character of the function which it represents—the circle of convergence passes through the singularity nearest to the centre. There is no such simple relation in the case of Dirichlet series. There is not necessarily any singularity on the line of convergence, and in fact $f(s)$ may be an integral function even though the abscissa of convergence of the series is finite. This is shown by the above example of the series for $(1-2^{1-s})\zeta(s)$. This is an integral function, since the pole of $\zeta(s)$ at $s = 1$ is cancelled by the zero of $1-2^{1-s}$. But the corresponding series converges for $\sigma > 0$ only.

On the other hand, the series for $\zeta(s)$, § 9.1 (2), has a singularity on its line of convergence. This is a particular case of the following theorem:

If $a_n \geqslant 0$ for all values of n, then the real point of the line of convergence is a singularity of $f(s)$.

The proof is similar to that of the corresponding theorem for power series (§ 7.21).

We may suppose without loss of generality that $\sigma_0 = 0$. Then if $s = 0$ is a regular point, the Taylor's series for $f(s)$, at the point $s = 1$, has a radius of convergence greater than 1. Hence we can find a negative value of s for which

$$f(s) = \sum_{\nu=0}^{\infty} \frac{(s-1)^\nu}{\nu!} f^{(\nu)}(1) = \sum_{\nu=0}^{\infty} \frac{(1-s)^\nu}{\nu!} \sum_{n=1}^{\infty} \frac{(\log n)^\nu a_n}{n}.$$

But every term in this repeated series is positive. Hence the order of summation may be inverted, and we obtain

$$f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n} \sum_{\nu=0}^{\infty} \frac{(1-s)^\nu (\log n)^\nu}{\nu!}$$

$$= \sum_{n=1}^{\infty} \frac{a_n}{n} e^{(1-s)\log n} = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

Thus the Dirichlet series is convergent for a negative value of s, contrary to hypothesis.

9.3. Asymptotic behaviour of the function as $t \to \infty$. *The function $f(s)$ is bounded in any half-plane included in the half-plane of absolute convergence.*

For
$$|f(s)| \leqslant \sum_{n=1}^{\infty} \frac{|a_n|}{n^\sigma} \leqslant \sum_{n=1}^{\infty} \frac{|a_n|}{n^\alpha}$$
for $\sigma \geqslant \alpha > \bar{\sigma}$, and all values of t.

If the series
$$\sum_{n=1}^{\infty} \frac{|a_n|}{n^{\bar{\sigma}}}$$
is convergent, we can take $\alpha = \bar{\sigma}$, and the function is bounded in the half-plane of absolute convergence. This is true, for example, of the function
$$f(s) = \sum_{n=2}^{\infty} \frac{1}{n^s \log^2 n}.$$
But in general the half-plane of absolute convergence is not a region where $f(s)$ is bounded, even if we exclude the neighbourhood of singularities on the line $\sigma = \bar{\sigma}$ (see § 9.32).

Even in the half-plane of absolute convergence, the behaviour of $f(\sigma+it)$ as $t \to \infty$ is, in general, rather complicated. Take, for example, a series with real positive coefficients in which
$$\sum_{n=3}^{\infty} a_n n^{-\sigma} < a_2 2^{-\sigma}$$
for a certain value of σ. Then
$$\mathbf{R}f(s) = \sum_{n=1}^{\infty} \frac{a_n \cos(t \log n)}{n^\sigma} > a_1 + \frac{a_2}{2^\sigma} - \sum_{n=3}^{\infty} \frac{a_n}{n^\sigma}$$
for $t = 2m\pi/\log 2$, $m = 1, 2, \ldots$. Also
$$\mathbf{R}f(s) < a_1 - \frac{a_2}{2^\sigma} + \sum_{n=3}^{\infty} \frac{a_n}{n^\sigma}$$
for $t = (2m+1)\pi/\log 2$, $m = 1, 2, \ldots$. Hence $\mathbf{R}f(s)$ oscillates as $t \to \infty$.

9.31. The following theorem is due to Dirichlet: *Given N real numbers $c_1, c_2, \ldots, c_N$, a positive integer q, and a positive number τ,*

we can find a number t in the range $\tau \leqslant t \leqslant \tau q^N$, and integers $x_1, \ldots, x_N$, such that

$$|tc_n - x_n| \leqslant \frac{1}{q} \qquad (n = 1, 2, \ldots, N).$$

The proof is based on the argument that if there are $m+1$ points in m regions, there must be at least one region which contains at least two points.

Consider the N-dimensional unit cube with a vertex at the origin and edges along the coordinate axes. Divide each edge into q equal parts, and thus the cube into q^N equal compartments. Consider the q^N+1 points in the cube

$$(\lambda c_1 - [\lambda c_1], \lambda c_2 - [\lambda c_2], \ldots, \lambda c_N - [\lambda c_N]),$$

where λ takes the values $0, \tau, \ldots, \tau q^N$. At least two of these points must lie in the same compartment. If they are given by $\lambda = \lambda_1$, $\lambda = \lambda_2$ $(\lambda_1 < \lambda_2)$, then there are integers $x_1, x_2, \ldots$ such that

$$(\lambda_2 - \lambda_1)c_n - x_n \leqslant 1/q \qquad (n = 1, 2, \ldots, N),$$

and so $t = \lambda_2 - \lambda_1$ gives the required result.

9.32. We can now deduce the following theorem.

If $f(s) = \sum a_n n^{-s}$, where $a_n \geqslant 0$ for every value of n, and where $\sum a_n n^{-\bar{\sigma}}$ is divergent, the function $f(s)$ is not bounded in the region $\sigma > \bar{\sigma}$, $|t| \geqslant t_0 > 0$.

That there may be no singularities on the boundary of this region is shown by the function $\zeta(s) = \sum n^{-s}$, which is regular except at $s = 1$. We make $|t| \geqslant t_0$ to exclude the neighbourhood of the point $s = \bar{\sigma}$, where there is a singularity.

We have, for every value of N, and $\sigma > \bar{\sigma}$,

$$f(s) = \sum_{n=1}^{N} \frac{a_n}{n^\sigma} e^{-it \log n} + \sum_{N+1}^{\infty} \frac{a_n}{n^s},$$

and so

$$|f(s)| \geqslant \mathbf{R} \sum_{n=1}^{N} \frac{a_n}{n^\sigma} e^{-it \log n} - \left| \sum_{N+1}^{\infty} \frac{a_n}{n^s} \right|$$

$$\geqslant \sum_{n=1}^{N} \frac{a_n}{n^\sigma} \cos(t \log n) - \sum_{N+1}^{\infty} \frac{a_n}{n^\sigma}.$$

By Dirichlet's theorem there is, for given N and q, a number t $(\tau \leqslant t \leqslant \tau q^N)$ and integers $x_1, \ldots, x_N$, such that

$$\left| \frac{t \log n}{2\pi} - x_n \right| \leqslant \frac{1}{q} \qquad (n = 1, 2, \ldots, N).$$

Hence $\cos(t\log n) \geqslant \cos(2\pi/q)$ for these values of n, and so

$$\sum_{n=1}^{N} \frac{a_n}{n^\sigma}\cos(t\log n) \geqslant \cos\frac{2\pi}{q}\sum_{n=1}^{N}\frac{a_n}{n^\sigma}$$

$$> \cos\frac{2\pi}{q} f(\sigma) - \sum_{N+1}^{\infty}\frac{a_n}{n^\sigma}.$$

Hence, taking $q=6$, say, so that $\cos 2\pi/q = \frac{1}{2}$,

$$|f(s)| > \tfrac{1}{2}f(\sigma) - 2\sum_{N+1}^{\infty}\frac{a_n}{n^\sigma}.$$

Given any positive number H, we can choose $\sigma - \bar{\sigma}$ so small that $f(\sigma) > 4H$ (since $f(\sigma) \to \infty$ as $\sigma \to \bar{\sigma}$). Having fixed σ, we can choose N so large that

$$\sum_{N+1}^{\infty}\frac{a_n}{n^\sigma} < \tfrac{1}{2}H.$$

Then $|f(s)| > H$, and the result follows.

9.33. In the half-plane of convergence, the function may, for certain values of t, become as large as a power of t. For example, the function $f(s) = (1-2^{1-s})\zeta(s)$, referred to in § 9.13, satisfies the inequality

$$|f(s)| > At^{\frac{1}{2}-\sigma}$$

for some arbitrarily large values of t, and values of σ between 0 and $\frac{1}{2}$.*

On the other hand, the function cannot have values greater than every power of t. This is shown by the following theorem.

We have

$$f(s) = O(|t|^{1-(\sigma-\sigma_0)+\epsilon})$$

as $|t| \to \infty$, *for any value of* σ *between* σ_0 *and* σ_0+1; *and also uniformly in the half-plane to the right of any such line.*

Suppose first that $\sum a_n$ is convergent. Then a_n and s_n are bounded. Now (§ 9.14 (1))

$$\sum_{1}^{N}\frac{a_n}{n^s} = \sum_{1}^{M}\frac{a_n}{n^s} + \sum_{M+1}^{N} s_n\left\{\frac{1}{n^s} - \frac{1}{(n+1)^s}\right\} - \frac{s_M}{(M+1)^s} + \frac{s_N}{(N+1)^s}.$$

If $\sigma > 0$, the last term tends to zero as $N \to \infty$, and we obtain

$$f(s) = \sum_{1}^{M}\frac{a_n}{n^s} + \sum_{M+1}^{\infty} s_n\left\{\frac{1}{n^s} - \frac{1}{(n+1)^s}\right\} - \frac{s_M}{(M+1)^s}.$$

* See Miscellaneous Examples, no. 18.

Hence, by § 9.11 (2), if $0<\sigma<1$,

$$|f(s)|<A\sum_{1}^{M}\frac{1}{n^\sigma}+A\frac{|s|}{\sigma}\sum_{M+1}^{\infty}\left\{\frac{1}{n^\sigma}-\frac{1}{(n+1)^\sigma}\right\}+\frac{A}{(M+1)^\sigma}$$
$$<AM^{1-\sigma}+AtM^{-\sigma}+A.$$

Taking $M=[t]$, we obtain

$$f(s)=O(t^{1-\sigma}) \qquad (0<\sigma<1),$$

and similarly $f(s)=O(t^{1-\alpha})$ $(0<\alpha<1,\ \sigma\geqslant\alpha)$. In the general case, the series $\sum a_n n^{-s}$ is convergent for $s=\sigma_0+\epsilon$, and we obtain the above case by changing the origin to this point. Hence the general result follows.

9.4. Functions of finite order. At this point we adopt a slightly different point of view. We have so far considered $f(s)$ as being defined by the series $\sum a_n n^{-s}$, and we have confined our attention to the half-plane of convergence of the series. It may, however, be possible to continue the function outside this half-plane. The function, so defined, may be regular in a wider half-plane; or it may be regular in a wider half-plane except for a certain finite region. We shall now consider the relations between a function defined in this way, and the Dirichlet series from which it originated.

The theorem of § 9.33 suggests that it will be particularly interesting to consider functions which satisfy the condition

$$f(s)=O(|t|^A)$$

for some positive value of A. A function which satisfies this condition for a particular value of σ is said to be of finite order for that value; if the condition is satisfied uniformly for $\sigma_1\leqslant\sigma\leqslant\sigma_2$, we say that the function is of finite order in this strip. Similarly we can define a function of finite order in a half-plane $\sigma\geqslant\sigma_1$.

We have seen that any function defined by a Dirichlet series is of finite order in a half-plane included in the half-plane of convergence. It may be of finite order outside this half-plane; for $\zeta(s)$, for example, $\sigma_0=1$; but (§ 9.13)

$$\zeta(s)=(1-2^{1-s})^{-1}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^s} \qquad (\sigma>0),$$

and hence, by § 9.33,

$$\zeta(s) = O(|t|^{1-\sigma+\epsilon}) \qquad (0 < \sigma \leqslant 1).$$

9.41. The function $\mu(\sigma)$. The lower bound μ of numbers ξ such that $f(s) = O(|t|^{\xi})$ is called the *order* of $f(s)$ for that particular value of σ. Thus μ is a function of σ.

The main properties of the function $\mu(\sigma)$ follow from the Phragmén-Lindelöf theorem proved in § 5.65. Suppose that $f(s)$ is regular and of finite order for $\sigma_1 \leqslant \sigma \leqslant \sigma_2$, $t \geqslant t_0$, and let $\mu(\sigma_1) = \mu_1$, $\mu(\sigma_2) = \mu_2$. Then, if ϵ is any positive number,

$$f(\sigma_1 + it) = O(t^{\mu_1+\epsilon}), \qquad f(\sigma_2 + it) = O(t^{\mu_2+\epsilon}).$$

Hence, by the theorem referred to,

$$f(s) = O(t^{k(\sigma)}) \qquad (\sigma_1 \leqslant \sigma \leqslant \sigma_2),$$

where $$k(\sigma) = \frac{(\sigma_2-\sigma)(\mu_1+\epsilon)+(\sigma-\sigma_1)(\mu_2+\epsilon)}{\sigma_2-\sigma_1}.$$

Making $\epsilon \to 0$, it follows that

$$\mu(\sigma) \leqslant \frac{(\sigma_2-\sigma)\mu_1+(\sigma-\sigma_1)\mu_2}{\sigma_2-\sigma_1} \qquad (\sigma_1 \leqslant \sigma \leqslant \sigma_2). \tag{1}$$

Hence *the function* $\mu(\sigma)$ *is convex downwards.*

It follows also that $\mu(\sigma)$ *is continuous* (§ 5.31).

Secondly, $\mu(\sigma) = 0$ for sufficiently large values of σ; for since $f(s)$ is bounded for $\sigma > \bar{\sigma}$, $\mu(\sigma) \leqslant 0$ for $\sigma > \bar{\sigma}$; on the other hand, if a_m is the first coefficient in the Dirichlet series which does not vanish, and $\bar{\sigma} < \alpha < \sigma$,

$$|f(s)| \geqslant \frac{|a_m|}{m^\sigma} - \sum_{n=m+1}^{\infty} \frac{|a_n|}{n^\sigma} \geqslant \frac{|a_m|}{m^\sigma} - (m+1)^{\alpha-\sigma} \sum_{n=m+1}^{\infty} \frac{|a_n|}{n^\alpha},$$

which can be made positive by taking σ large enough. Thus $|f(s)|$, considered as a function of t, has a positive lower bound if σ is large enough. Hence $\mu(\sigma) \geqslant 0$, so that in fact $\mu(\sigma) = 0$.

If now $\mu(\sigma)$ were negative for any value σ_1 in the region where $f(s)$ is of finite order, it would follow from (1), with σ_2 so large that $\mu_2 = 0$, that $\mu(\sigma) < 0$ for $\sigma_1 < \sigma < \sigma_2$; and we have shown that this is impossible if σ_2 is large enough. *Hence* $\mu(\sigma)$ *is never negative.*

In particular, $\mu(\sigma) = 0$ for $\sigma > \bar{\sigma}$; for we have already shown that $\mu(\sigma) \leqslant 0$ for $\sigma > \bar{\sigma}$.

Again, take $\sigma_2 > \bar{\sigma}$ in (1), so that $\mu_2 = 0$. Then if $\mu_1 > 0$,

$$\mu(\sigma) \leqslant \frac{\sigma_2 - \sigma}{\sigma_2 - \sigma_1} \mu_1 < \mu_1 \qquad (\sigma > \sigma_1).$$

Hence $\mu(\sigma)$ is a steadily decreasing function of σ.

9.42. Perron's formula. We next require an expression for the sum s_n as an integral. This is a particular case of the following theorem.

If x is not an integer, c is any positive number, and $\sigma > \sigma_0 - c$, then

$$\sum_{n<x} \frac{a_n}{n^s} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f(s+w) \frac{x^w}{w} \, dw. \tag{1}$$

Suppose first that $\sigma > \bar{\sigma} - c$. Then the series for $f(s+w)$ is absolutely and uniformly convergent, and we have

$$\frac{1}{2\pi i} \int_{c-iU}^{c+iT} f(s+w) \frac{x^w}{w} \, dw = \frac{1}{2\pi i} \int_{c-iU}^{c+iT} \sum_1^\infty \frac{a_n}{n^{s+w}} \frac{x^w}{w} \, dw$$

$$= \frac{1}{2\pi i} \sum_1^\infty \frac{a_n}{n^s} \int_{c-iU}^{c+iT} \left(\frac{x}{n}\right)^w \frac{dw}{w}. \tag{2}$$

Now by § 3.126

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \left(\frac{x}{n}\right)^w \frac{dw}{w} = \begin{matrix} 1 & (n < x) \\ 0 & (n > x). \end{matrix}$$

It is therefore sufficient to prove that we can replace U and T in (2) by ∞; that is, we want

$$\lim_{T\to\infty} \sum_1^\infty \frac{a_n}{n^s} \int_{c+iT}^{\infty+iT} \left(\frac{x}{n}\right)^w \frac{dw}{w} = 0,$$

with a similar result for U.

