Mathematics Department Stanford University
Math 51H – Integrals

There is a class of functions on which one 'knows' what the integral should be. Namely, if $f : [a,b] \to \mathbb{R}$ is an affine function, i.e. $f(t) = \alpha t + \beta$ for some $\alpha, \beta \in \mathbb{R}$, then one should have

$$\int_a^b f = \Big(\alpha \frac{a+b}{2} + \beta\Big)(b-a)$$

since $b-a$ is the length of the interval, and $\alpha\frac{a+b}{2} + \beta$ is the 'average' value of the affine function: $\frac{f(a)+f(b)}{2}$. More generally, one should have that the integral is additive for $a < b < c$:

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

This gives that one also knows what the integral must be for piecewise affine functions: if $f : [a,b] \to \mathbb{R}$ and $a = t_0 < t_1 < \ldots < t_n = b$, with $f(t) = \alpha_i t + \beta_i$, $t \in [t_{i-1}, t_i]$, with $\alpha_i, \beta_i \in \mathbb{R}$, then we should have

$$\int_a^b f = \sum_{i=1}^n \Big(\alpha_i \frac{t_i + t_{i-1}}{2} + \beta_i\Big)(t_i - t_{i-1}).$$

Note that this is just like the more standard formula for piecewise constant (thus typically discontinuous) functions, but because we can remain in the setting of continuous functions, this is more convenient for us.

Now, it should not be completely 'obvious' that there can be a notion of integral that satisfies this. The issue if that f is piecewise affine, there may be many ways of choosing the t_i; for instance, if one such partition works, one could always refine it by adding more division points. One then has to show that the result is independent of the particular choice of division points (subject to f being affine on each subinterval). This is quite easy though (good exercise); if one has two such partitions, one can take their common refinement (consisting of all the division points in either), and then show that the integrals defined using the original two partitions are equal to that defined using their common refinement, and thus to each other.

Once one knows that the integral is well-defined on the set $D = D([a,b])$ of piecewise affine functions, one gets the basic properties of the integral very easily for functions $f \in D$:

1. linearity: $\int_a^b (c_1 f_1 + c_2 f_2) = c_1 \int_a^b f_1 + c_2 \int_a^b f_2,\ c_j \in \mathbb{R}, f_j \in D$,
2. positivity: if $f \in D$, $f \geq 0$ (i.e. $f(t) \geq 0$ for all $t \in [a,b]$) then $\int_a^b f \geq 0$,
3. boundedness: if $f \in D$ then $|\int_a^b f| \leq \int_a^b |f| \leq (b-a)\|f\|$, where $\|f\|$ is the norm on continuous functions: $\|f\| = \sup\{|f(t)| :\ t \in [a,b]\}$.
4. additivity: if $a < b < c$ and $f \in D([a,c])$ then $\int_a^b f + \int_b^c f = \int_a^c f$.

Note that all properties but the last refer to a single interval $[a,b]$, which is why we are using D simply there in the notation.

Notice that positivity plus linearity imply that if $f, g \in D$, $f \geq g$ (i.e. $f(t) \geq g(t)$ for all $t \in [a,b]$) then $\int_a^b f \geq \int_a^b g$. Indeed, $\int_a^b f - \int_a^b g = \int_a^b (f-g) \geq 0$ where the first equality is linearity of $\int_a^b$ and the second is positivity using $f - g \geq 0$.

In fact, notice that if f is valued in a normed vector space $(V, \|.\|_V)$ (but still defined on an interval), the definition of the integral, as a V-valued map, for piecewise affine functions still works with the same proof, and one still has linearity, additivity, as well as boundedness provided the latter is interpreted as $\|\int_a^b f\|_V \leq \int_a^b \|f\|_V \leq (b-a)\|f\|$, where now $\|f\| = \sup\{\|f(t)\|_V :\ t \in [a,b]\}$. In fact, all the arguments below go through in this setting, provided V is in addition *complete*, i.e. Cauchy sequences converge.

We also know from the problem set that continuous functions $f \in C([a,b])$ can be approximated by elements of D: for all $\varepsilon > 0$ there exists $g \in D$ such that $\|f - g\| < \varepsilon$. (One says that D is **dense** in $C([a,b])$.) This gives the basic idea for getting a definition of the integral of continuous functions.

