Introduction We are given a continuous linear functional $L : C[0, 1] \to \mathbb{R}$. We seek a function α of bounded variation on $[0, 1]$ such that $Lf = \int_0^1 f(x)\, d\alpha(x)$ for all $f \in C[0, 1]$. We are *not* in the extended-reals context!

It is true for all pairs V, W of normed vector spaces that a *linear* mapping $T : V \to W$ is continuous at one point if and only if T is continuous at all points of V. Proof of this is left to you! In particular, if T is continuous then T is continuous at 0. Then, given $\epsilon > 0$ there exists $\delta > 0$ such that $\|v\|_V < \delta \Rightarrow \|Tv\|_W < \epsilon$. Now suppose that $0 \neq v \in V$ is arbitrary. Then (important device!) $\|(\delta v/2\|v\|_V)\|_V = \delta/2 < \delta$. Hence $\left\| T\left(\frac{\delta v}{2\|v\|_V}\right)\right\|_W < \epsilon$, so that

$$\|Tv\|_W \leq \frac{2\epsilon}{\delta}\|v\|_V \quad \text{for all} \quad v \in V, \quad \text{since the inequality holds for } 0 \in V \text{ as well.}$$

Thus whenever a linear $T : V \to W$ is continuous there exists $C \geq 0$ such that $\|Tv\|_W \leq C\|v\|_V$ for all $v \in V$. We denote by $\|T\|_{V,W}$ the infimum of the set of all such C. Because it is the *infimum* of all such C, $\|Tv\|_W \leq \|T\|_{V,W}\|v\|_V$ for all $v \in V$, and we call it the norm of T.
This defines a norm on the vector space of all continuous linear mappings $T : V \to W$ (verify!).

Exercise: *Show that* $\|T\|_{V,W} = \sup_{\|v\|_V \leq 1} \|Tv\|_W$ *and that if* $\|Tv\|_W \leq C\|v\|_V$ *for some* C *and for all* $v \in V$, *then* T *is continuous.*

For us, $V = C[0, 1]$ and $\|f\|_V = \max\{|f(x)| : x \in [0, 1]\}$; $W = \mathbb{R}$ and $\|x\|_W = |x|$. But we will simply write $\|f\|$ and $\|L\|$ for the norms of our functions f and linear functionals L. "Linear functional" is the name used for a linear mapping from a vector space to its scalars (scalar field).

Examples: Let $L_0 f := f(0)$. Then $|L_0 f| \leq \|f\|$, so L_0 (also known as the "Dirac delta function") is a continuous linear functional. We will use $L_{1/2} f := f(1/2)$ and $L_1 f := f(1)$ later. $L_{1/2}$ and L_1 are also Dirac deltas, at different base points. Another example: $Lf := \int_0^1 f(x)\, d\alpha(x)$, where α is a function of bounded variation on $[0, 1]$.

This, our preliminary version of the Riesz Representation Theorem for the continuous linear functionals on $C[0, 1]$, asserts that *every* continuous linear functional on $C[0, 1]$ is given by $\int_0^1 f(x)\, d\alpha(x)$ for some function α of bounded variation on $[0, 1]$.

Let us put $\beta(x) := [0 < x \leq 1]$, where the notation $[P(x)]$, where $P(x)$ is a mathematical statement with variable x, denotes the number 1 if $P(x)$ is TRUE and denotes 0 if if $P(x)$ is FALSE. This notation, due to Iverson and Knuth, American Math. Monthly 99(1992) 403 - 426, is yet another way of writing a characteristic function. Thus $\beta(x) = 1$ if $x > 0$ and $\beta(0) = 0$. Then β is a function of bounded variation on $[0, 1]$. Given a partition π of $[0, 1]$ and an associated selection vector ξ, $RS(f, \pi, \xi, \beta) = f(\xi_1) \to f(0)$ as the mesh size of π tends to zero. That is, $L_0 f = \int_0^1 f(x)\, d\beta(x)$. Left to you: find functions like β that represent $L_{1/2}$ and L_1 in this manner.

A Special Case Suppose that whenever $f \geq 0$ (here $f \in C[0, 1]$, an implicit assumption, one of many that will be made without mention) we have $Lf \geq 0$. In particular, if $f \geq g \geq 0$ then $Lf \geq Lg$.