Now for a fixed x,

$$\int_{c+iT}^{c+i\infty} \left(\frac{x}{n}\right)^w \frac{dw}{w} = -\left(\frac{x}{n}\right)^{c+iT} \frac{1}{(\log x/n)(c+iT)} + \frac{1}{\log x/n} \int_{c+iT}^{c+i\infty} \left(\frac{x}{n}\right)^w \frac{dw}{w^2}$$

$$= O\left(\frac{1}{n^c T}\right) + O\left(\frac{1}{n^c} \int_T^\infty \frac{dv}{c^2+v^2}\right) = O\left(\frac{1}{n^c T}\right).$$

Hence
$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} \int_{c+iT}^{c+i\infty} \left(\frac{x}{n}\right)^w \frac{dw}{w} = O\left(\frac{1}{T} \sum_{n=1}^{\infty} \frac{|a_n|}{n^{\sigma+c}}\right),$$
and the result (for $\sigma > \bar{\sigma}-c$) follows.

Suppose next that $\sigma_0 - c < \sigma \leqslant \bar{\sigma}-c$. Let $\alpha > \bar{\sigma}-\sigma$, and consider the integral
$$\int f(s+w) \frac{x^w}{w} dw$$
taken round the rectangle formed by the lines $\mathbf{R}(w) = c$, $\mathbf{R}(w) = \alpha$, $\mathbf{I}(w) = -U$, $\mathbf{I}(w) = T$. By the theorem of § 9.33, the integrand is
$$O(t^{-(\sigma+c-\sigma_0)+\epsilon}),$$
and so the integrals along the horizontal sides tend to zero as $U \to \infty$, $T \to \infty$. The integrand is regular inside the rectangle, and so, by Cauchy's theorem,
$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f(s+w) \frac{x^w}{w} dw = \frac{1}{2\pi i} \int_{\alpha-i\infty}^{\alpha+i\infty} f(s+w) \frac{x^w}{w} dw.$$
Since $\sigma > \bar{\sigma}-\alpha$, the right-hand side is equal to $\sum_{n<x} a_n n^{-s}$, by the first part. This completes the proof.

The particular case $s = 0$ is
$$\sum_{n<x} a_n = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f(w) \frac{x^w}{w} dw \qquad (c > \sigma_0). \tag{3}$$
This is Perron's formula.

9.43. There are several other formulae of the same type as Perron's. One which we shall use later is
$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} e^{-(n\delta)^\lambda} = \frac{1}{2\pi i\lambda} \int_{c-i\infty}^{c+i\infty} \Gamma\left(\frac{w}{\lambda}\right) f(s+w) \delta^{-w} dw,$$
where $\delta > 0$, $\lambda > 0$, and $c > 0$, $c > \bar{\sigma}-\sigma$. To prove this, write the right-hand side as
$$\frac{1}{2\pi i\lambda} \int_{c-i\infty}^{c+i\infty} \Gamma\left(\frac{w}{\lambda}\right) \sum_{n=1}^{\infty} \frac{a_n}{n^{s+w}} \delta^{-w} dw,$$
and observe that we can invert the order of summation and

integration by 'absolute convergence'. We therefore obtain

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s} \frac{1}{2\pi i\lambda} \int_{c-i\infty}^{c+i\infty} \Gamma\left(\frac{w}{\lambda}\right)(n\delta)^{-w}\, dw,$$

and the result follows by the calculus of residues.

9.44. The theorem of § 9.42 enables us to obtain a result of the opposite type to the previous ones—we can pass from the order of the function to the convergence of the series.

The Dirichlet series is convergent in the half-plane where $f(s)$ is regular and $\mu(\sigma) = 0$.

Let s be a point in the interior of this half-plane, and let δ be a positive number so small that $\sigma-\delta$ is still in the same half-plane. Let $c > \bar{\sigma}-\sigma+1$ (so that the simpler case of the theorem of § 9.42 can be used). We deform the contour of § 9.42 (1) into the form

$$c-i\infty, \quad c-iT, \quad -\delta-iT, \quad -\delta+iT, \quad c+iT, \quad c+i\infty,$$

where $T > |t|$. In doing so we pass over a pole at $w = 0$, with residue $f(s)$. Hence

$$\sum_{n<x} \frac{a_n}{n^s} - f(s)$$
$$= \frac{1}{2\pi i}\left\{\int_{c-i\infty}^{c-iT} + \int_{c-iT}^{-\delta-iT} + \int_{-\delta-iT}^{-\delta+iT} + \int_{-\delta+iT}^{c+iT} + \int_{c+iT}^{c+i\infty}\right\} f(s+w)\frac{x^w}{w}\, dw.$$

Since we are in the half-plane where $\mu(\sigma) = 0$, we have $f(s) = O(|t|^\epsilon)$ for every positive ϵ. Hence

$$\int_{-\delta-iT}^{-\delta+iT} f(s+w)\frac{x^w}{w}\, dw = \int_{-T}^{T} O\{(|t|+|v|)^\epsilon\}\frac{x^{-\delta}\, dv}{\sqrt{(\delta^2+v^2)}} = O(x^{-\delta}T^\epsilon),$$

and
$$\int_{-\delta+iT}^{c+iT} f(s+w)\frac{x^w}{w}\, dw = \int_{-\delta}^{c} O(T^\epsilon)\frac{x^c}{T}\, du = O(x^c T^{\epsilon-1}).$$

A similar result holds for the integral over $(c-iT, -\delta-iT)$. Finally, as in § 9.42,

$$\int_{c+iT}^{c+i\infty} f(s+w)\frac{x^w}{w}\, dw = \sum_{n=1}^{\infty}\frac{a_n}{n^s}\int_{c+iT}^{c+i\infty}\left(\frac{x}{n}\right)^w \frac{dw}{w}$$
$$= O\left(\frac{x^c}{T}\sum_{n=1}^{\infty}\frac{|a_n|}{n^{\sigma+c}|\log x/n|}\right).$$

We may suppose without loss of generality that x is half an odd integer. Then

$$|\log x/n| > \log\{(n+\tfrac{1}{2})/n\} > A/n.$$

Hence the above expression is

$$O\left(\frac{x^c}{T}\sum_{n=1}^{\infty}\frac{|a_n|}{n^{\sigma+c-1}}\right) = O\left(\frac{x^c}{T}\right),$$

since $\sigma+c-1 > \bar{\sigma}$.

A similar result holds for the integral over $(c-i\infty, c-iT)$; and adding, we obtain

$$\sum_{n<x} a_n n^{-s} - f(s) = O(x^{-\delta}T^{\epsilon}) + O(x^c T^{\epsilon-1}).$$

Taking $T = x^{2c}$, this is

$$O(x^{-\delta+2c\epsilon}) + O(x^{-c+2c\epsilon}),$$

which tends to zero as $x \to \infty$ if $\epsilon < \delta/2c$ and $\epsilon < \frac{1}{2}$. This proves the theorem.

A more general theorem of the same type is given in Landau's *Handbuch*, § 238, Satz 57.

9.45. Let σ_ϵ be the abscissa limiting the half-plane where $f(s)$ is regular and of the form $O(t^\epsilon)$. Then we have proved that

$$\sigma_0 \leqslant \sigma_\epsilon \leqslant \bar{\sigma}.$$

It is not easy to give an example where these numbers are all different. There is some reason to suppose that, for the function $(1-2^{1-s})\zeta(s)$, we have $\sigma_\epsilon = \frac{1}{2}$, so that the three numbers are 0, $\frac{1}{2}$, and 1 respectively. But this has not been proved.

9.5. The mean-value formula. *If* $\sigma > \bar{\sigma}$,

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|f(s)|^2\,dt = \sum_{n=1}^{\infty}\frac{|a_n|^2}{n^{2\sigma}}.$$

For
$$|f(s)|^2 = \sum_{m=1}^{\infty}\frac{a_m}{m^{\sigma+it}}\sum_{n=1}^{\infty}\frac{\bar{a}_n}{n^{\sigma-it}}$$
$$= \sum_{n=1}^{\infty}\frac{|a_n|^2}{n^{2\sigma}} + \sum\sum_{m\neq n}\frac{a_m\bar{a}_n}{m^\sigma n^\sigma}\left(\frac{n}{m}\right)^{it},$$

the series being absolutely convergent, and uniformly convergent in any finite t-range. Hence we may integrate term by

term, and obtain

$$\frac{1}{2T}\int_{-T}^{T}|f(s)|^2\,dt=\sum_{n=1}^{\infty}\frac{|a_n|^2}{n^{2\sigma}}+\sum\sum_{m\neq n}\frac{a_m\bar{a}_n}{m^\sigma n^\sigma}\frac{2\sin(T\log n/m)}{2T\log n/m}.$$

The factor involving T is bounded for all T, m, and n, so that the double series converges uniformly with respect to T; and each term tends to zero as $T\to\infty$. Hence the sum tends to zero as $T\to\infty$, and the result follows.

9.51. The mean-value half-plane. Let σ_m be the least number such that $f(s)$ is regular and of finite order, and the mean-value formula holds, for every σ greater than σ_m. We shall call the half-plane $\sigma>\sigma_m$ the mean-value half-plane. This expression is justified by the following theorem:*

If $f(s)$ is regular and of finite order for $\sigma\geqslant\alpha$, and

$$\frac{1}{2T}\int_{-T}^{T}|f(\alpha+it)|^2\,dt \tag{1}$$

is bounded as $T\to\infty$, then

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|f(\sigma+it)|^2\,dt=\sum_{n=1}^{\infty}\frac{|a_n|^2}{n^{2\sigma}} \tag{2}$$

for $\sigma>\alpha$, and uniformly in any strip $\alpha<\sigma_1\leqslant\sigma\leqslant\sigma_2$.

Starting with the formula of § 9.43, and moving the contour to $\mathbf{R}(w)=\alpha-\sigma$, where $\sigma>\alpha$, we pass a pole at $w=0$, with residue $\lambda f(s)$; and if $\lambda>\sigma-\alpha$, no other pole is passed. Hence

$$\sum_{n=1}^{\infty}\frac{a_n}{n^s}e^{-(n\delta)^\lambda}-f(s)=\frac{1}{2\pi i\lambda}\int_{\alpha-\sigma-i\infty}^{\alpha-\sigma+i\infty}\Gamma\left(\frac{w}{\lambda}\right)f(s+w)\delta^{-w}\,dw$$

$$=O\left\{\delta^{\sigma-\alpha}\int_{-\infty}^{\infty}e^{-A|v|}|f\{\alpha+i(t+v)\}|\,dv\right\}$$

by the asymptotic formula for the Γ-function (§ 4.42, ex. (i)). Now if $|t|<T$,

$$\int_{2T}^{\infty}e^{-Av}|f\{\alpha+i(t+v)\}|\,dv=O\left(\int_{2T}^{\infty}e^{-Av}v^A\,dv\right)=O(e^{-AT}),$$

* Carlson (1). The theorem is analogous to Parseval's theorem for Fourier series (§ 13.54).

and a similar result holds for the integral over $(-\infty, -2T)$. Also, by Schwarz's inequality,*

$$\left\{\int_{-2T}^{2T} e^{-A|v|}|f\{\alpha+i(t+v)\}|\,dv\right\}^2$$
$$\leqslant \int_{-2T}^{2T} e^{-A|v|}|f\{\alpha+i(t+v)\}|^2\,dv \int_{-2T}^{2T} e^{-A|v|}\,dv$$
$$< A\int_{-2T}^{2T} e^{-A|v|}|f\{\alpha+i(t+v)\}|^2\,dv.$$

Hence

$$\left|\sum \frac{a_n}{n^s}e^{-(n\delta)^\lambda}-f(s)\right|^2$$
$$< A\delta^{2\sigma-2\alpha}\int_{-2T}^{2T} e^{-A|v|}|f\{\alpha+i(t+v)\}|^2\,dv + A\delta^{2\sigma-2\alpha}e^{-AT},$$

and, integrating with respect to t over $(-T, T)$,

$$\int_{-T}^{T}\left|\sum \frac{a_n}{n^s}e^{-(n\delta)^\lambda}-f(s)\right|^2 dt$$
$$< A\delta^{2\sigma-2\alpha}\int_{-2T}^{2T} e^{-A|v|}\,dv \int_{-T}^{T}|f\{\alpha+i(t+v)\}|^2\,dt + O(\delta^{2\sigma-2\alpha}).$$

Now

$$\int_{-T}^{T}|f\{\alpha+i(t+v)\}|^2\,dt = \int_{-T+v}^{T+v}|f(\alpha+it)|^2\,dt = O(T)$$

uniformly for $|v| < 2T$. Hence

$$\frac{1}{2T}\int_{-T}^{T}\left|\sum \frac{a_n}{n^s}e^{-(n\delta)^\lambda}-f(s)\right|^2 dt = O(\delta^{2\sigma-2\alpha})$$

uniformly with respect to T. Hence†

$$\left\{\frac{1}{2T}\int_{-T}^{T}\left|\sum \frac{a_n}{n^s}e^{-(n\delta)^\lambda}\right|^2 dt\right\}^{\frac{1}{2}} - \left\{\frac{1}{2T}\int_{-T}^{T}|f(s)|^2\,dt\right\}^{\frac{1}{2}} = O(\delta^{\sigma-\alpha}) \quad (3)$$

uniformly with respect to T.

* See *P.M.* Ch. VII, ex. 42; or § 12.41 below.

† By Minkowski's inequality (§ 12.43), but only in the case where $p = 2$ and the functions are continuous.

If $\delta > 0$, the series $\sum a_n n^{-s} e^{-(n\delta)^\lambda}$ is absolutely convergent, and so, by § 9.5,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left| \sum \frac{a_n}{n^s} e^{-(n\delta)^\lambda} \right|^2 dt = \sum \frac{|a_n|^2}{n^{2\sigma}} e^{-2(n\delta)^\lambda} \tag{4}$$

Taking, say, $\delta = 1$, it follows from (3) and (4) that

$$\frac{1}{2T} \int_{-T}^{T} |f(s)|^2 \, dt < A.$$

Hence by (3), with any positive δ,

$$\sum \frac{|a_n|^2}{n^{2\sigma}} e^{-2(n\delta)^\lambda} < A,$$

and so, since δ may be as small as we please,

$$\sum \frac{|a_n|^2}{n^{2\sigma}}$$

is convergent, and

$$\lim_{\delta\to 0} \sum \frac{|a_n|^2}{n^{2\sigma}} e^{-2(n\delta)^\lambda} = \sum \frac{|a_n|^2}{n^{2\sigma}}.$$

Given ϵ, we can choose δ so that the absolute value of the left-hand side of (3) is less than ϵ, for all values of T, and so that

$$\left| \left\{ \sum \frac{|a_n|^2}{n^{2\sigma}} e^{-2(n\delta)^\lambda} \right\}^{\frac{1}{2}} - \left\{ \sum \frac{|a_n|^2}{n^{2\sigma}} \right\}^{\frac{1}{2}} \right| < \epsilon.$$

Having fixed δ, we can, by (4), choose T_0 so large that

$$\left| \left\{ \frac{1}{2T} \int_{-T}^{T} \left| \sum \frac{a_n}{n^s} e^{-(n\delta)^\lambda} \right|^2 dt \right\}^{\frac{1}{2}} - \left\{ \sum \frac{|a_n|^2}{n^{2\sigma}} e^{-2(n\delta)^\lambda} \right\}^{\frac{1}{2}} \right| < \epsilon.$$

for $T > T_0$. Then

$$\left| \left\{ \frac{1}{2T} \int_{-T}^{T} |f(s)|^2 \, dt \right\}^{\frac{1}{2}} - \left\{ \sum \frac{|a_n|^2}{n^{2\sigma}} \right\}^{\frac{1}{2}} \right| < 3\epsilon$$

for $T > T_0$, and the theorem is proved.

9.52. *If $|f(s)|^2$ has a mean-value for $\sigma = \alpha$, then the Dirichlet series is absolutely convergent for $\sigma > \alpha + \frac{1}{2}$. In symbols,*

$$\bar{\sigma} \leqslant \sigma_m + \tfrac{1}{2}.$$

It follows from the above theorem that

$$\sum \frac{|a_n|^2}{n^{2\alpha+2\epsilon}}$$

is convergent for every positive ϵ. Now

$$\left\{\sum_{n=1}^{N}\left|\frac{a_n}{n^s}\right|\right\}^2 \leqslant \sum_{n=1}^{N}\frac{|a_n|^2}{n^{2\alpha+2\epsilon}}\sum_{n=1}^{N}\frac{1}{n^{2\sigma-2\alpha-2\epsilon}}$$

and this is bounded if ϵ is small enough and $\sigma-\alpha>\frac{1}{2}$. This result was obtained in another way by Hardy (10).

9.53. *If $f(s)$ is bounded for $\sigma>\alpha$, then $\sum |a_n|^2 n^{-2\alpha}$ is convergent; if $|f(s)|\leqslant M$, then*

$$\sum \frac{|a_n|^2}{n^{2\alpha}} \leqslant M^2.$$

This also follows from the theorem of § 9.51. For

$$\sum \frac{|a_n|^2}{n^{2\sigma}} = \lim \frac{1}{2T}\int_{-T}^{T} |f(s)|^2\,dt \leqslant M^2$$

for every $\sigma>\alpha$; and making $\sigma\to\alpha$ the result follows.