Suppose $f \in C([a,b])$, and let $f_n \in D$ be such that $\lim f_n = f$, with the limit taken in the metric space $C([a,b])$. Explicitly, this says that for all $\varepsilon > 0$ there is N such that $n \geq N$ implies $\|f_n - f\| < \varepsilon$. Notice that by the approximability mentioned in the previous paragraph, such sequences *exist*. We would like to say that $\lim \int_a^b f_n$ exists and is independent of the choice of the particular sequence chosen to converge to f; if we do this we can *define* $\int_a^b f = \lim \int_a^b f_n$.

Now, as in a metric space convergent sequences are Cauchy, if $f_n \in D$ is such that $\lim f_n = f$, then $\{f_n\}_{n=1}^\infty$ is Cauchy. Thus, for all $\varepsilon > 0$ there is N such that $n, m \geq N$ imply $\|f_n - f_m\| < \varepsilon$. This implies that $\int_a^b f_n$ is a Cauchy sequence in $\mathbb{R}$:

$$\Big|\int_a^b f_n - \int_a^b f_m\Big| = \Big|\int_a^b (f_n - f_m)\Big| \leq (b-a)\|f_n - f_m\|$$

where the first equality is the linearity of the integral on D and the second inequality is the boundedness. But this implies that $\{\int_a^b f_n\}_{n=1}^\infty$ is Cauchy: given $\varepsilon > 0$ choose N such that $n, m \geq N$ implies $\|f_n - f_m\| < \frac{\varepsilon}{b-a}$; then $n, m \geq N$ implies $|\int_a^b f_n - \int_a^b f_m| < \varepsilon$. Since $\mathbb{R}$ is complete, $\lim \int_a^b f_n$ exists, which was the first item to be checked.

The second item to check is that if we have two sequences in D converging to f, say $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$, then $\lim \int_a^b f_n = \lim \int_a^b g_n$. But this is easy. For under this assumption the alternating sequence $f_1, g_1, f_2, g_2, f_3, \ldots$, i.e. the sequence $\{h_n\}_{n=1}^\infty$ with $h_{2k-1} = f_k$, $h_{2k} = g_k$, also converges to f as follows easily from the definition, thus by what we have shown $\lim \int_a^b h_n$ exists, but then all subsequences of $\{h_n\}_{n=1}^\infty$ converge to this very same limit, so $\lim \int_a^b f_n = \lim \int_a^b g_n$.

Putting these together, we can *define* $\int_a^b f$ for $f \in C([a,b])$: take any sequence $\{f_n\}_{n=1}^\infty$ in D with $\lim_{n\to\infty} f_n = f$ (with convergence in $C([a,b])$, and then

$$\int_a^b f = \lim_{n\to\infty} \int_a^b f_n.$$

Note that this argument remains valid if the the target space $\mathbb{R}$ of f is replaced by a *complete* normed vector space V, such as $\mathbb{R}^m$. Complete normed vector spaces are called **Banach spaces**.

Then it is straightforward to check that the integral inherits the properties from D:

1. linearity: $\int_a^b (c_1 f_1 + c_2 f_2) = c_1 \int_a^b f_1 + c_2 \int_a^b f_2,\ c_j \in \mathbb{R}, f_j \in C([a,b])$,
2. positivity: if $f \in C([a,b])$, $f \geq 0$ (i.e. $f(t) \geq 0$ for all $t \in [a,b]$) then $\int_a^b f \geq 0$,
3. boundedness: if $f \in C([a,b])$ then $|\int_a^b f| \leq \int_a^b |f| \leq (b-a)\|f\|$, where $\|f\|$ is the norm on continuous functions: $\|f\| = \sup\{|f(t)| : t \in [a,b]\}$.
4. additivity: if $a < b < c$ and $f \in C([a,c])$ then $\int_a^b f + \int_b^c f = \int_a^c f$.

For instance, linearity follows from the linearity of the limit and of the integral on D: if $\lim_{n\to\infty} f_n^{(j)} = f_j$, $j = 1, 2$ with $f_n^{(j)} \in D$, then

$$\begin{aligned} c_1 \int_a^b f_1 + c_2 \int_a^b f_2 &= c_1 \lim \int_a^b f_n^{(1)} + c_2 \lim \int_a^b f_n^{(2)} \\ &= \lim (c_1 \int_a^b f_n^{(1)} + c_2 \int_a^b f_n^{(2)}) \\ &= \lim \int_a^b (c_1 f_n^{(1)} + c_2 f_n^{(2)}) = \int (c_1 f_1 + c_2 f_2), \end{aligned}$$

where the first equality is the definition of the integral, the second is the linearity of the limit in $\mathbb{R}$, the third the linearity of the integral on D, and the fourth is the linearity of the limit in the vector space $C([a,b])$, i.e. that $\{c_1 f_n^{(1)} + c_2 f_n^{(2)}\}_{n=1}^\infty$ converges to $c_1 f_1 + c_2 f_2$ in $C([a,b])$. (Note that in *any* normed vector space the limit is linear, with the same proof as in the case of $\mathbb{R}$.)