If we assume that $f(s)$ is bounded, the analysis of § 9.51 can, of course, be very much simplified.

9.54. Another consequence of these theorems is that a strip in which $f(s)$ is bounded, but in which the Dirichlet series is not absolutely convergent, can be at most of breadth $\frac{1}{2}$.

9.55. *The Dirichlet series converges in the half-plane in which $f(s)$ is regular and of finite order, and*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|f(s)|^2\,dt$$

exists. That is, $\sigma_0\leqslant\sigma_m$.

We have first to deduce an 'order' result for $f(s)$ from the given mean-value result.

We have $f(s)=O(|t|^{\frac{1}{2}})$ uniformly in any strip $\alpha\leqslant\sigma\leqslant\beta$, where $\alpha>\sigma_m$.

Let s be a point of the strip (α,β), R a number less than 1 and less than $\alpha-\sigma_m$, independent of t. Then, if $0<\rho<R$,

$$f(s)=\frac{1}{2\pi i}\int_{|z-s|=\rho}\frac{f(z)}{z-s}\,dz=\frac{1}{2\pi}\int_0^{2\pi} f(s+\rho e^{i\phi})\,d\phi.$$

Hence, by Schwarz's inequality,

$$|f(s)|^2 \leqslant \frac{1}{4\pi^2}\int_0^{2\pi} d\phi \int_0^{2\pi} |f(s+\rho e^{i\phi})|^2\, d\phi = \frac{1}{2\pi}\int_0^{2\pi} |f(s+\rho e^{i\phi})|^2\, d\phi.$$

Multiplying by ρ, and integrating with respect to ρ from 0 to R, we have

$$\tfrac{1}{2}R^2|f(s)|^2 \leqslant \frac{1}{2\pi}\int_0^R\int_0^{2\pi} |f(s+\rho e^{i\phi})|^2 \rho\, d\rho d\phi$$

$$< \frac{1}{2\pi}\int_{\sigma-R}^{\sigma+R}\int_{t-R}^{t+R} |f(x+iy)|^2\, dxdy < \frac{1}{2\pi}\int_{\sigma-R}^{\sigma+R} dx \int_{-|t|-1}^{|t|+1} |f(x+iy)|^2\, dy.$$

Now
$$\int_{-|t|-1}^{|t|+1} |f(x+iy)|^2\, dy = O(t)$$

uniformly in x; hence

$$\tfrac{1}{2}R^2|f(s)|^2 = O(t),$$

which gives the required result.

We use the same contour integral as in § 9.44, s and $\sigma-\delta$ now being in the half-plane $\sigma > \sigma_m$. Then

$$\left|\int_{-\delta-iT}^{-\delta+iT} f(s+w)\frac{x^w}{w}\, dw\right| \leqslant x^{-\delta}\left\{\int_{-T}^{T}\frac{|f(s+w)|^2}{\sqrt{(\delta^2+v^2)}}\, dv \int_{-T}^{T}\frac{dv}{\sqrt{(\delta^2+v^2)}}\right\}^{\frac{1}{2}}.$$

Let
$$\phi(v) = \int_0^v |f(\sigma+u+iy)|^2\, dy = O(v).$$

Then

$$\int_0^T \frac{|f|^2}{\sqrt{(\delta^2+v^2)}}\, dv = \frac{\phi(T)}{\sqrt{(\delta^2+T^2)}} + \int_0^T \frac{v\,\phi(v)}{(\delta^2+v^2)^{\frac{3}{2}}}\, dv$$

$$= O(1) + \int_{-T}^{T}\frac{O(v^2)}{(\delta^2+v^2)^{\frac{3}{2}}}\, dv = O(\log T).$$

Similarly the integral over $(-T, 0)$ is $O(\log T)$. Hence

$$\int_{-\delta-iT}^{-\delta+iT} f(s+w)\frac{x^w}{w}\, dw = O(x^{-\delta}\log T).$$

Also, by the lemma,

$$\int_{-\delta+iT}^{c+iT} f(s+w)\frac{x^w}{w}\,dw = \int_{-\delta}^{c} O(\sqrt{T})\frac{x^c}{T}\,du = O(x^c T^{-\frac{1}{2}}),$$

with a similar result for the integral over $(c-iT, -\delta-iT)$. As in § 9.44, the remaining integrals are $O(x^c T^{-1})$. Hence

$$\sum_{n<x}\frac{a_n}{n^s}-f(s) = O(x^{-\delta}\log T)+O(x^c T^{-\frac{1}{2}}),$$

and, taking $T = x^{2c+1}$, this tends to 0 as $x \to \infty$, so that the result follows.

9.6. The uniqueness theorem. A function $f(s)$ can have at most one representation as a Dirichlet series. More precisely, *if*

$$\sum_1^\infty \frac{a_n}{n^s} = \sum_1^\infty \frac{b_n}{n^s}$$

in any region of values of s, then $a_n = b_n$ for all values of n.

For the series $\sum (a_n-b_n)n^{-s}$ is uniformly convergent in a region including part of the given region and extending arbitrarily far to the right; and so its sum is the same analytic function, viz. 0, in the whole region. But, if m is the first value of n for which $a_n \neq b_n$,

$$|\sum (a_n-b_n)n^{-s}| \geqslant |a_m-b_m|m^{-\sigma}-\sum_{m+1}^\infty |a_n-b_n|n^{-\sigma},$$

and, as in § 9.41, this is positive if σ is large enough. This leads to a contradiction, and so proves the theorem.

9.61. The zeros of $f(s)$. The above argument shows that *$f(s)$ always has a half-plane free from zeros.*

The problem of the distribution of the zeros of any given $f(s)$ is usually a very difficult one, and the results for different functions may be very different. For example, it is supposed that all the complex zeros of the function

$$\frac{1}{1^s}-\frac{1}{2^s}+\frac{1}{3^s}-\ldots = (1-2^{1-s})\zeta(s)$$

lie on the lines $\sigma = 1$ (zeros of $1-2^{1-s}$) and $\sigma = \frac{1}{2}$ (zeros of $\zeta(s)$); the zeros on $\sigma = 1$ are easily identified, but the remaining statement has never been proved.

On the other hand, it is known that the function $\zeta'(s)/\zeta(s)$,

which is represented by an absolutely convergent Dirichlet series for $\sigma > 1$, has no zeros in a certain half-plane $\sigma > E$ $(E > 1)$, and zeros on lines $\sigma = \sigma'$ which are dense everywhere in the interval $1 < \sigma < E$.

It is interesting to compare the general problem with the particular case where $a_n = 0$ except when n is a power of 2. Then the function is of the form

$$f(s) = \sum_{n=0}^{\infty} \frac{b_n}{2^{ns}} = \sum_{n=0}^{\infty} b_n z^n,$$

where $z = 2^{-s}$. The series is a power series as well as a Dirichlet series. To each zero z_ν of the power series corresponds a sequence of zeros

$$s_{\mu,\nu} = -\frac{\log z_\nu + 2\mu\pi i}{\log 2} \qquad (\mu = 0, \pm 1, \pm 2, \ldots)$$

of $f(s)$. If z_0 is the zero of smallest modulus (other than 0), $f(s)$ has no zero to the right of the line

$$\sigma = \frac{\log 1/|z_0|}{\log 2},$$

there being an infinity of zeros on this line.

9.62. The function $N(\sigma, T)$. Let t_0 be a positive number such that $f(s)$ is regular for $t \geqslant t_0$ and σ sufficiently large, and let $N(\sigma, T)$ be the number of zeros $\sigma' + it'$ of $f(s)$ such that $\sigma' > \sigma$, $t_0 < t' < T$. Then we have the following theorems:

9.621. *If $f(s)$ is of finite order for $\sigma \geqslant \alpha$, then*

$$N(\sigma, T) = O(T \log T) \qquad (\sigma > \alpha).$$

We can find a number β so large that $|f(s)|$ has positive lower and upper bounds on the line $\sigma = \beta$. Let $0 < \delta < \frac{1}{2}(\beta - \alpha)$. We apply Jensen's theorem to the circle with centre $\beta + in\delta$ and radius $\beta - \alpha$. If $n(r)$ is the number of zeros of $f(s)$ in the circle $|s - (\beta + in\delta)| \leqslant r$, Jensen's theorem gives

$$\int_0^{\beta-\alpha} \frac{n(r)}{r}\, dr = \frac{1}{2\pi} \int_0^{2\pi} \log|f\{\beta + in\delta + (\beta - \alpha)e^{i\theta}\}|\, d\theta - \log|f(\beta + in\delta)|.$$

Now $f(s) = O(t^A)$ for $\sigma \geqslant \alpha$, and so

$$\log|f\{\beta + in\delta + (\beta - \alpha)e^{i\theta}\}| = \log|O\{(n\delta + \beta - \alpha)^A\}| < K \log n,$$

where K depends on α, β, δ,... only. Also $\log|f(\beta+in\delta)| = O(1)$; hence

$$\int_0^{\beta-\alpha} \frac{n(r)}{r}\,dr < K\log n.$$

But

$$\int_0^{\beta-\alpha} \frac{n(r)}{r}\,dr \geqslant n(\beta-\alpha-\delta)\int_{\beta-\alpha-\delta}^{\beta-\alpha} \frac{dr}{r} > Kn(\beta-\alpha-\delta),$$

and $n(\beta-\alpha-\delta)$ is, if δ is small enough, greater than the number of zeros in the strip

$$\sigma \geqslant \alpha+2\delta, \qquad (n-\tfrac{1}{2})\delta \leqslant t < (n+\tfrac{1}{2})\delta.$$

Denoting this number by ν_n, we have therefore

$$\nu_n < K\log n.$$

Hence $$N(\alpha+2\delta, T) \leqslant \sum_{t_0/\delta<n<T/\delta} \nu_n < KT\log T,$$

and the theorem follows.

9.622. *If $f(s)$ is bounded for $\sigma \geqslant \alpha$,*

$$N(\sigma, T) = O(T) \qquad (\sigma > \alpha).$$

The proof is similar to the previous one, but here the factor $\log T$ obviously does not occur. The example at the end of § 9.61 shows that we may have $N(\sigma, T) > AT$.

9.623. *If $f(s)$ has a mean value for $\sigma = \alpha$, and is of finite order for $\sigma \geqslant \alpha$, then* $$N(\sigma, T) = O(T) \qquad (\sigma > \alpha).$$

We use the following lemma:

If $\phi(t)$ is a positive continuous function in (a, b),

$$\frac{1}{b-a}\int_a^b \log\phi(t)\,dt \leqslant \log\left\{\frac{1}{b-a}\int_a^b \phi(t)\,dt\right\}.$$

Divide the interval (a, b) into n equal parts by the points $a = x_0, x_1, \ldots, x_n = b$. We have

$$\{\phi(x_1)\phi(x_2)\ldots\phi(x_n)\}^{1/n} \leqslant \{\phi(x_1)+\phi(x_2)+\ldots+\phi(x_n)\}/n.$$

Hence $$\frac{1}{n}\sum \log\phi(x_\nu) \leqslant \log\left\{\frac{1}{n}\sum \phi(x_\nu)\right\},$$

i.e.

$$\frac{1}{b-a}\sum (x_\nu - x_{\nu-1})\log\phi(x_\nu) \leqslant \log\left\{\frac{1}{b-a}\sum (x_\nu - x_{\nu-1})\phi(x_\nu)\right\}.$$

Making $n \to \infty$, the result follows.

The theorem may be deduced from Jensen's theorem by an elaboration of the argument of § 9.621, but it is more convenient to use the theorem of § 3.8. Applying it to the function $f(s)$ and the rectangle $(\alpha, \beta; t_0, T)$, we have, on taking real parts,

$$2\pi\int_\alpha^\beta N(\sigma, T)\,d\sigma = \int_{t_0}^T \log|f(\alpha+it)|\,dt -$$
$$-\int_{t_0}^T \log|f(\beta+it)|\,dt + \int_\alpha^\beta \arg f(\sigma+iT)\,d\sigma - \int_\alpha^\beta \arg f(\sigma+it_0)\,d\sigma. \tag{1}$$

Applying the lemma to the first term on the right of (1), we have

$$\frac{1}{T-t_0}\int_{t_0}^T \log|f(\alpha+it)|\,dt = \frac{1}{2}\frac{1}{T-t_0}\int_{t_0}^T \log|f(\alpha+it)|^2\,dt$$
$$\leqslant \tfrac{1}{2}\log\left\{\frac{1}{T-t_0}\int_{T_0}^T |f(\alpha+it)|^2\,dt\right\} < A$$

by hypothesis. Thus the term in question is less than AT.

Secondly, as in § 9.621, $\log|f(\beta+it)|$ is bounded if β is large enough. Hence, if β is suitably chosen, the second term on the right of (1) is $O(T)$.

To deal with the third term, suppose first that $f(s)$ is real for real s. We can take β so large that $\mathbf{R}\{f(s)\}$ does not vanish on $\sigma = \beta$. Then, as in § 3.56, $\arg f(s)$ is bounded on $\sigma = \beta$, and, on $t = T$, $\arg f(s) = O(q)$, where q is the number of times $\mathbf{R}\{f(s)\}$ vanishes on $t = T$, $\alpha \leqslant \sigma < \beta$. Now on $t = T$

$$\mathbf{R}\{f(s)\} = \tfrac{1}{2}\{f(\sigma+iT)+f(\sigma-iT)\} = g(\sigma),$$

say, and q is the number of zeros of $g(z)$ on the real z-axis such that $\alpha \leqslant z \leqslant \beta$. Since $g(z) = O(T^A)$, it follows from Jensen's theorem as in § 9.621 that $q = O(\log T)$. Hence the third term on the right of (1) is $O(\log T)$.

If $f(s)$ is not real on the real axis, we can consider instead the function

$$f_1(s) = \sum \frac{a_n}{n^s} \sum \frac{\bar{a}_n}{n^s} = f(s)\bar{f}(s)$$

and apply the same proof to this.

Finally, the last term on the right of (1) is a constant. Hence

$$\int_\alpha^\beta N(\sigma, T)\, d\sigma = O(T).$$

But $$\int_\alpha^\beta N(\sigma, T)\, d\sigma \geqslant \int_\alpha^{\alpha+\delta} N(\sigma, T)\, d\sigma \geqslant \delta N(\alpha+\delta, T),$$

and the result now follows.

9.7. Representation of functions by Dirichlet series. What sort of function can be represented by a Dirichlet series?

It would take us much too far to give anything like an adequate answer to this question, but we can give some indications. It is not difficult to see that a Dirichlet series can only represent functions of a very special kind.

If $f(s)$ is representable by a Dirichlet series, it must, in the first place, be regular and bounded in a certain half-plane (viz. $\sigma \geqslant \bar{\sigma}+\epsilon$). Further, it must have a mean-value

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f(\sigma+it)|^2\, dt$$

for all sufficiently large values of σ, and the value of the limit must decrease steadily as σ increases.

Again, if $f(s) = \sum a_n n^{-s}$, and x is real, then for $\sigma > \bar{\sigma}$

$$\frac{1}{2T}\int_{-T}^{T} f(s)x^s\, dt = \frac{1}{2T}\int_{-T}^{T}\left(\sum_{n=1}^{\infty}\frac{a_n}{n^s}\right)x^s\, dt = \frac{x^\sigma}{2T}\sum_{n=1}^{\infty}\frac{a_n}{n^\sigma}\int_{-T}^{T}\left(\frac{x}{n}\right)^{it} dt$$

$$= a_x + x^\sigma \sum_{n\neq x} \frac{a_n}{n^\sigma}\frac{2\sin(T\log x/n)}{2T\log x/n},$$

the term a_x occurring if x is a positive integer only. The last series, being uniformly convergent in T, tends to zero as $T\to\infty$. Hence

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} f(s)x^s\, dt = \begin{matrix} a_x & (x \text{ a positive integer}), \\ 0 & (\text{otherwise}). \end{matrix}$$

This, therefore, is a necessary condition for $f(s)$ to have the form $\sum a_n n^{-s}$ (the formulae are due to Hadamard). It is, however, not sufficient. But it shows what special properties a function representable by a Dirichlet series must have.

If the Dirichlet series reduces to a single term, say $f(s) = ak^{-s}$, then $f(s)$ is periodic, with period $2\pi i/\log k$. The general Dirichlet series with period $2\pi i/\log k$ is $\sum_{n=0}^{\infty} b_n k^{-ns}$. If we insert other terms, the property of periodicity disappears; but $f(s)$ always retains a certain more general property, which resembles that of periodicity, and any such function is said to be 'almost periodic'. It is in the study of almost periodic functions that answers to the question which we have raised are to be found. We have no space to go into this question further. But we may say roughly that, if an almost periodic function takes a certain value, it repeats this value, not *exactly*, but *approximately*, an infinity of times; and the points where it does this are distributed in much the same way as the periods $(a, 2a, 3a, \ldots)$ of a periodic function.

The theory of almost periodic functions is due to H. Bohr (1), (2), (3).

MISCELLANEOUS EXAMPLES

1. Prove that, if $\phi(x) = \sum_{n=1}^{\infty} a_n e^{-nx}$, then

$$f(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} x^{s-1}\phi(x)\, dx$$

(i) for $\sigma > 0$, $\sigma > \bar{\sigma}$, (ii) for $\sigma > 0$, $\sigma > \sigma_0$.

2. If $0 < \theta < 2\pi$, the function $f(s)$, defined for $\sigma > 0$ by the equation

$$f(s) = \sum_{n=1}^{\infty} \frac{e^{in\theta}}{n^s},$$

is an integral function.

[Use ex. 1 and proceed as in the case of $\zeta(s)$.]

3. The functions defined for $\sigma > 1$ by the series

$$\sum_{n=1}^{\infty} \frac{e^{ain^b}}{n^s} \quad (a > 0,\ 0 < b < 1), \qquad \sum_{n=1}^{\infty} \frac{e^{ai(\log n)^2}}{n^s} \quad (a > 0),$$

are both integral functions. [Hardy (7), (10).]

4. A function represented by a Dirichlet series cannot tend to a limit (in the half-plane of absolute convergence) as $t \to \infty$, unless it is a constant.

[If $f(s) = \sum_1^\infty a_n n^{-s}$, then for $\sigma > \bar{\sigma}$

$$\lim \frac{1}{2T} \int_{-T}^{T} f(s)\, dt = a_1, \qquad \lim \frac{1}{2T} \int_{-T}^{T} |f(s)|^2\, dt = \sum a_n^2.$$

Hence, if $f(s) \to a$, $\quad a_1 = a, \quad \sum |a_n|^2 = |a|^2.$

Hence $\quad |a_2|^2 + |a_3|^2 + \ldots = 0,$

i.e. $a_2 = 0$, $a_3 = 0, \ldots$. Hence $f(s) = a_1$.]

5. Show that

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} \qquad (\sigma > 1),$$

where $\mu(1) = 1$, $\mu(n) = (-1)^r$ if n is the product of r different primes, and otherwise $\mu(n) = 0$. Show also that

$$\frac{\zeta(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{|\mu(n)|}{n^s}.$$

[The infinite product for $\zeta(s)$ is given in § 1.44, ex. 1.]

6. Verify the formulae*

$$\{\zeta(s)\}^2 = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}, \qquad \frac{\{\zeta(s)\}^3}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{d(n^2)}{n^s}, \qquad \frac{\{\zeta(s)\}^4}{\zeta(2s)} = \sum_{n=1}^{\infty} \frac{\{d(n)\}^2}{n^s},$$

where $d(n)$ denotes the number of divisors of n, and $\sigma > 1$.

[If the expression of n in prime factors is

$$n = p_1^{m_1} p_2^{m_2} \ldots p_r^{m_r},$$

then $$d(n) = (m_1+1)(m_2+1)\ldots(m_r+1).$$

Hence $$\sum \frac{d(n^2)}{n^s} = \prod_p \sum_{m=0}^{\infty} \frac{(2m+1)}{p^{ms}}$$

and $$\sum \frac{\{d(n)\}^2}{n^s} = \prod_p \sum_{m=0}^{\infty} \frac{(m+1)^2}{p^{ms}}.]$$

7. Verify the formulae

$$\zeta(s)\zeta(s-a) = \sum_{n=1}^{\infty} \frac{\sigma_a(n)}{n^s} \qquad (\sigma > 1, \sigma > a+1),$$

* A number of other formulae of this kind are given by Pólya and Szegö, *Aufgaben*, VIII. Abschn., nos. 49–64.

and*

$$\frac{\zeta(s)\zeta(s-a)\zeta(s-b)\zeta(s-a-b)}{\zeta(2s-a-b)} = \sum_{n=1}^{\infty} \frac{\sigma_a(n)\sigma_b(n)}{n^s}$$

$$(\sigma > 1, \sigma > a+1, \sigma > b+1, \sigma > a+b+1),$$

where $\sigma_a(n)$ denotes the sum of the ath powers of the divisors of n.

[The second formula follows from the identity

$$\frac{1-p^{a+b-2s}}{(1-p^{-s})(1-p^{a-s})(1-p^{b-s})(1-p^{a+b-s})}$$

$$= \frac{1}{(1-p^a)(1-p^b)} \sum_{m=0}^{\infty} \frac{(1-p^{(m+1)a})(1-p^{(m+1)b})}{p^{ms}}.]$$

8. Let $d_k(n)$, where $k = 2, 3, \ldots$, denote the number of ways of expressing n as a product of k factors, the order of the factors being taken into account. Then

$$\sum_{n=1}^{\infty} \frac{d_k(n)}{n^s} = \{\zeta(s)\}^k \qquad (\sigma > 1);$$

and

$$\sum_{n=1}^{\infty} \frac{\{d_k(n)\}^2}{n^s} = \{\zeta(s)\}^k \prod_p \left\{P_{k-1}\left(\frac{1+p^{-s}}{1-p^{-s}}\right)\right\} \qquad (\sigma > 1),$$

where p runs through all prime numbers, and $P_n(z)$ is the Legendre polynomial of degree n. [Titchmarsh (8).]

9. Show that, if $f(s)$ has the period $2\pi i/\log k$, Hadamard's formulae for the coefficients a_n (§ 9.7) are equivalent to Laurent's formulae for the coefficients in a power series.

10. A necessary and sufficient condition that a function $f(s)$ should be of the form

$$\sum_{n=0}^{\infty} \frac{b_n}{k^{ns}}$$

is that $f(s)$ should be regular and bounded for sufficiently large values of σ. and have the period $2\pi i/\log k$.

11. If $a_n = 0$ unless n is a power of k, then

$$\sigma_0 = \bar{\sigma} = \sigma_\epsilon = \sigma_m.$$

12. The function $f(s) = \sum_{m=0}^{\infty} 2^{-m!s}$ has the line $\sigma = 0$ as a natural boundary. [See § 4.71.]

13. The function $f(s) = \sum p^{-s}$, where p runs through all prime numbers, has the line $\sigma = 0$ as a natural boundary.

[This is a more recondite example than the previous one; see Landau and Walfisz (1).]

* Ramanujan (1). B. M. Wilson (1).

14. The function

$$f(s) = \sum_{n=1}^{\infty} \frac{\{d_k(n)\}^r}{n^s}$$

is meromorphic if $r = 1$, or if $r = 2$, $k = 2$; for other values of r and k it has the line $\sigma = 0$ as a natural boundary. [Estermann (1).]

15. Show that, for the function $\zeta(s)$,

$$\mu(\sigma) = 0 \ (\sigma \geqslant 1), \qquad = 1-2\sigma \ (\sigma \leqslant 0),$$

and that $\mu(\sigma) \leqslant 1-\sigma$ for $0 < \sigma < 1$.

[The result for $\sigma < 0$ follows from the functional equation for $\zeta(s)$. The actual value of $\mu(\sigma)$ for $0 < \sigma < 1$ is not known.]

16. Calculate the mean value

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f(s)|^2 \, dt \qquad (\sigma > 1)$$

for the functions $f(s) = \zeta(s)$, $1/\zeta(s)$, $\{\zeta(s)\}^2$.

17. Show that, if $f(s)$ is unbounded on any line $\sigma = \alpha$ in the half-plane where it is of finite order, it is also unbounded on every line $\sigma = \beta < \alpha$ in the same half-plane.

18. Show that the function $f(s) = (1-2^{1-s})\zeta(s)$ is unbounded on every line $\sigma = \alpha \leqslant 1$; and that $t^{\sigma-\frac{1}{2}}f(s)$ is unbounded on every line $\sigma = \alpha$, where $0 < \alpha < \frac{1}{2}$.

[The theorem of § 9.32 shows that $\zeta(s)$, and so also $(1-2^{1-s})\zeta(s)$, is unbounded for $\sigma > 1$, $|t| > 1$. The theorem of § 9.41 then gives the first result, and the second result then follows from the functional equation for $\zeta(s)$, § 4.44, and the asymptotic formula for the Γ-function, § 4.42.]

19. If $s_n = \sum_{\nu=1}^{n} a_\nu$ is bounded, then $f(s) = \sum a_n n^{-s}$ is regular for $\sigma > 0$; and, if $f(s)$ has a pole on $\sigma = 0$, it is at most of the first order.

$\Big[$If $\phi(u) = \sum_{\nu \leqslant u} a_\nu$, we have

$$f(s) = s \int_1^{\infty} \frac{\phi(u)}{u^{s+1}} du = O\left(s \int_1^{\infty} \frac{du}{u^{\sigma+1}}\right) = O\left(\frac{s}{\sigma}\right).\Big]$$

20. If $s_n \sim n$, then $f(s) \sim 1/(s-1)$ as $s \to 1$ by real values greater than 1.

If $s_n \sim n \log^k n$, where k is a positive integer, then

$$f(s) \sim \frac{k!}{(s-1)^{k+1}}.$$

CHAPTER X

THE THEORY OF MEASURE AND THE LEBESGUE INTEGRAL

10.1. Riemann integration. In the theory of analytic functions we have used the familiar definition of an integral due to Riemann. In the theory of functions of a real variable, however, Riemann's definition has been almost entirely superseded by a more general one, due to Lebesgue.

Lebesgue's definition enables us to integrate functions for which Riemann's method fails; but this is only one of its advantages. The new theory gives us a command over the whole subject which was previously lacking. It deals, so to speak, automatically with many of the limiting processes which present difficulties in the Riemann theory. At this early stage it is difficult to say anything more precise.

Let us begin by recalling the definition of the Riemann integral of a bounded function. Suppose that $f(x)$ is bounded in the interval (a, b); we subdivide this interval by means of the points $x_0, x_1, \ldots, x_n$, so that

$$a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b.$$

Let m_ν, M_ν be the lower and upper bounds of $f(x)$ in the interval $x_\nu < x \leqslant x_{\nu+1}$, and let

$$s = \sum_{\nu=0}^{n-1} m_\nu(x_{\nu+1} - x_\nu), \qquad S = \sum_{\nu=0}^{n-1} M_\nu(x_{\nu+1} - x_\nu).$$

When the number of division-points is increased indefinitely so that the greatest interval $x_{\nu+1} - x_\nu$ tends to zero, each of the sums s and S tends to a limit. If the limits are the same, their common value is the Riemann integral

$$\int_a^b f(x)\, dx.$$

In certain cases, e.g. if $f(x)$ is continuous, we can say definitely that this integral exists.

Suppose in particular that $f(x)$ takes the values 0 and 1 only, say $f(x) = 1$ in a set E, and $f(x) = 0$ elsewhere. Then it is easily seen that s is equal to the sum of the lengths of those intervals throughout which $f(x) = 1$, i.e. intervals consisting entirely of

points of E; while S is the sum of lengths of intervals which include any point of E. If the set E consists of a finite number of intervals, there is no difficulty in proving that s and S tend to the same limit, viz. the sum of the lengths of the intervals of E.

The Riemann integral of such a function ($f(x) = 1$ in E, 0 elsewhere) may be called the *extent* of the set E. *Extent* is thus a generalization of the *length* of an interval. The extent of E, if it exists, is written $e(E)$, so that

$$e(E) = \int_a^b f(x)\, dx.$$

Whether the extent exists or not, the limits of s and S exist. These limits are called the interior and exterior extents* of E, and are written $e_i(E)$, $e_e(E)$.

The function $f(x)$ is called the *characteristic function* of the set E.

It is easy to define a set which has no extent. Let E be the set of all rational values of x in (a, b). Since every interval contains both rational and irrational numbers, we have $m_\nu = 0$, $M_\nu = 1$, for all modes of division and all values of ν. Hence $s = 0$, $S = b - a$, and consequently

$$e_i(E) = 0, \qquad e_e(E) = b - a.$$

The extent of this set is therefore undefined, and the characteristic function $f(x)$ has no Riemann integral.

In the general case we may say that the definition of the extent of E depends on the consideration of certain sets of intervals related to E, the number of such intervals being always finite.

Lebesgue's generalization is in the first place a generalization of extent; and it consists fundamentally in removing the restriction that our sets of intervals must be finite. Before we can introduce it formally we must make some further remarks about sets of points.

10.2. Sets of points. For the fundamental ideas concerning sets of points we refer to Hardy's *Pure Mathematics*, Chapter I.

We usually denote sets of points by E, E_1,... and suppose

* The exterior extent is sometimes called the content.

them all to lie within a finite interval (a, b). We denote by CE the complement of E, i.e. the set of all points of the interval (a, b) which do not belong to E.

If E_1 and E_2 are two sets, we denote by E_1+E_2 the set of all points belonging to E_1 or E_2, and by E_1E_2 the set of all points belonging to both E_1 and E_2. The notation is suggested by the fact that, if $f_1(x)$, $f_2(x)$ are the characteristic functions of E_1 and E_2, then $f_1(x)f_2(x)$ is the characteristic function of E_1E_2; while, if E_1 and E_2 have no common points, $f_1(x)+f_2(x)$ is the characteristic function of E_1+E_2.

Note that $$C(E_1+E_2) = CE_1 . CE_2.$$

The notation extends in an obvious way to any finite number of sets; also, if there are an infinity of given sets $E_1, E_2, \ldots$, then $E_1+E_2+\ldots$ denotes the set of points belonging to any of the given sets, and $E_1E_2\ldots$ denotes the set of points belonging to each of the given sets.

By $E_1 < E_2$ we mean that every point of E_1 is a point of E_2. Two sets are said to 'overlap' if they have common points.

An infinite set of points is said to be *enumerable* if it is possible to define a one-to-one correspondence between the points of the set and the integers 1, 2, 3,...; that is, we must be able to arrange the points in a sequence $x_1, x_2, x_3, \ldots$ such that every point occupies a definite place in the sequence. For example, the set of numbers $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ is enumerable; so is the set $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$.

The set of all proper rational fractions is enumerable; for we can arrange them as follows:

$$\tfrac{1}{2}, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{4}, \tfrac{3}{4}, \ldots$$

taking the denominators in order of magnitude, then the numerators.

The 'sum' of two enumerable sets is enumerable; for if E_1 consists of the points $x_1, x_2, \ldots$, and E_2 of $\xi_1, \xi_2, \ldots$, then all points of E_1+E_2 are given by the sequence

$$x_1, \xi_1, x_2, \xi_2, \ldots.$$

A similar argument applies to any finite number of enumerable sets. Further, *the sum of an enumerable infinity of enumerable sets is enumerable*; for let the sets be $E_1, E_2, \ldots$, and let E_n consist of the points

$$x_{1,n}, x_{2,n}, \ldots, x_{m,n}, \ldots.$$

We can arrange the double infinity of points $x_{m.n}$ as a single infinity in various ways, e.g. by taking together points for which $m+n=k$ $(k=2,3,\ldots)$, and in each such group taking m increasing; thus

$$x_{1,1},\ x_{1,2},\ x_{2,1},\ x_{1,3},\ x_{2,2},\ x_{3,1},\ x_{1,4},\ldots.$$

This proves the theorem.

Finally *a sub-set of an enumerable set is enumerable.* For any sub-set of x_1, x_2, $x_3,\ldots$ clearly has a first member, a second member, a third member, and so on, and this gives the required enumeration.

10.201. The reader might begin to suspect that all sets were enumerable; but this is not the case. *The set of all numbers between* 0 *and* 1 *is not enumerable.*

To prove this, suppose on the contrary it were possible to arrange all such numbers in a sequence x_1, $x_2,\ldots$. Suppose each such number expressed as an infinite decimal ('terminating' decimals end with an infinity of 0's; we exclude a recurring 9). We then form a new decimal ξ, such that, for every value of n, the nth term in the decimal for ξ exceeds by 1 the nth term in the decimal for x_n, if it is 0, 1,..., 7, and is 0 if it is 8 or 9. This rule defines ξ completely, and ξ does not end with a recurring 9. But ξ is a number between 0 and 1, and is different from any of the numbers x_n. This contradicts the assumption that the sequence x_n contains *all* the numbers between 0 and 1.

A similar argument applies to any interval. We call all the points of an interval a *continuum*. Our result is that *a continuum is not enumerable.*

10.202. A point ξ is called a 'limit-point' of a set E if, however small δ may be, there are points of E, other than ξ, in the interval $(\xi-\delta,\xi+\delta)$. (See *P.M.*, p. 30, where a limit-point is called a 'point of accumulation'.)

A set which contains all its limit-points is called a *closed* set. Thus an interval together with its end-points is a closed set. Such an interval is called a closed interval.

An *open* interval is an interval without its end-points. An *open* set is the complement of a closed set with respect to an open interval.

An open set consists of an enumerable set of non-overlapping

open intervals. For let E be an open set, and let x be a point of E. Then, for sufficiently small values of δ, the interval $(x, x+\delta)$ consists entirely of points of E; for otherwise x would be a limit-point of CE, so that CE would not be closed. Let δ_1 be the upper bound of values of δ with this property. Then ξ belongs to E for $x \leqslant \xi < x+\delta_1$; but $x+\delta_1$ is not a point of E, since, if it were, the interval of points of E would extend beyond it, by the above argument.

Similarly there is a number δ_2 such that ξ is in E for $x-\delta_2 < \xi \leqslant x$, while $x-\delta_2$ is not in E.