The proof of additivity is completely similar.

To see the boundedness, we first want to note that if $f_n \to f$ with $f_n \in D$, then $|f_n| \to |f|$. To see this, we first need a lemma, which is useful even in normed vector spaces.

Lemma 1 *If V is a normed vector space with norm $\|.\|_V$ then*

$$|\|v\|_V - \|w\|_V| \leq \|v - w\|_V$$

Remark 1 *The lemma implies that $\|.\|_V$ is a uniformly continuous function $V \to \mathbb{R}$. Indeed, given $\varepsilon > 0$ let $\delta = \varepsilon$; then $\|v - w\|_V < \delta$ implies $|\|v\|_V - \|w\|_V| < \varepsilon$.*

Proof of the lemma: We have

$$\|v\|_V \leq \|v - w\|_V + \|w\|_V, \ \ \|w\|_V \leq \|w - v\|_V + \|v\|_V,$$

and $\|v - w\|_V = \|w - v\|_V$, so

$$-\|v - w\|_V \leq \|v\|_V - \|w\|_V \leq \|v - w\|_V,$$

which is exactly the statement of the lemma. $\square$

Returning to our setting, since $|.| : \mathbb{R} \to \mathbb{R}$ is continuous (this being a special case of Remark 1) and the composite of continuous functions is continous, $|f_n|$ is a continuous function. Further,

$$||f_n(t)| - |f(t)|| \leq |f_n(t) - f(t)|, \tag{1}$$

and so

$$\||f_n| - |f|\| = \sup\{||f_n(t)| - |f(t)|| : t \in [a,b]\} \leq \sup\{|f_n(t) - f(t)| : t \in [a,b]\} = \|f_n - f\|,$$

where the inequality between the sups holds since (1) shows that for every element of the first set there is a $\geq$ element of the second, hence the sups have the same inequality (every bound for the second set is a bound for the first, so in particular the sup of the second set is, which is thus greater than or equal to the least upper bound for the first). This shows that $|f_n| \to |f|$ in $C([a,b])$. But now the first part of boundedness is easy: by boundedness in D, $|\int_a^b f_n| \leq \int_a^b |f_n|$; be the definition of the integral (plus the continuity of $|.| : \mathbb{R} \to \mathbb{R}$) the first converges to $|\int_a^b f|$, the second to $\int_a^b |f|$, so the result follows from the fact that if one has two convergent sequences $\{a_n\}_{n=1}^\infty$, $\{b_n\}_{n=1}^\infty$ with $a_n \leq b_n$ for all n (or indeed simply for all sufficiently large n) then $\lim a_n \leq \lim b_n$.

To see the second inequality in the boundedness, let $\varepsilon > 0$. Then there exists N such that $n \geq N$ implies $\sup |f_n| \leq \sup |f| + \varepsilon$, and so the second inequality for boundedness on D shows that

$$\int_a^b |f_n| \leq (b-a) \sup |f_n| \leq (b-a)(\sup |f| + \varepsilon)$$

for $n \geq N$. But the left hand side converges to $\int_a^b |f|$, while the right hand side is a constant (thus converges to itself), so as above one concludes that

$$\int_a^b |f| \leq (b-a)(\sup |f| + \varepsilon). \tag{2}$$

As $\varepsilon > 0$ was arbitrary, this shows that

$$\int_a^b |f| \leq (b-a) \sup |f|.$$

(The detailed argument would be: if $\int_a^b |f| > (b-a) \sup |f|$, take $\varepsilon = \frac{1}{2(b-a)} (\int_a^b |f| - (b-a) \sup |f|) > 0$ for a contradiction with (2): they yield $\varepsilon \leq \varepsilon/2$.)

Finally positivity follows the same way as the bound we just observed: if $f \geq 0$ and $f_n \to f$, then for $\varepsilon > 0$ there exists N such that for $n \geq N$, $f_n > -\varepsilon$, and thus $\int_a^b f_n \geq -\varepsilon \int_a^b 1 = -(b-a)\varepsilon$ (note that $1 \in D$, and $\int_a^b 1 = b-a$ from the definition), so $\int_a^b f \geq -(b-a)\varepsilon$; since $\varepsilon > 0$ was arbitrary, $\int_a^b f \geq 0$.