Thus x is a point of an open interval $(x-\delta_2, x+\delta_1)$ of points of E.

Similarly all points of E fall into open intervals. To arrange these intervals as an enumerable sequence, take first the interval, if there is one, greater than $\frac{1}{2}(b-a)$; next, in the order in which they occur on the line, those whose length is $\leqslant \frac{1}{2}(b-a)$ and $> \frac{1}{3}(b-a)$; and so on. Every interval of E has a definite place in this enumeration.

The 'sum' of two open sets is an open set. For if $E = E_1+E_2$, and E_1 and E_2 are open, every point of E is an interior point of an interval of points of E.

The same argument shows that *the sum of any finite number, or of an enumerable infinity, of open sets is open.* In particular (the converse of the above theorem), *the sum of an infinity of open intervals is an open set.*

Also *if E_1 and E_2 are open sets, then E_1E_2 is open.* For a point of E_1E_2 is an interior point of intervals both of E_1 and of E_2; and so it is not a limit-point of $C(E_1E_2)$, which consists of points of either CE_1 or CE_2.

This argument cannot be extended to an infinity of sets; e.g. if E_n is the open interval $-1/n < x < 1/n$, then $E_1E_2\ldots$ is the single point $x = 0$.

10.21. The measure of a set of points. We are now in a position to define a new generalization of 'length'. Instead of starting from a finite number of intervals, we start from an open set, which may contain an infinity of intervals.

The measure of an open set is defined to be the sum of the lengths of its intervals. This sum is, in general, the sum of an infinite series. It is always convergent, since the sum of any

finite number of terms is the sum of the lengths of a finite number of non-overlapping intervals, all contained in an interval (a, b), and so is not greater than $b-a$. Hence the measure of any open set contained in (a, b) does not exceed $b-a$.

The exterior measure of a set E is the lower bound of the measures of all open sets which contain E. It is denoted by $m_e(E)$. It is clear that

$$0 \leqslant m_e(E) \leqslant b-a,$$

and that, if $E_1 \prec E_2$, then $m_e(E_1) \leqslant m_e(E_2)$.

The interior measure, $m_i(E)$, is defined by the formula

$$m_i(E) = b-a-m_e(CE).$$

If $m_i(E) = m_e(E)$, then the set E is said to be measurable, and the common value of $m_i(E)$ and $m_e(E)$ is called its measure, and is denoted by $m(E)$.

We have also

$$m_i(CE) = b-a-m_e(E).$$

If E is measurable, so that $m_i(E) = m_e(E)$, it follows that $m_i(CE) = m_e(CE)$. Hence CE is measurable, and

$$m(E)+m(CE) = b-a.$$

Notice that we have given two definitions of the measure of an open set, one direct and one indirect. It will appear before long that they are equivalent. Meanwhile, in arguments involving open sets, we use the direct definition.

10.22. *For any set E we have*

$$m_i(E) \leqslant m_e(E).$$

For, by the definition of exterior measure, there are open sets O and O', including E and CE respectively, and such that

$$m(O) < m_e(E)+\epsilon,$$
$$m(O') < m_e(CE)+\epsilon.$$

If $\epsilon > 0$, every point of the interval $(a+\epsilon, b-\epsilon)$ is an interior point of an interval of O or of O'; and so, by the Heine-Borel theorem,* we can select from these intervals a finite set, say Q, which together include $(a+\epsilon, b-\epsilon)$. Then plainly

$$m(Q) \geqslant b-a-2\epsilon$$

and

$$m(Q) \leqslant m(O)+m(O').$$

* *P.M.* § 105. In the proof there given we start with an interval ending at a, whereas here there is an interval including a. This does not affect the proof.

Combining these inequalities we have

$$b-a < m_e(E)+m_e(CE)+4\epsilon.$$

Making $\epsilon \to 0$, it follows that

$$b-a \leqslant m_e(E)+m_e(CE),$$

which is equivalent to the result stated.

If $m_e(E)=0$, it follows that $m_i(E)=0$. Hence E is measurable, and its measure is zero.

10.23. We now come to the two fundamental theorems in the theory of measure.

First fundamental theorem. *If $E_1, E_2,\ldots, E_n,\ldots$ are measurable sets, then the set $E=E_1+E_2+E_3+\ldots$ is measurable, and*

$$m(E) \leqslant m(E_1)+m(E_2)+\ldots.$$

If $E_1, E_2,\ldots$ do not overlap, then the equality holds. (Otherwise the series may diverge.)

Second fundamental theorem. *If $E_1, E_2,\ldots$ are measurable sets, then the set $E_1E_2E_3\ldots$ is measurable.*

That is, the set of points belonging to any of the sets $E_1, E_2,\ldots$ is measurable, and so is the set of points belonging to all of them.

We shall begin by proving two lemmas on open sets, the first of which is the first fundamental theorem for open sets. We next prove a general theorem on exterior measure, and deduce from it the first fundamental theorem for the case where the sets do not overlap. Then we obtain the second theorem for two sets, and use it to complete the first theorem. Finally we use this result to complete the second theorem.

10.24. *If $O_1, O_2,\ldots$ are open sets (overlapping or not), and*

$$O=O_1+O_2+O_3+\ldots,$$

then

$$m(O) \leqslant m(O_1)+m(O_2)+\ldots. \tag{1}$$

We assume the convergence of the series on the right, since otherwise the theorem is meaningless.

Let the intervals of O_n be $(a_{m,n}, b_{m,n})$ $(m=1, 2,\ldots)$, and let those of O be (A_k, B_k) $(k=1, 2,\ldots)$. Let ϵ be a positive number less than $\frac{1}{2}(B_k-A_k)$. Then every point of the interval $(A_k+\epsilon, B_k-\epsilon)$ is an interior point of one of the intervals $(a_{m,n}, b_{m,n})$ which make up (A_k, B_k). If $\sum_k$ denotes a summation over these intervals, it follows from the Heine-Borel theorem, as in the previous proof, that

$$B_k - A_k - 2\epsilon \leqslant \sum_k (b_{m,n} - a_{m,n}).$$

Making $\epsilon \to 0$, we obtain

$$B_k - A_k \leqslant \sum_k (b_{m,n} - a_{m,n}), \tag{2}$$

and, summing with respect to k,

$$m(O) \leqslant \sum_{k=1}^{\infty} \sum_k (b_{m,n} - a_{m,n}). \tag{3}$$

Since a convergent double series of positive terms can be summed in any manner, the right-hand side of (3) can be rearranged in the form

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (b_{m,n} - a_{m,n}) = \sum_{n=1}^{\infty} m(O_n).$$

This proves the theorem.

If none of the sets overlap, each interval (A_k, B_k) coincides with one interval $(a_{m,n}, b_{m,n})$, and the inequalities (2) and (3), and so also (1), become equalities.

An enumerable set is measurable, and its measure is zero. For let the set be $x_1, x_2, \ldots$. Include x_1 in an open interval of length ϵ. If this does not include x_2, we can include x_2 in an interval of length $\frac{1}{2}\epsilon$; and so generally x_n in an interval of length $\epsilon/2^n$. Thus the given set can be included in an open set of measure not greater than 2ϵ. Since ϵ may be as small as we please, the exterior measure of the set is zero. Hence its measure is zero.

10.241. *If O and O' are open sets which together include all points of the interval (a, b), then*

$$m(OO') \leqslant m(O) + m(O') - (b-a).$$

By the Heine-Borel theorem we can select finite sets of the intervals of O and O', say Q from O and Q' from O', such that Q and Q' together include the whole interval $(a+\epsilon, b-\epsilon)$; and we may, by adding further intervals if necessary, suppose that

$$O = Q + R, \qquad O' = Q' + R',$$

where $m(R) < \epsilon$, $m(R') < \epsilon$. Now

$$OO' < QQ' + R + R',$$

so that by the previous lemma

$$m(OO') \leqslant m(QQ') + m(R) + m(R') < m(QQ') + 2\epsilon.$$

But $m(Q)+m(Q')-m(QQ') \geqslant b-a-2\epsilon$, from elementary considerations, and $m(O) \geqslant m(Q)$, $m(O') \geqslant m(Q')$. Making $\epsilon \to 0$, the result follows.*

10.25. *If E_1, E_2,... are any sets, and*

$$E = E_1+E_2+\ldots,$$

then
$$m_e(E) \leqslant m_e(E_1)+m_e(E_2)+\ldots.$$

We can enclose E_n in an open set O_n such that

$$m(O_n) < m_e(E_n)+\frac{\epsilon}{2^n}.$$

Summing with respect to n, and using the result of § 10.24,

$$m(O) \leqslant m(O_1)+m(O_2)+\ldots < m_e(E_1)+m_e(E_2)+\ldots+\epsilon.$$

But O is an open set which includes E. Hence

$$m_e(E) \leqslant m(O).$$

Hence
$$m_e(E) < m_e(E_1)+m_e(E_2)+\ldots+\epsilon,$$

and, making $\epsilon \to 0$, the result follows.

10.26. *If E_1, E_2,... are non-overlapping measurable sets, and*

$$E = E_1+E_2+\ldots,$$

then E is measurable, and

$$m(E) = m(E_1)+m(E_2)+\ldots.$$

We may suppose that all the sets are included in (a, b).

(i) Consider first the case of two sets, $E = E_1+E_2$. We know already that

$$m_e(E) \leqslant m_e(E_1)+m_e(E_2) = m(E_1)+m(E_2).$$

Hence it is sufficient to prove that

$$m_i(E) \geqslant m(E_1)+m(E_2),$$

i.e. that
$$m_e(CE) \leqslant m(CE_1)+m(CE_2)-(b-a).$$

Now we can include CE_1, CE_2, in open sets O_1, O_2, such that

$$m(O_1) < m(CE_1)+\epsilon, \qquad m(O_2) < m(CE_2)+\epsilon.$$

Since E_1 and E_2 have no common points, CE_1 and CE_2 together include the whole interval, and hence so do O_1 and O_2. Hence

$$m(O_1O_2) \leqslant m(O_1)+m(O_2)-(b-a).$$

* Actually the two sides are equal. This follows in due course from the first fundamental theorem.

But O_1O_2 includes CE. Hence

$$m_e(CE) \leqslant m(O_1O_2) \leqslant m(O_1)+m(O_2)-(b-a)$$
$$< m(CE_1)+m(CE_2)+2\epsilon-(b-a),$$

and, making $\epsilon \to 0$, the result follows.

(ii) The theorem for any finite number of sets follows by repeated application of (i).

(iii) In the case of an infinity of sets, we have, for all values of n,

$$m(E_1)+m(E_2)+\ldots+m(E_n) = m(E_1+\ldots+E_n) \leqslant b-a.$$

Hence $\sum m(E_n)$ is convergent.

Let $S_n = E_1+\ldots+E_n$. Then $CE < CS_n$, so that

$$m_e(CE) \leqslant m_e(CS_n) = m(CS_n) = b-a-m(E_1)-\ldots-m(E_n).$$

Making $n \to \infty$ we obtain

$$m_e(CE) \leqslant b-a-\sum m(E_n),$$

i.e.

$$m_i(E) \geqslant \sum m(E_n).$$

Combining this with § 10.25, the result follows.

In particular, taking $E_1, E_2, \ldots$ to be open intervals, it follows that any open set is measurable in the general sense, and that the two definitions of the measure of an open set agree. Also any closed set, as the complement of an open set, is measurable.

If E_1 and E_2 are measurable sets, E_1 being included in E_2, then E_2-E_1 is measurable.

For

$$C(E_2-E_1) = E_1+CE_2.$$

10.27. *If E and F are measurable sets, so is EF.*

Let both sets be included in (a,b), and suppose first that F is an interval (α,β). Let E_1 be the part of E in (α,β), E_2 the remainder. Similarly, if O is an open set containing E, let $O = O_1+O_2$. O_1 and O_2 are open sets containing respectively E_1 and E_2, if we neglect the points α and β, as we obviously may; and clearly

$$m(O) = m(O_1)+m(O_2).$$

Taking lower bounds,

$$m_e(E) = m_e(E_1)+m_e(E_2). \tag{1}$$

Similarly, if $e = CE = e_1+e_2$,

$$m_e(e) = m_e(e_1)+m_e(e_2). \tag{2}$$

But, since E is measurable,

$$m_e(E)+m_e(e) = b-a, \tag{3}$$

and by § 10.25

$$m_e(E_2)+m_e(e_2) \geqslant m_e(E_2+e_2) = b-a-(\beta-\alpha). \tag{4}$$

From (1), (2), (3) and (4) it follows that

$$m_e(E_1)+m_e(e_1) \leqslant \beta-\alpha,$$

and hence E_1 is measurable.

The result is therefore proved if F is an interval, and so, by the previous theorem, if F is an open set. In the general case we can include F in an open set O, and CF in O', so that $m(O)+m(O') < b-a+\epsilon$. Then

$$EF < EO, \qquad C(EF) = CF+F.CE < O'+O.CE,$$

so that

$$\begin{aligned} m_e(EF)+m_e\{C(EF)\} &\leqslant m(EO)+m(O')+m(O.CE) \\ &= m(O)+m(O') < b-a+\epsilon. \end{aligned}$$

Making $\epsilon \to 0$, $m_e(EF)+m_e\{C(EF)\} \leqslant b-a$, whence the result.

If E_1 and E_2 are measurable, the set E of points belonging to E_2 but not to E_1 is measurable.

For $E = E_2.CE_1$.

10.28. We can now complete the proofs of the fundamental theorems. Let E_1, E_2,... be any sets, overlapping or not, and let E be their sum. Let

$$E_2' = E_2.CE_1, \qquad E_3' = E_3.C(E_1+E_2'),$$
$$E_4' = E_4.C(E_1+E_2'+E_3'),$$

and so on. Then E_1, E_2', E_3',... are non-overlapping measurable sets, and $E = E_1+E_2'+E_3'+\ldots$. Hence E is measurable, by § 10.26, and the proof of the first fundamental theorem is completed, the inequality then stated following from § 10.25.

Again, if $F = E_1E_2E_3\ldots$, then

$$CF = CE_1+CE_2+\ldots.$$

Hence, by what has just been proved, CF is measurable, and so F is measurable. This proves the second fundamental theorem.

10.29. Limiting sets. *If E_1, E_2,... are measurable sets, each contained in the following one, and E is their sum, then*

$$\lim_{n\to\infty} m(E_n) = m(E).$$

For the sets E_2-E_1, $E_3-E_2,\ldots$ are measurable and non-overlapping, and

$$E = E_1+(E_2-E_1)+(E_3-E_2)+\ldots,$$

so that

$$m(E) = m(E_1)+m(E_2-E_1)+\ldots$$
$$= \lim\{m(E_1)+m(E_2-E_1)+\ldots+m(E_n-E_{n-1})\} = \lim m(E_n).$$

The set E is called the *outer limiting set* of the sets E_1, $E_2,\ldots$.

If each of the sets E_1, $E_2,\ldots$ contains the next, and $E = E_1E_2\ldots$, then

$$\lim_{n\to\infty} m(E_n) = m(E).$$

This follows by complementary sets from the previous theorem. In this case the set E is called the *inner limiting set.*

Unlike most of the theorems on the measure of sets, the first of these results holds if 'measure' is replaced by 'exterior measure', whether the sets are measurable or not. This remark will be useful in the next chapter, where it happens to be inconvenient to verify that certain sets are measurable.

If E is the outer limiting set of a sequence E_n, then

$$\lim_{n\to\infty} m_e(E_n) = m_e(E).$$

Let E_n be included in an open set O_n such that

$$m(O_n) < m_e(E_n)+\epsilon.$$

Let $S_n = O_nO_{n+1}O_{n+2}\ldots$, and let $S = S_1+S_2+\ldots$. Then $E_n < S_n < O_n$, $E < S$, and $S_n < S_{n+1}$, so that S is the outer limiting set of the sets S_n (this is not necessarily true for O_n, which is why we introduce S_n). Hence

$$m_e(E) \leqslant m(S) = \lim m(S_n) < \lim m_e(E_n)+\epsilon,$$

and, making $\epsilon \to 0$, $m_e(E) \leqslant \lim m_e(E_n)$. But since a set which includes E also includes E_n, $m_e(E) \geqslant m_e(E_n)$ for every n. This proves the theorem.

10.291. Cantor's ternary set. The following set of points, defined by Cantor, has many interesting properties.

Divide the interval $(0, 1)$ into three equal parts, and remove the interior of the middle part. Next subdivide each of the two remaining parts into three equal parts, and remove the interiors of the middle parts of each of them; and repeat this process indefinitely. Thus at the pth step we remove 2^{p-1} intervals.

We denote these intervals, from left to right, by $\delta_{p,k}$, where k runs from 1 to 2^{p-1}. For each k the length of $\delta_{p,k}$ is 3^{-p}.

Let E be the set of points which remain. Then E is the set of points represented by the infinite decimals

$$\cdot a_1 a_2 \ldots a_n \ldots (3)$$

in the scale of 3 (indicated by the final figure), where the numbers $a_1, a_2, \ldots$ take the values 0 or 2 only, never the value 1; for example, E includes $\frac{2}{3} = \cdot 200\ldots$, and also $\frac{1}{3}$, which can be represented as $\cdot 0222\ldots$. In fact the first step described above removes from the interval all points for which the first figure is a 1 (except $\cdot 100\ldots = \cdot 022\ldots$); the second step removes all remaining points for which the second figure is a 1 (except $\cdot 010\ldots = \cdot 0022\ldots$, and $\cdot 210\ldots = \cdot 2022\ldots$); and so on. Notice also that the end-points of the intervals $\delta_{p,k}$ consist of all decimals $\cdot a_1 a_2 \ldots (3)$, where the digits after a certain point are all 0's or all 2's. This is obviously true for $\delta_{1,1}$; then $\delta_{2,1}$, $\delta_{2,2}$ are obtained by taking the first decimal as 0 or 2 and then the rest as the decimals corresponding to the ends of $\delta_{1,1}$; and so on. Thus the general form of the end-points of a $\delta_{p,k}$ is

$$\cdot a_1 \ldots a_m 0222\ldots (3), \qquad \cdot a_1 \ldots a_m 2000\ldots (3).$$

The set E is not enumerable; this may be proved in the same way that it was proved that the continuum was not enumerable. On the other hand, the measure of E is zero; for

$$m(E) = 1 - \sum m(\delta_{p,k}) = 1 - \sum_{p=1}^{\infty} \frac{2^{p-1}}{3^p} = 0.$$

We shall refer to this set again in § 11.72.

Example. Prove that the measure of the set of points in the interval (0, 1) representing numbers whose expressions as infinite decimals do not contain some particular digit (say 7) is zero.

10.3. Measurable functions. Let $f(x)$ be a bounded function of x in the interval $a \leqslant x \leqslant b$. We denote by $E(f > c)$ the set of points in (a, b) where $f(x) > c$; and similarly with other inequalities.

The function $f(x)$ is said to be measurable if any one of the sets

$$E(f \geqslant c), \qquad E(f < c), \qquad E(f > c), \qquad E(f \leqslant c)$$

is measurable for all values of c.

Any one of these four conditions implies the other three.

Suppose, for example, that the first holds. The second follows by complementary sets. Hence also the sets

$$E_n = E\left(f < c + \frac{1}{n}\right) \qquad (n = 1, 2, \ldots)$$

are all measurable. Hence the set

$$(E_1 - E_2) + (E_2 - E_3) + \ldots = E(c < f < c+1)$$

is measurable. Hence

$$E(f = c) = E(f \geqslant c) - E(f \geqslant c+1) - E(c < f < c+1)$$

is measurable, and the result clearly follows from this.

10.31. General properties of measurable functions.

(i) *Let f be a measurable function, k a constant. Then $k+f$, kf, and in particular $-f$, are measurable.*

This is obvious.

(ii) *If f and ϕ are measurable functions, the set $E(f > \phi)$ is measurable.*

If $f > \phi$, there is a rational number r such that $f > r > \phi$. Hence

$$E(f > \phi) = \sum_r E(f > r) \,.\, E(\phi < r)$$

where r runs through all rational numbers. Hence the result.

(iii) *If f and ϕ are measurable, so are $f+\phi$ and $f-\phi$.*

For

$$E(f+\phi > c) = E(f > c-\phi)$$

and the result follows from (ii). Similarly for $f-\phi$.

(iv) *If f and ϕ are measurable, so is $f\phi$.*

The function $\{f(x)\}^2$ is measurable, for, if $c > 0$,

$$E(f^2 > c) = E(f > \sqrt{c}) + E(f < -\sqrt{c}).$$

The general theorem then follows from the fact that

$$f\phi = \tfrac{1}{4}(f+\phi)^2 - \tfrac{1}{4}(f-\phi)^2.$$

(v) *If $f_n(x)$ is a sequence of measurable functions, then*

$$\varliminf_{n\to\infty} f_n(x), \qquad \varlimsup_{n\to\infty} f_n(x),$$

supposed finite, are measurable. In particular, if the sequence tends to a limit, the limit is measurable.

Let $f(x) = \varlimsup f_n(x)$. Let c be any real number, let

$$E_{m,n} = E\left(f_n > c + \frac{1}{m}\right) + E\left(f_{n+1} > c + \frac{1}{m}\right) + \ldots,$$

and let $E_m = E_{m,1} E_{m,2} E_{m,3} \ldots$. By the fundamental theorems

$E_{m,n}$ and E_m are measurable. Now E_m, the set of points common to all the sets $E_{m,n}$, is the set where $f_\nu > c+1/m$ for arbitrarily large values of ν. Hence

$$f = \overline{\lim} f_\nu \geqslant c + \frac{1}{m} > c$$

in E_m. Let $E = E_1 + E_2 + E_3 + \ldots$. Then E is measurable, and $f > c$ at all points of E. Conversely, if $f(x) > c$, then there is an integer m such that $f_\nu(x) > c + 1/m$ for arbitrarily large values of ν, and so x belongs to one of the sets E_m. Hence $E = E(f > c)$, which proves the theorem.

(vi) *A continuous function is measurable.* For if $f(x)$ is continuous, it is easily seen that $E(f \leqslant c)$ is closed. Hence $E(f > c)$ is open, and so measurable.

All the ordinary functions of analysis may be obtained by limiting processes from continuous functions, and so are measurable. The same thing is true of some of the more artificial functions. For example,

$$\lim_{n\to\infty} \{\cos m!\pi x\}^{2n}$$

is the limit of a continuous function, and is equal to 1 if $m!x$ is an integer, and otherwise is zero. If x is rational, $m!x$ is an integer if m is large enough. Hence

$$f(x) = \lim_{m\to\infty} \lim_{n\to\infty} \{\cos m!\pi x\}^{2n}$$

is equal to 1 if x is rational, and to 0 otherwise. The fact that this function is measurable has, of course, been proved more directly (§ 10.22).

10.4. The Lebesgue integral of a bounded function. We are now in a position to define the Lebesgue integral of any bounded measurable function.

If $f(x)$ is the characteristic function of a set E, i.e. $f(x) = 1$ in E and 0 elsewhere, a natural definition of the integral is

$$\int_a^b f(x)\, dx = m(E).$$

If $f(x) = k$ in E and 0 elsewhere, then we take

$$\int_a^b f(x)\, dx = km(E).$$

In the general case, let α and β be the lower and upper bounds of $f(x)$. As in the case of Riemann integration, the integral is defined as the limit of the sum; but this time the sum is obtained by dividing up the interval of variation of $f(x)$. We take numbers $y_0, y_1, \ldots, y_{n+1}$ such that

$$\alpha = y_0 < y_1 < y_2 < \ldots < y_{n-1} < y_n = \beta.$$

Let e_ν be the set where $y_\nu \leqslant f(x) < y_{\nu+1}$ $(\nu = 0, \ldots, n-1)$, and e_n the set where $f(x) = \beta$. Since $f(x)$ is measurable, all the sets e_ν are measurable. Putting $y_{n+1} = \beta$, let

$$s = \sum_{\nu=0}^{n} y_\nu m(e_\nu), \qquad S = \sum_{\nu=0}^{n} y_{\nu+1} m(e_\nu).$$

The Lebesgue integral of $f(x)$ over (a, b) is the common limit of the sums s and S when the number of division-points y_ν is increased indefinitely, so that the greatest value of $y_{\nu+1} - y_\nu$ tends to zero.

To justify the definition we have to prove that the two limits exist and are equal.

Suppose the interval (α, β) divided up in two different ways, each difference $y_{\nu+1} - y_\nu$ in each way being less than ϵ. Let the sums formed in these two ways be s, S and s', S'. Then

$$S - s = \sum_{\nu=0}^{n} (y_{\nu+1} - y_\nu) m(e_\nu) \leqslant \epsilon \sum_{\nu=0}^{n} m(e_\nu) = \epsilon(b-a),$$

and similarly $S' - s' \leqslant \epsilon(b-a)$.

We now divide up the interval (α, β) by taking all the division-points of the first two ways at once. This gives two more sums, s'' and S''. Now the insertion of a new division-point does not decrease a lower sum or increase an upper sum; for example, if we insert a point η between y_ν and $y_{\nu+1}$ we have

$$y_\nu m(e_\nu) \leqslant y_\nu m\{E(y_\nu \leqslant f < \eta)\} + \eta m\{E(\eta \leqslant f < y_{\nu+1})\},$$

so that the lower sum is not decreased. Applying this principle repeatedly, we obtain

$$s \leqslant s'', \qquad s' \leqslant s'',$$

and similarly

$$S'' \leqslant S, \qquad S'' \leqslant S'.$$

It follows that the intervals (s, S) and (s', S') have points in common, e.g. all points of the interval (s'', S''). Hence the numbers s, s', S, S' all lie within an interval of length $2\epsilon(b-a)$. The existence and equality of the limits then follow from the general principle of convergence.

10.41. Comparison with Riemann's definition. Perhaps the most obvious difference to the beginner is that, in Lebesgue's definition, we divide up the interval of variation of the function instead of the interval of integration. This, however, is comparatively unimportant. What is essential is that we use the general theory of 'measure' of sets instead of the more limited theory of 'extent'. It would be possible to build up an integral from integrals of characteristic functions, but using extent instead of measure. This would be substantially equivalent to Riemann's definition. On the other hand, it is possible to define an integral equivalent to Lebesgue's by dividing up the interval of integration in a suitable way.

In both Riemann's and Lebesgue's definitions we have upper and lower sums which tend to limits. In the Riemann case the two limits are not necessarily the same, and the function is only integrable if they are the same. In the Lebesgue case the two limits are necessarily the same, their equality being a consequence of the assumption that the function is measurable.

Lebesgue's definition is more general than Riemann's. For the characteristic function of the set of rational points has a Lebesgue integral, but not a Riemann integral; and we shall see later that, if a function has a Riemann integral, then it also has a Lebesgue integral, and the two are equal.

We use the same notation

$$\int_a^b f(x)\,dx$$

for a Lebesgue integral as we have done for a Riemann integral. When it is necessary to distinguish a Riemann integral from a Lebesgue integral, we shall denote the former by

$$R\int_a^b f(x)\,dx.$$

10.42. Integral over any measurable set. Let E be any measurable set contained in an interval (a, b). The integral of $f(x)$ over E may be defined in the same way as the integral over an interval. The sets e_ν of § 10.4 are now the sub-sets of E where $y_\nu \leqslant f(x) < y_{\nu+1}$; the proof of the existence of the

integral is practically unchanged. The integral is written

$$\int_E f(x)\,dx.$$

Any integral over a set of measure zero is zero. For all the sets e_ν are of measure zero, and so the sums s and S are always 0.

We might also define the integral by putting $f(x)=0$ in CE, and then using the definition of the integral over an interval. It is easily seen that the two definitions are equivalent.

10.43. Henceforward we shall assume that all sets and functions introduced are measurable, without always saying so explicitly.

10.44. Elementary properties of the integral of a bounded function.

(i) **The mean-value theorem.** *If* $\alpha \leqslant f(x) \leqslant \beta$, *then*

$$\alpha m(E) \leqslant \int_E f(x)\,dx \leqslant \beta m(E).$$

For it is easily seen that $\alpha m(E) \leqslant s \leqslant \beta m(E)$, and the result follows in the limit.

(ii) *The integral is additive for a finite number or for an enumerable infinity of non-overlapping sets included in a finite interval. That is, if*

$$E = E_1 + E_2 + \ldots,$$

then

$$\int_E f(x)\,dx = \int_{E_1} f(x)\,dx + \int_{E_2} f(x)\,dx + \ldots.$$

Suppose first that there are two sets, E_1 and E_2. Inserting division-points y_ν, the sets E, E_1, E_2 are divided into sub-sets e_ν, e_ν^1, e_ν^2, such that

$$m(e_\nu) = m(e_\nu^1) + m(e_\nu^2).$$

Hence

$$\int_{E_1} + \int_{E_2} = \lim \sum y_\nu m(e_\nu^1) + \lim \sum y_\nu m(e_\nu^2)$$
$$= \lim \sum y_\nu m(e_\nu) = \int_E.$$

Similarly for any finite number of sets.

If there are an infinity of sets, let S_n be the sum of the first n, R_n the remainder. Then

$$\int_E = \int_{S_n} + \int_{R_n}.$$

But, by the mean-value theorem, if $|f(x)| \leqslant M$, then

$$\left|\int_{R_n} f(x)\,dx\right| \leqslant M m(R_n),$$

and this tends to zero as $n \to \infty$, since the series $\sum m(E_n)$ is convergent. Hence

$$\int_E = \lim \int_{S_n} = \int_{E_1} + \int_{E_2} + \dots.$$

(iii) *If, in a set E, $f(x) \leqslant \phi(x)$, then*

$$\int_E f(x)\,dx \leqslant \int_E \phi(x)\,dx.$$

Take division-points y_ν, and define the sets e_ν by means of $f(x)$. Then, in e_ν, $\phi(x) \geqslant f(x) \geqslant y_\nu$. Hence

$$\int_E \phi(x)\,dx = \sum \int_{e_\nu} \phi(x)\,dx \geqslant \sum y_\nu m(e_\nu).$$

The right-hand side tends to $\int_E f(x)\,dx$, whence the result follows.

(iv) *The integral of the sum of a finite number of bounded measurable functions is the sum of the integrals of the separate functions.*

In the first place, if k is a constant,

$$\int_E (f+k)\,dx = \int_E f\,dx + \int_E k\,dx = \int_E f\,dx + km(E).$$

For calculate the sum s relative to $f(x)$ with the scale $y_0, y_1, \dots$, and the sum s' relative to $f(x)+k$ with the scale $y_0+k, y_1+k, \dots$.

Then $$s' = \sum (y_\nu + k)m(e_\nu) = s + km(E),$$

and the result follows in the limit.

Now consider any two functions $f(x)$ and $\phi(x)$. We have

$$\begin{aligned}\int_E \{f(x)+\phi(x)\}\,dx &= \sum \int_{e_\nu} (f+\phi)\,dx \\ &\geqslant \sum \int_{e_\nu} (y_\nu+\phi)\,dx \\ &= s + \int_E \phi\,dx\end{aligned}$$

by what has just been proved. Similarly, replacing y_ν by $y_{\nu+1}$, we obtain

$$\int_E (f+\phi)\,dx \leqslant S + \int_E \phi\,dx.$$

The result now follows in the limit.

The result for any finite number of functions is obtained by repeated application of the result for two functions.

(v) *If k is a constant,*

$$\int_E k f(x)\,dx = k \int_E f(x)\,dx.$$

This is obvious if $k = 0$. If $k > 0$, calculate the second integral

with the scale y_ν, and the first with the scale ky_ν. Then the sets e_ν are the same in each case, and $s = ks'$, whence the result.

(vi) *We have*

$$\left|\int_E f(x)\,dx\right| \leqslant \int_E |f(x)|\,dx.$$

Let E_1 be the set where $f(x) \geqslant 0$, E_2 the set where $f(x) < 0$.

Then
$$\int_E f\,dx = \int_{E_1} f\,dx - \int_{E_2} |f|\,dx,$$
$$\int_E |f|\,dx = \int_{E_1} f\,dx + \int_{E_2} |f|\,dx,$$

and the result is obvious.

(vii) A relation which holds except in a set of measure zero is said to hold *almost everywhere.*

Two functions which are equal almost everywhere have the same integral.

Let $f(x) = \phi(x)$ at all points of E, except in a set e of measure zero. Then

$$\int_E (f-\phi)\,dx = \int_e (f-\phi)\,dx + \int_{E.Ce} (f-\phi)\,dx.$$

The first term is zero because $m(e) = 0$, and the second because the integrand is everywhere zero. Hence

$$\int_E f\,dx = \int_E \phi\,dx.$$

(viii) *If* $f(x) \geqslant 0$ *and* $\int_E f(x)\,dx = 0$, *then* $f(x) = 0$ *almost everywhere in* E.

Let $E_0 = E(f = 0)$, and

$$E_n = E\big(M/(n+1) < f \leqslant M/n\big), \qquad n = 1, 2, \ldots,$$

where M is the upper bound of f. Then $E = E_0 + E_1 + E_2 + \ldots$; and

$$m(E_n) \leqslant \frac{n+1}{M} \int_{E_n} f\,dx \leqslant \frac{n+1}{M} \int_E f\,dx = 0.$$

Thus $m(E_n) = 0$ for $n = 1, 2, \ldots$, and the result follows.