In summary, we have shown the (much harder) existence part of the following theorem:

Theorem 1 *For $a < b$, $a, b \in \mathbb{R}$, there exists a unique linear map $I_a^b : C([a,b]) \to \mathbb{R}$ such that I_a^b extends $\int_a^b$ from D (i.e. $I_a^b f = \int_a^b f$, $f \in D$) and such that there is a constant C such that*

$$|I_a^b f| \leq C\|f\|, \ f \in C([a,b]). \tag{3}$$

One writes $\int_a^b f = I_a^b f$ even for $f \in C([a,b])$.

Further, this unqiue linear map satisfies

1. *linearity: $I_a^b(c_1 f_1 + c_2 f_2) = c_1 I_a^b f_1 + c_2 I_a^b f_2$, $c_j \in \mathbb{R}, f_j \in C([a,b])$,*
2. *positivity: if $f \in C([a,b])$, $f \geq 0$ (i.e. $f(t) \geq 0$ for all $t \in [a,b]$) then $I_a^b f \geq 0$,*
3. *boundedness: if $f \in C([a,b])$ then $|I_a^b f| \leq I_a^b |f| \leq (b-a)\|f\|$, where $\|f\|$ is the norm on continuous functions: $\|f\| = \sup\{|f(t)| : t \in [a,b]\}$.*
4. *additivity: if $a < b < c$ and $f \in C([a,c])$ then $I_a^b f + I_b^c f = I_a^c f$.*

With the exception of positivity, the same holds for continuous functions with values in a Banach space V, where boundedness reads as

$$\|I_a^b f\|_V \leq I_a^b \|f\|_V \leq (b-a)\|f\|,$$

with the first integral being that of V-valued functions, the second that of $\mathbb{R}$-valued functions.

As already mentioned, we have proved existence. Uniqueness is very easy: if I_a^b is a linear map with (3), and $f_n \in D$ with $\lim f_n = f$, then $|I_a^b f_n - I_a^b f| = |I_a^b(f_n - f)| \leq C\|f_n - f\|$, so $\lim I_a^b f_n = I_a^b f$ (because given $\varepsilon > 0$ choose N such that $n \geq N$ implies $\|f_n - f\| < \frac{\varepsilon}{C+1}$; then $n \geq N$ implies $|I_a^b f_n - I_a^b f| \leq C\|f_n - f\| < \varepsilon$). If I_a^b in addition extends $\int_a^b$ from D (i.e. agrees with the latter on D), then we get $I_a^b f = \lim \int_a^b f_n$, and thus I_a^b is determined by $\int_a^b$ on D. *This completes the proof of the theorem.*

You should think of this theorem as a statement that *all properties of the integral are already encoded in its properties on piecewise affine functions.*

This is already a nice achievement, but let us now prove the fundamental theorem of calculus. For this purpose it is convenient to define $\int_a^a f = 0$, with which definition all properties, including the additivity, still hold.

Theorem 2 *(Fundamental theorem of calculus, part I) Suppose $f : [a,b] \to \mathbb{R}$ is continuous, and let $F(t) = \int_a^t f$, $t \in [a,b]$. Then F is C^1, and $F' = f$.*

Note that the proof given below goes through equally well when f is valued in a Banach space. Notice also that this says that the integral gives an antiderivative.

Proof: If we show F is differentiable with $F' = f$, continuity follows from that of f. So it suffices to show that F is differentiable with $F' = f$.

In order to do so, let $t \in [a, b)$ and consider $h > 0$; the case of $t \in (a, b]$ and $h < 0$ is similar. Then (for $h < b - t$)

$$F(t+h) - F(h) = \int_a^{t+h} f - \int_a^t f = \int_t^{t+h} f$$

by the additivity, and

$$\int_t^{t+h} f = \int_t^{t+h} f(t) + \int_t^{t+h} (f - f(t)) = hf(t) + \int_t^{t+h} (f - f(t)),$$

where the first equality is the linearity of the integral and the second is the integral of a constant function directly from the definition of D. Thus,

$$|F(t+h) - F(t) - hf(t)| = \left| \int_t^{t+h} (f - f(t)) \right| \leq h \sup\{|f(s) - f(t)| : \ s \in [t, t+h]\},$$

where the inequality is the boundedness statement for the integral on $[t, t+h]$. By the continuity of f at t, given $\varepsilon > 0$ there exists $\delta > 0$ such that $|s - t| < \delta$ implies $|f(s) - f(t)| < \varepsilon$; thus for $0 < h < \min(\delta, b - t)$ we have

$$|F(t+h) - F(t) - hf(t)| \leq \varepsilon h.$$

Together with the analogous argument for $h < 0$, this is exactly the definition of differentiability at t with derivative f, proving the theorem. $\square$