10.5. Lebesgue's convergence theorem (theorem of bounded convergence). *Let $f_n(x)$ be a sequence of measurable functions such that $|f_n(x)| \leqslant M$ for all values of n, when x is in a set E, and let*

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for all values of x in E. Then

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx.$$

Since sets of measure zero can be omitted from the integrals, it is sufficient that the conditions should hold *almost everywhere.*

Since $|f_n(x)| \leqslant M$ for each n, $|f(x)| \leqslant M$. Hence $f(x)$ is integrable, and we have to prove that

$$\lim \int_E \{f(x)-f_n(x)\}\,dx = 0.$$

Let $g_n = |f-f_n|$, let ϵ be any positive number, and let

$$E_1 = E(\epsilon > g_1, g_2, \ldots), \qquad E_2 = E(g_1 \geqslant \epsilon > g_2, g_3, \ldots),$$
$$E_3 = E(g_2 \geqslant \epsilon > g_3, g_4, \ldots),$$

and so on. Then the sets E_k are measurable; they are non-overlapping, since $g_k \geqslant \epsilon$ in E_{k+1}, but not in $E_1, \ldots, E_k$, so that E_{k+1} has no point in common with $E_1, \ldots, E_k$; and every point of E belongs to some E_k; for $g_n(x) \to 0$ for every x, so that to every x corresponds a first number k such that $g_k, g_{k+1}, \ldots$ are all less than ϵ, and then x belongs to E_k.

It follows that

$$\int_E g_n\,dx = \int_{E_1} g_n\,dx + \int_{E_2} g_n\,dx + \ldots.$$

Now $g_n < \epsilon$ in $E_1, \ldots, E_n$, and $g_n \leqslant 2M$ everywhere. Hence

$$\int_E g_n\,dx \leqslant \epsilon\{m(E_1)+\ldots+m(E_n)\}+2M\{m(E_{n+1})+\ldots\}.$$

Making $n \to \infty$, it follows that

$$\overline{\lim} \int_E g_n\,dx \leqslant \epsilon m(E).$$

Hence, making $\epsilon \to 0$, it follows that

$$\lim \int_E g_n\,dx = 0,$$

and the theorem follows.

The theorem is not true for Riemann integrals, because the function $f(x)$ is not necessarily integrable in Riemann's sense, even if each $f_n(x)$ is. For example, let $r_1, r_2, \ldots$ be the rational points in $(0, 1)$, and let $f_n(x) = 1$ if $x = r_1, r_2, \ldots$ or r_n, and $f_n(x) = 0$ elsewhere. Then

$$R\int_0^1 f_n(x)\,dx = 0$$

for every n; but $f(x)=1$ for every rational x, and $f(x)=0$ for irrational x, so that $f(x)$ is not integrable in Riemann's sense.

10.51. The theorem of bounded convergence may be stated as a theorem on term-by-term integration of series. *If the series*

$$u_1(x)+u_2(x)+\ldots$$

converges in a set E to $s(x)$, and its partial sums

$$s_n(x)=u_1(x)+\ldots+u_n(x)$$

are bounded for all values of n, when x is in E, then

$$\int_E s(x)\,dx=\int_E u_1(x)\,dx+\int_E u_2(x)\,dx+\ldots.$$

This is the final form of the theorem of bounded convergence proved for Riemann integrals in § 1.76.

10.52. Egoroff's theorem.* *If a sequence of functions converges to a finite limit almost everywhere in a set E, then, given δ, we can find a set of measure greater than $m(E)-\delta$ in which the sequence converges uniformly.*

Let $f_n(x)$ be the sequence, let E' be the set where $f_n(x)$ converges, say to $f(x)$, and let $g_n=|f-f_n|$.

Let $\epsilon_1,\ldots,\epsilon_r,\ldots$ be a sequence of positive numbers tending to zero. Let $S_{n,r}$ be the sub-set of E' where $g_\nu<\epsilon_r$ for $\nu\geqslant n$. Then each of the sets $S_{1,r}, S_{2,r},\ldots$ is contained in the next, and their outer limiting set (§ 10.29) is E', since $g_\nu\to 0$ everywhere in E'. Hence we can determine $n(r)$ so that

$$m(E'-S_{n(r),r})<\frac{\delta}{2^r}.$$

Let $$S=S_{n(1),1}S_{n(2),2}\ldots S_{n(r),r}\ldots.$$

Then, in S, $g_n<\epsilon_r$ $(n\geqslant n(r))$ for all values of r, i.e. $g_n\to 0$ uniformly in S; and

$$m(E-S)=m(E'-S)\leqslant\sum_{r=1}^{\infty}m(E'-S_{n(r),r})<\sum_{r=1}^{\infty}\frac{\delta}{2^r}=\delta.$$

This proves the theorem.

Example. Use Egoroff's theorem to prove Lebesgue's convergence theorem.

10.6. *If $f(x)$ has a Riemann integral over (a,b), then it has a Lebesgue integral over the same interval, and the two are equal.*

The result is easily proved if we assume that $f(x)$ is measurable;

* Egoroff (1).

for then it certainly has a Lebesgue integral. Dividing up the interval (a, b) by the points $x_0, x_1, \ldots, x_n$, and denoting by m_ν, M_ν the lower and upper bounds of $f(x)$ in $x_\nu < x \leqslant x_{\nu+1}$, we have

$$\sum_{\nu=0}^{n-1} m_\nu(x_{\nu+1}-x_\nu) \leqslant \sum_{\nu=0}^{n-1} \int_{x_\nu}^{x_{\nu+1}} f(x)\,dx \leqslant \sum_{\nu=0}^{n-1} M_\nu(x_{\nu+1}-x_\nu).$$

The middle term is the Lebesgue integral, while each of the extreme terms tends to the Riemann integral. Hence they are equal.

To prove that $f(x)$ is necessarily measurable if it has a Riemann integral, let

$$\phi(x) = m_\nu \quad (x_\nu < x \leqslant x_{\nu+1}), \qquad \Phi(x) = M_\nu \quad (x_\nu < x \leqslant x_{\nu+1}).$$

Then

$$\sum_{\nu=0}^{n-1} m_\nu(x_{\nu+1}-x_\nu) = \int_a^b \phi(x)\,dx, \qquad \sum_{\nu=0}^{n-1} M_\nu(x_{\nu+1}-x_\nu) = \int_a^b \Phi(x)\,dx.$$

Consider now an enumerable infinity of modes of division of the interval (a, b) such that $\max(x_{\nu+1}-x_\nu) \to 0$; and let each set of division-points contain the previous set. Let E be the set of all the division-points. E is enumerable and so of measure zero, and so may be neglected in integration. At any point x not in E, $\phi(x)$ does not decrease, and $\Phi(x)$ does not increase, as we insert division points. Hence $\phi(x) \to m(x)$, $\Phi(x) \to M(x)$, where $m(x)$ and $M(x)$ are the 'lower and upper bounds of $f(x)$ at x', i.e. the limits of the lower and upper bounds in indefinitely small intervals containing x. Also $\phi(x)$ and $\Phi(x)$ are measurable, and hence so are $m(x)$ and $M(x)$; and, by Lebesgue's convergence theorem,

$$\lim \int_a^b \phi(x)\,dx = \int_a^b m(x)\,dx, \qquad \lim \int_a^b \Phi(x)\,dx = \int_a^b M(x)\,dx.$$

But if $f(x)$ has a Riemann integral, each of these limits is equal to it. Hence

$$\int_a^b \{M(x)-m(x)\}\,dx = 0.$$

Since $M(x) \geqslant m(x)$ it follows by § 10.44 (viii) that $M(x) = m(x)$ almost everywhere; and since $M(x) \geqslant f(x) \geqslant m(x)$ it follows that $f(x) = m(x)$ almost everywhere. Hence $f(x)$ is measurable.

10.7. The Lebesgue integral of an unbounded function. Let $f(x)$ be an unbounded measurable function, and suppose first that $f(x) \geqslant 0$. Let $\{f(x)\}_n$, or simply $(f)_n$, denote $f(x)$ at points where $f(x) \leqslant n$, but n where $f(x) > n$. Then $\{f(x)\}_n$ is bounded and measurable, and so integrable. We define the integral of $f(x)$ over the set E to be the limit, if it exists, of the integral of $\{f(x)\}_n$,

$$\int_E f(x)\,dx = \lim_{n\to\infty} \int_E \{f(x)\}_n\,dx.$$

For a positive function $f(x)$ to be integrable over E, it is clearly necessary and sufficient that

$$\int_E \{f(x)\}_n\,dx$$

should be bounded.

The integral of a negative function may be defined in a similar way. In the general case, let $f(x) \geqslant 0$ in E_1, $f(x) < 0$ in E_2. Then we define the integral of $f(x)$ by the equation

$$\int_E f(x)\,dx = \int_{E_1} f(x)\,dx + \int_{E_2} f(x)\,dx.$$

A function which is integrable in this sense is 'absolutely integrable', i.e. $|f(x)|$ is also integrable. In fact it is clear that

$$\int_E |f(x)|\,dx = \int_{E_1} f(x)\,dx - \int_{E_2} f(x)\,dx.$$

It would of course be possible to define integrals which are not absolutely convergent; but we shall see that integrals of the above kind preserve all the characteristic properties of integrals of bounded functions, whereas this would not be true of non-absolutely convergent integrals.

We shall henceforth use the word 'integrable' to describe any function, bounded or unbounded, which has an integral in the above sense.

The use of the expression 'infinity', introduced in § 5.701, is also very convenient here. For example, if

$$\int_E \{f(x)\}_n\,dx$$

tends to infinity with n, we write

$$\int_E f(x)\,dx = \infty.$$

Examples. (i) Show that $\int_0^1 x^{-a}\,dx$ exists as a Lebesgue integral, and is equal to $1/(1-a)$, if $0 < a < 1$; but is infinite if $a \geqslant 1$.

[The Lebesgue definition of the integral is

$$\lim_{n\to\infty}\left\{\int_0^{n^{-1/a}} n\,dx + \int_{n^{-1/a}}^1 x^{-a}\,dx\right\},$$

and the results are the same as in the elementary theory.]

(ii) More generally, let $f(x)$ be positive, and bounded in $(\epsilon, 1)$ for every positive ϵ. Then

$$\int_0^1 f(x)\,dx = \lim_{\epsilon\to 0}\int_\epsilon^1 f(x)\,dx$$

in the sense that both sides are finite and equal, or both infinite.

(iii) The function

$$f(x) = \frac{d}{dx}\left(x^2\sin\frac{1}{x^2}\right) = 2x\sin\frac{1}{x^2} - \frac{2}{x}\cos\frac{1}{x^2}$$

is not integrable in Lebesgue's sense over $(0, 1)$.

[The function is continuous over $(\epsilon, 1)$, and $\lim_{\epsilon\to 0}\int_\epsilon^1 f(x)\,dx$ exists. But

$$\int_0^1 |f(x)|\,dx = \infty;$$

for

$$|f(x)| \geqslant \frac{2}{x}\left|\cos\frac{1}{x^2}\right| - 2x \geqslant \frac{1}{x} - 2x$$

in each of the intervals $\{(2n+\frac{1}{3})\pi\}^{-\frac{1}{2}} \leqslant x \leqslant \{(2n-\frac{1}{3})\pi\}^{-\frac{1}{2}}$, and it is easily seen from this that

$$\int_0^1 \{|f(x)|\}_n\,dx > A\log n.]$$

(iv) Let $f(x)$ be any measurable function in E, and let e_n be the sub-set of E where $n-1 \leqslant f(x) < n$. Then the necessary and sufficient condition that $f(x)$ should be integrable over E is that $\sum_{n=-\infty}^{\infty} |n|m(e_n)$ should be convergent.

(v) We might define the integral of a positive unbounded function $f(x)$ by taking $\{f(x)\}^n = f(x)$ if $f(x) \leqslant n$, and otherwise $\{f(x)\}^n = 0$, and substituting $\{f(x)\}^n$ for $\{f(x)\}_n$ in Lebesgue's definition. Show that this definition is equivalent to that of Lebesgue.

(vi) If $|f(x)| \leqslant \phi(x)$, and $\phi(x)$ is integrable over E, then $f(x)$ is integrable over E.

(vii) If $f(x)$ is integrable over E, and E_n is the part of E where $|f(x)| \geqslant n$, then $m(E_n) = o\,(1/n)$.

(viii) If $f(x) = 0$ at every point of Cantor's ternary set, and $f(x) = p$ in each of the complementary intervals of length 3^{-p}, then

$$\int_0^1 f(x)\,dx$$

exists in Lebesgue's sense and is equal to 3.

10.71. Elementary properties of integrals. *The integral is additive, i.e. if E_1, E_2,... are non-overlapping sets, and $E = E_1 + E_2 + \ldots$, then*

$$\int_E f\,dx = \int_{E_1} f\,dx + \int_{E_2} f\,dx + \ldots.$$

We may suppose without loss of generality that $f \geqslant 0$; for if the result is true for positive functions it is true similarly for negative functions, and so by addition in the general case. This remark simplifies many of our proofs.

We define $(f)_n$ as before. The integral of $(f)_n$ is additive, so that

$$\int_E (f)_n\,dx = \sum_{k=1}^{\infty} \int_{E_k} (f)_n\,dx \leqslant \sum_{k=1}^{\infty} \int_{E_k} f\,dx.$$

Now make $n \to \infty$. If there are only a finite number of sets, the result follows (from the equality). If there are an infinity of sets we obtain (from the inequality)

$$\int_E f\,dx \leqslant \sum \int_{E_k} f\,dx.$$

But for any value of K

$$\int_E (f)_n\,dx \geqslant \sum_{k=1}^{K} \int_{E_k} (f)_n\,dx.$$

Making $n \to \infty$ first, and then $K \to \infty$, we obtain

$$\int_E f\,dx \geqslant \sum \int_{E_k} f\,dx.$$

Hence the result. (Notice the analogy with the proof given in § 1.62 that a double series of positive terms may be summed by rows or by columns to the same sum.)

10.72. *The sum of a finite number of integrable functions is integrable, and the integral of the sum is the sum of the integrals of the separate functions.*

It is sufficient to consider two functions, say $f(x)$ and $g(x)$. Suppose first that they are both positive, and let $\phi = f + g$. Then

$$(\phi)_n \leqslant (f)_n + (g)_n \leqslant (\phi)_{2n}.$$

Hence

$$\int_E (\phi)_n\,dx \leqslant \int_E (f)_n\,dx + \int_E (g)_n\,dx \leqslant \int_E (\phi)_{2n}\,dx,$$

and making $n \to \infty$

$$\int_E \phi\,dx \leqslant \int_E f\,dx + \int_E g\,dx \leqslant \int_E \phi\,dx.$$

which gives the required result.

If $f \geqslant 0$, $g < 0$, consider the set where $\phi \geqslant 0$. Here

$$f = \phi + (-g),$$

and the result follows from the previous case. Similarly where $\phi < 0$ we consider $-g = f + (-\phi)$.

Having proved the result for the sum and difference of positive functions, the general result now follows.

10.73. The following results can easily be deduced from the corresponding results for bounded functions:

(i) *If k is a constant,*

$$\int_E kf\,dx = k\int_E f\,dx.$$

(ii)
$$\left|\int_E f\,dx\right| \leqslant \int_E |f|\,dx.$$

(iii) *Two functions which are equal almost everywhere have the same integral.*

(iv) *If $f(x) \geqslant 0$, $\int_E f(x) = 0$, then $f(x) = 0$ almost everywhere in E.*

(v) *If $f(x)$ is integrable over E, and $E_1, E_2, \ldots$ is a sequence of sets contained in E such that $m(E_k) \to 0$, then $\int_{E_k} f(x)\,dx \to 0$, and indeed uniformly for all such sequences of sets.*

For, supposing, as we may, that $f(x) \geqslant 0$, choose n so that

$$\int_E [f(x) - \{f(x)\}_n]\,dx < \epsilon.$$

Having fixed n, we have

$$\int_{E_k} \{f(x)\}_n\,dx \leqslant nm(E_k) < \epsilon \qquad (k > k_0).$$

Hence

$$\begin{aligned}\int_{E_k} f(x)\,dx &= \int_{E_k} \{f(x)\}_n\,dx + \int_{E_k} [f(x) - \{f(x)\}_n]\,dx \\ &\leqslant \int_{E_k} \{f(x)\}_n\,dx + \int_E [f(x) - \{f(x)\}_n]\,dx \\ &< 2\epsilon \qquad (k > k_0),\end{aligned}$$

and the result follows.

Example. Let $f(x)$ be integrable, and $\phi(x)$ integrable in Riemann's sense, over (a, b). Dividing up the interval (a, b) by points x_ν as in § 10.1, prove that, as $\max(x_{\nu+1}-x_\nu) \to 0$,

$$\lim \sum_{\nu=0}^{n-1} \phi(x_\nu) \int_{x_\nu}^{x_{\nu+1}} f(x)\, dx = \int_a^b \phi(x) f(x)\, dx.$$

[Titchmarsh (1).]

10.8. The general convergence theorem of Lebesgue. *If $f_n(x)$ is a sequence of functions such that $|f_n(x)| \leqslant F(x)$, where $F(x)$ is integrable over E, for all values of n and all values of x in E, and*

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for all values of x in E, then

$$\lim_{n\to\infty} \int_E f_n(x)\, dx = \int_E f(x)\, dx.$$

As usual, it is sufficient that the conditions should hold almost everywhere. The proof is almost the same as that of the theorem of bounded convergence. We define the sets E_n as before; by § 10.71 the series

$$\sum \int_{E_n} F(x)\, dx$$

is convergent, and we have

$$\int_E g_n\, dx \leqslant \epsilon\{m(E_1)+\ldots+m(E_n)\}+$$
$$+2\int_{E_{n+1}} F(x)\, dx + 2\int_{E_{n+2}} F(x)\, dx + \ldots.$$

Making $n \to \infty$ it follows that

$$\overline{\lim} \int_E g_n\, dx \leqslant \epsilon m(E),$$

and the result now follows as before.