In order to do the second half of the fundamental theorem of calculus, we just need an observation:

Proposition 1 *If $f \in C([a, b])$ is differentiable on (a, b) and $f'(t) = 0$ for all $t \in (a, b)$ then f is constant, i.e. $f(t) = f(a)$ for all $t \in [a, b]$.*

Proof: Let $t_1 < t_2$, $t_1, t_2 \in [a, b]$. By the mean value theorem on $[t_1, t_2]$, there is $c \in (t_1, t_2)$ such that $f'(c) = \frac{f(t_2) - f(t_1)}{t_2 - t_1}$; by the assumption $f'(c) = 0$, so $f(t_1) = f(t_2)$. Thus, for all $t \in [a, b]$, $f(t) = f(a)$, so f is a constant function. $\square$

Theorem 3 *(Fundamental theorem of calculus, part II) If $f \in C^1([a, b])$ then $f(b) - f(a) = \int_a^b f'$.*

This is the theorem that is actually used to evaluate integrals explicitly: given a function $g \in C([a, b])$ find $f \in C^1([a, b])$ such that $f' = g$, and then $f(b) - f(a) = \int_a^b g$. Note that part I of the fundamental theorem says that the indefinite integral of g is such a function, but that does not give an explicit evaluation, i.e. you need a different method of finding the antiderivative for the explicit calculation. For instance, if $g(t) = t^n$, then you check that $f(t) = \frac{1}{n+1} t^{n+1}$ satisfies $f'(t) = g(t)$ using the product rule for differentiation, and then apply the second part of the fundamental theorem of calculus to find $\int_a^b g$.

Proof: Let $F(t) = \int_a^t f'$. By the first part of the fundamental theorem of calculus proven above, F is continuously differentiable with $F' = f'$. Thus, the function $g = f - F$ satisfies $g \in C^1([a, b])$ and $g' = 0$. By the previous proposition, $g(b) = g(a)$, i.e. $f(b) - F(b) = f(a) - F(a)$, i.e.

$$f(b) - f(a) = F(b) - F(a) = \int_a^b f',$$

completing the proof. $\square$

Note that this gives also part II of the fundamental theorem of calculus for $\mathbb{R}^n$-valued (or more generally, with values in a finite dimensional normed vector space) functions since it suffices to show that the individual components of f satisfy the theorem, but that is exactly the the above statement. It *does not* give part II of the fundamental theorem of calculus for functions with values in a Banach space because we cannot argue on a component-by-component basis. However, more advanced techniques in fact give the conclusion even in that case. In fact, in the case of f taking values in a complete inner product space (Hilbert space) V is simply to notice that for any $v \in V$,

$$v \cdot \int_a^b f = \int_a^b v \cdot f,$$

again because it is easy to see for piecewise affine functions and the equality is preserved under limits. The fundamental theorem of calculus is applicable to the real valued function $v \cdot f$, so

$$v \cdot \int_a^b f' = \int_a^b v \cdot f' = \int_a^b (v \cdot f)' = v \cdot f(b) - v \cdot f(a) = v \cdot (f(b) - f(a)).$$

Thus,

$$v \cdot \left(\int_a^b f' - (f(b) - f(a)) \right) = 0$$

for all $v \in V$; taking $v = \int_a^b f' - (f(b) - f(a))$ shows $\int_a^b f' - (f(b) - f(a) = 0$, and thus part II of the fundamental theorem of calculus works also for functions with values in Hilbert spaces.

Finally, integration by parts is a simple consequence of the fundamental theorem of calculus and the product rule for differentiation: $(fg)' = f'g + fg'$; indeed, for $f, g \in C^1([a, b])$,

$$f(b)g(b) - f(a)g(a) = (fg)(b) - (fg)(a) = \int_a^b (fg)' = \int_a^b (f'g + fg') = \int_a^b f'g + \int_a^b fg',$$

and now a simple rearrangement gives the formula:

$$\int_a^b f'g = fg|_a^b - \int_a^b fg', \qquad fg|_a^b = f(b)g(b) - f(a)g(a).$$