The above theorem enables us to prove a new theorem on term-by-term integration of series. *We may multiply a boundedly convergent series by any integrable function, and integrate term by term.* For if $s_n(x)$ is the nth partial sum of the series, and $|s_n(x)| \leqslant M$, and $\phi(x)$ is the integrable function, we have

$$|\phi(x) s_n(x)| \leqslant M |\phi(x)|,$$

which is integrable, and may be taken as the $F(x)$ of the above proof.

10.81. The following theorem is often useful. Its original form is due to Fatou.*

If $f_n(x) \geqslant 0$ for all values of n, and x in E, and $f_n(x) \to f(x)$ as $n \to \infty$, then

$$\int_E f(x)\,dx \leqslant \varliminf_{n\to\infty} \int_E f_n(x)\,dx.$$

The statement implies that, if the right-hand side is finite, then $f(x)$ is finite almost everywhere and integrable; while, if $f(x)$ is not integrable, or is infinite in a set of positive measure, then

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \infty.$$

It is easily seen that, with the usual notation,

$$\lim_{n\to\infty} \{f_n(x)\}_k = \{f(x)\}_k.$$

Hence, by the theorem of bounded convergence,

$$\lim_{n\to\infty} \int_E \{f_n(x)\}_k\,dx = \int_E \{f(x)\}_k\,dx.$$

But
$$\int_E \{f_n(x)\}_k\,dx \leqslant \int_E f_n(x)\,dx,$$

and hence
$$\varliminf_{n\to\infty} \int_E f_n(x)\,dx \geqslant \int_E \{f(x)\}_k\,dx.$$

Making $k \to \infty$, the result follows at once if $f(x)$ is finite almos everywhere, the set where $f(x)$ is infinite being omitted from th integral. If $f(x) = \infty$ in a set e of positive measure, then

$$\int_E \{f(x)\}_k\,dx \geqslant km(e)$$

for all values of k, and the result follows.

10.82. A convergence theorem for monotonic sequence

Let $f_1(x)$, $f_2(x)$,... be a sequence of positive integrable function non-decreasing for every value of x in E. Let $f(x)$ be the lim finite or infinite, of the sequence. Then

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx$$

in the following sense:

(i) *if the left-hand side is finite, then $f(x)$ is finite almost eve where and integrable, and the equality holds;*

* Fatou (1), p. 375.

(ii) *if the right-hand side is finite, so is the left-hand side, and equality holds;*

(iii) *if the left-hand side is infinite, then $f(x)$ is not integrable or is infinite in a set of positive measure;*

(iv) *the converse of* (iii) *holds.*

If the left-hand side is finite, so is the right-hand side, by Fatou's theorem; and equality in cases (i) and (ii) follows from Lebesgue's convergence theorem, since $f_n(x) \leqslant f(x)$. Then (iii) follows from (ii) and (iv) from (i).

10.83. We can now put the theorem of § 1.77 on integration of series into a more satisfactory form.

If $u_n(x) \geqslant 0$ for all values of n and x, then

$$\int_a^b \{\sum u_n(x)\}\, dx = \sum \int_a^b u_n(x)\, dx,$$

provided that either side is convergent.

For the partial sum $s_n(x) = u_1(x) + \ldots + u_n(x)$ is positive, and non-decreasing for every value of x.

In particular, the convergence of the right-hand side implies the convergence of $\sum u_n(x)$ for almost all values of x.

We have still to consider the case where the range of integration is infinite; but as we have not yet discussed infinite Lebesgue integrals of this kind, we must postpone the complete result until the end of the next section.

10.9. Integrals over an infinite range. Let $f(x)$ be a function which is integrable over the interval (a, b), for all finite values of b. Let $f_1(x) = f(x)$ where $f(x) \geqslant 0$, and $f_1(x) = 0$ elsewhere; and let $f_2(x) = -f(x)$ where $f(x) < 0$, and $f_2(x) = 0$ elsewhere. Then

$$\int_a^b f(x)\, dx = \int_a^b f_1(x)\, dx - \int_a^b f_2(x)\, dx.$$

Each integral on the right is a non-decreasing function of b, and so tends to a finite limit or to positive infinity as $b \to \infty$. We write

$$\int_a^\infty f_1(x)\, dx = \lim_{b\to\infty} \int_a^b f_1(x)\, dx, \qquad \int_a^\infty f_2(x)\, dx = \lim_{b\to\infty} \int_a^b f_2(x)\, dx,$$

CHAPTER XI

DIFFERENTIATION AND INTEGRATION

11.1. Introduction. The 'fundamental theorem of the integral calculus' is that differentiation and integration are inverse processes. This general principle may be interpreted in two different ways. If $f(x)$ is integrable, the function

$$F(x) = \int_a^x f(t)\,dt \tag{1}$$

is called the indefinite integral of $f(x)$; and the principle asserts that

$$F'(x) = f(x). \tag{2}$$

On the other hand, if $F(x)$ is a given function, and $f(x)$ is defined by (2), the principle asserts that

$$\int_a^x f(t)\,dt = F(x) - F(a). \tag{3}$$

The main object of this chapter is to consider in what sense these theorems are true.

As in elementary theory, (2) follows from (1) for every value of x for which $f(x)$ is continuous. For we can choose h_0 so small that $|f(t)-f(x)| < \epsilon$ for $|t-x| \leqslant h_0$; and then

$$\left|\frac{F(x+h)-F(x)}{h} - f(x)\right| = \left|\frac{1}{h}\int_x^{x+h} \{f(t)-f(x)\}\,dt\right| \leqslant \epsilon \quad (|h| < h_0),$$

by the mean-value theorem. This proves (2).

However, in the Lebesgue theory we consider functions which are in general discontinuous, so that the above argument does not apply to them. Actually the interesting question is, not whether (2) holds at particular points, but whether it is true in general; and to this we can give a satisfactory answer.

If $f(x)$ is any integrable function, its indefinite integral $F(x)$ has almost everywhere a finite differential coefficient equal to $f(x)$.

The problem of deducing (3) from (2) is much more difficult. We require in the first place that $F'(x)$ should exist at any rate almost everywhere, and, as we shall see in § 11.22, this is not necessarily so. Secondly, if $F'(x)$ exists we require that it should

be integrable. If we were relying on the Riemann theory, we should find a fundamental difficulty here; for Volterra has shown by an example* that $F'(x)$ may exist everywhere and be bounded, and yet not be integrable in Riemann's sense. In the Lebesgue theory, a differential coefficient is measurable, and so integrable if it is bounded. But, if it is unbounded, it is not necessarily integrable in the Lebesgue sense. The problem has received a satisfactory answer, but it requires a more general process, known as totalization, or Denjoy integration, which we have not space to consider here. The result is that if $F'(x)$ is finite everywhere, then (3) follows from (2) if the integral is taken in the Denjoy sense.

11.2. Differentiation throughout an interval. The ordinary functions of analysis are differentiable in general, i.e. for most values of the variable, though there may be special points at which they are not differentiable. The exceptional points are usually isolated. This seems to have created the impression at one time that a continuous function necessarily has a differential coefficient in general. It was, however, shown by Weierstrass that this is quite untrue. *There is a continuous function which has no differential coefficient anywhere.*

Nevertheless, the idea that an 'ordinary function' has a differential coefficient in general is correct, if we attach this vague expression to a different class of functions. We shall see that it is true in the sense that *a monotonic function has a finite differential coefficient almost everywhere.*

We shall first consider non-differentiable functions, and then proceed to the constructive side of the theory.

11.21. Continuous non-differentiable functions. There are many simple examples of continuous functions which are not differentiable at particular points; for example, if $f(x) = |x|$, the ratio

$$\frac{f(h)-f(0)}{h}$$

tends to different limits, 1 and -1, as $h \to 0$ by positive or negative values; and if $f(x) = x \sin 1/x$ $(x \neq 0)$, $f(0) = 0$, the ratio does not tend to any definite limit.

We can next, by a method known as the condensation of

* Hobson, vol. i, p. 461.

singularities, construct continuous functions which are not differentiable in a set which is everywhere dense, for example in the set of rational points. Let $r_1, r_2, \ldots$ denote the rational numbers between 0 and 1, and let

$$F(x) = \sum_{n=1}^{\infty} a_n f(x - r_n),$$

where $f(x)$ has an assigned singularity at $x = 0$, and the coefficients a_n tend to zero sufficiently rapidly. Then $F(x)$ will have the assigned singularity at every rational point. For example,

$$F(x) = \sum_{n=1}^{\infty} \frac{|x - r_n|}{3^n}$$

is continuous, since the series is uniformly convergent; but it is not differentiable at any rational point; for

$$\frac{F(r_k+h) - F(r_k)}{h} = \sum_{n=1}^{k-1} \frac{|r_k+h-r_n| - |r_k-r_n|}{h.3^n} + \frac{|h|}{h.3^k} + \\ + \sum_{k+1}^{\infty} \frac{|r_k+h-r_n| - |r_k-r_n|}{h.3^n};$$

and as $h \to 0$ the first term tends to a limit, the second term tends to $\pm 1/3^k$ according as $h > 0$ or $h < 0$, and, if $|h| < 1$, the third term does not exceed

$$\sum_{k+1}^{\infty} \frac{1}{3^n} = \frac{1}{2 \,.\, 3^k}$$

in absolute value. Hence $F'(r_k)$ does not exist.

To obtain functions which are *everywhere* non-differentiable we have to use quite different methods. The first example of such a function was given by Weierstrass.

11.22. Weierstrass's non-differentiable function. This function is defined by the series

$$f(x) = \sum_{n=0}^{\infty} b^n \cos(a^n \pi x),$$

where $0 < b < 1$, and a is an odd positive integer. The series is uniformly convergent in any interval, so that $f(x)$ is everywhere continuous. On the other hand, if $ab > 1$, the series obtained by term-by-term differentiation is divergent. This in

itself does not prove that $f(x)$ is not differentiable, but it suggests possibilities in this direction. We shall prove that *if* $ab > 1+\frac{3}{2}\pi$, *the function has no finite differential coefficient for any value of* x.

We have

$$\frac{f(x+h)-f(x)}{h} = \sum_{n=0}^{\infty} b^n \frac{\cos\{a^n\pi(x+h)\}-\cos(a^n\pi x)}{h}$$

$$= \sum_{n=0}^{m-1} + \sum_{m}^{\infty} = S_m + R_m,$$

say. Now

$$|\cos\{a^n\pi(x+h)\}-\cos(a^n\pi x)| = |a^n\pi h \sin\{a^n\pi(x+\theta h)\}| \leqslant a^n\pi|h|,$$

so that

$$|S_m| \leqslant \sum_{n=0}^{m-1} \pi a^n b^n = \pi\frac{a^m b^m - 1}{ab-1} < \pi\frac{a^m b^m}{ab-1}.$$

We next obtain a lower limit for R_m, giving h a particular value. We can write

$$a^m x = \alpha_m + \xi_m,$$

where α_m is an integer, and $-\frac{1}{2} \leqslant \xi_m < \frac{1}{2}$. Let

$$h = \frac{1-\xi_m}{a^m}.$$

Then

$$0 < h \leqslant \frac{3}{2a^m}.$$

Also $\quad a^n\pi(x+h) = a^{n-m}.a^m\pi(x+h) = a^{n-m}\pi(\alpha_m+1).$

Since a is odd, it follows that

$$\cos\{a^n\pi(x+h)\} = (-1)^{a^{n-m}(\alpha_m+1)} = (-1)^{\alpha_m+1}.$$

Again

$$\cos(a^n\pi x) = \cos\{a^{n-m}\pi(\alpha_m+\xi_m)\} = \cos(a^{n-m}\pi\alpha_m)\cos(a^{n-m}\pi\xi_m)$$
$$= (-1)^{\alpha_m}\cos(a^{n-m}\pi\xi_m).$$

Hence

$$R_m = \frac{(-1)^{\alpha_m+1}}{h}\sum_{n=m}^{\infty} b^n\{1+\cos(a^{n-m}\pi\xi_m)\}.$$

All the terms of this series are positive, and hence, taking the first term only,

$$|R_m| > \frac{b^m}{|h|} > \frac{2}{3}a^m b^m.$$

Hence

$$\left|\frac{f(x+h)-f(x)}{h}\right| \geqslant |R_m|-|S_m| > \left(\frac{2}{3}-\frac{\pi}{ab-1}\right)a^m b^m.$$

If $ab > 1+\frac{3}{2}\pi$, the factor in brackets is positive; and when $m \to \infty$, $h \to 0$, and the expression on the right tends to infinity. Hence $\{f(x+h)-f(x)\}/h$ takes arbitrarily large values, so that $f'(x)$ does not exist or is not finite.

The graph of the function may be said to consist of an infinity of infinitesimal crinkles; but it is almost impossible to form any definite picture of it which does not obscure its essential feature.*

11.23. The following example of a continuous non-differentiable function is due to van der Waerden.† The function is similar to Weierstrass's, but the result is obtained in quite a different way.

Let $f_n(x)$ denote the distance between x and the nearest number of the form $m/10^n$, where m is an integer. Then the function

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is a continuous non-differentiable function.

Each $f_n(x)$ is continuous; and $|f_n(x)| < 10^{-n}$, so that the series is uniformly convergent. Hence $f(x)$ is continuous.

Let x be any number in the interval $(0, 1)$, and suppose it expressed as a decimal. If the qth figure is 4 or 9, let $x' = x-10^{-q}$; otherwise let $x' = x+10^{-q}$. Then if $n < q$, the nearest number $m/10^n$ is the same for x and x', and x and x' lie on the same side of it; while if $n \geqslant q$, the numbers $m/10^n$ and $m'/10^n$ corresponding to x and x' differ by $x-x'$. These rules may be verified by considering simple examples, such as $q = 2$, $x = \cdot 326$, $\cdot 346$, or $\cdot 396$.

It follows that

$$\begin{aligned} f_n(x')-f_n(x) &= \pm(x'-x) && (n < q) \\ &= 0 && (n \geqslant q). \end{aligned}$$

Hence $$f(x')-f(x) = \sum_{n=1}^{q-1} \pm(x'-x) = p(x'-x),$$

* For further properties of this function see Hardy (7), where the same result is obtained for $ab > 1$. A general method of constructing continuous non-differentiable functions is given by Knopp (2).

† Van der Waerden (1).

where p is an integer, and is odd or even with $q-1$. Hence $\{f(x')-f(x)\}/(x'-x)$ cannot tend to a finite limit as $x' \to x$.

11.3. The four derivates of a function. Whether the differential coefficient

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

exists or not, the four expressions

$$\overline{\lim_{h\to +0}} \frac{f(x+h)-f(x)}{h}, \qquad \underline{\lim_{h\to +0}} \frac{f(x+h)-f(x)}{h},$$

$$\overline{\lim_{h\to -0}} \frac{f(x+h)-f(x)}{h}, \qquad \underline{\lim_{h\to -0}} \frac{f(x+h)-f(x)}{h}$$

always have a meaning, being either finite, or positive or negative infinity. They are called the upper and lower derivates on the right, and the upper and lower derivates on the left, respectively. We shall denote them by

$$D^+f(x), \qquad D_+f(x), \qquad D^-f(x), \qquad D_-f(x)$$

respectively, the sign referring to that of h in the above ratio, and its position corresponding to the 'lower' or 'upper' limit. If $D^+f = D_+f$, the function is said to have a right-hand derivative, if $D^-f = D_-f$, a left-hand derivative. The necessary and sufficient condition for the existence of the ordinary differential coefficient is that all the derivates should be equal.

We denote the left-hand and right-hand derivatives, when they exist, by $f'_-(x)$ and $f'_+(x)$.

Examples. (i) The function $\sqrt{x^2}$, where the positive value of the square root is always taken, has different left-hand and right-hand derivatives at $x = 0$.

(ii) Let $f(x) = x \sin 1/x$ $(x \neq 0)$, 0 $(x = 0)$. Then at $x = 0$

$$D_+f = -1, \qquad D^+f = 1, \qquad D_-f = -1, \qquad D^-f = 1.$$

(iii) Let
$$f(x) = \begin{cases} ax\sin^2 1/x + bx\cos^2 1/x & (x > 0) \\ 0 & (x = 0) \\ a'x\sin^2 1/x + b'x\cos^2 1/x & (x < 0), \end{cases}$$

where $a < b$, $a' < b'$. Then at $x = 0$

$$D_+f = a, \qquad D^+f = b, \qquad D_-f = a', \qquad D^-f = b'.$$

(iv) If $f(x)$ is continuous in (a, b), and one of its derivates is non-negative in the interval, then $f(a) \leqslant f(b)$.

[Let $D^+f \geqslant 0$, for example. Suppose that $f(b)-f(a) < -\epsilon(b-a)$, and let $\phi(x) = f(x)-f(a)+\epsilon(x-a)$. Then $\phi(b) < 0$. Also $\phi(x) > 0$ for some