We now define $\varphi_n(x, t) := \min\{1, \max\{0, -n(t-(x+\frac{1}{n}))\}\}$ for $x \in [0, 1]$ and $t \in \mathbb{R}$. Thus $\varphi_n(x, t)$ is continuous, as a function of (x, t), and it can be checked that

$$\varphi_n(x, t) := \begin{cases} 1, & \text{if } t \leq x; \\ 0, & \text{if } t \geq x + \frac{1}{n}; \\ 1 - n(t - x), & \text{if } x < t < x + \frac{1}{n}. \end{cases}$$

We next set $\varphi_n(x) \in C[0, 1]$ equal to $\varphi_n(x, t)$ restricted to the interval $[0, 1]$, so that $\varphi_n(x)(t) := \varphi_n(x, t)$ when $t \in [0, 1]$. The functions $\varphi_n(x)$ are continuous "approximations" of the characteristic function of the interval $[0, x]$.

Next, we define our first, approximate, attempt at constructing the desired function α:

$$a_n(x) := L(\varphi_n(x)) \quad \text{(here, } L \text{ acts on } \varphi_n(x) \text{ as a function of } t \in [0, 1]\text{).}$$

We remark in passing that if $L = L_0$, then $a_n(x) \equiv 1$. What is $a_n(x)$ when $L = L_{1/2}$ or L_1?

If $n \le m$ then $\varphi_n(x) \ge \varphi_m(x)$ for all $t \in [0,1]$. Thus $L(1) \ge a_n(x) \ge a_m(x) \ge 0$ for all $x \in [0,1]$, so we define

$$\alpha_0(x) := \lim_{n\to\infty} a_n(x), \quad \text{for } x \in [0,1].$$

If $0 \le x \le y \le 1$ then $\varphi_n(x) \le \varphi_n(y)$ for all $t \in [0,1]$. Thus α_0 is increasing (we let only one n at a time become infinite):

$$\alpha_0(x) \le a_n(x) \le a_n(y), \quad \text{whence } \alpha_0(x) \le a_n(y), \quad \text{so that } 0 \le \alpha_0(x) \le \alpha_0(y) \le L(1).$$

We now want to show that for each $f \in C[0,1]$, $Lf = \int_0^1 f(x)\,d\alpha_0(x)$. This involves several steps. To begin, we step back from thinking about L and look at a useful method of approximating our functions f.

Since f is uniformly continuous we can find, given $\eta > 0$, a number $\delta > 0$ such that $|x-y| < \delta \Rightarrow |f(x)-f(y)| < \eta/2$. We choose a partition π of $[0,1]$ whose mesh size is less than $\delta/2$. We pick some n so large that $1/n < \delta/2$, *and* $1/n < \min|x_i - x_{i-1}|$, $1 \le i \le n$, as well. Now $\varphi_n(x_1) + \sum_{k=2}^{N_\pi}(\varphi_n(x_k) - \varphi_n(x_{k-1})) \equiv 1$, and each summand is non-negative and at most 1, so if we define $f_{n,\pi} := f(\xi_1)\varphi_n(x_1) + \sum_{k=2}^{N_\pi} f(\xi_k)(\varphi_n(x_k) - \varphi_n(x_{k-1}))$ we have

$$f(t) - f_{n,\pi}(t) = (f(t) - f(\xi_1))\varphi_n(x_1)(t) + \sum_{k=2}^{N_\pi}(f(t) - f(\xi_k))(\varphi_n(x_k) - \varphi_n(x_{k-1}))(t). \tag{1}$$

The support of $\varphi_n(x_k) - \varphi_n(x_{k-1})$ is $[x_{k-1}, x_k + \frac{1}{n}] \subseteq [x_{k-1}, x_{k+1}]$. The support of $\varphi_n(x_1)$ is $[x_0, x_1 + \frac{1}{n}] \subseteq [x_0, x_2]$. For a first i we have $x_{i-1} \le t \le x_i$. Then $(f(t) - f(\xi_k))(\varphi_n(x_k) - \varphi_n(x_{k-1}))(t) = 0$ unless $k = i-1$ or $k = i$. But then $|f(t) - f(\xi_k)| < \eta/2$. Hence $|f(t) - f_{n,\pi}(t)| < \eta$, so $\|f - f_{n,\pi}\| < \eta$.

We have shown that, if the mesh size of π is less than $\delta/2$, then for all n sufficiently large, depending on δ and on π, $\|f - f_{n,\pi}\| < \eta$.

Now we return to L. We have $|Lf - Lf_{n,\pi}| < \eta\|L\|$. We can write

$$\begin{aligned} Lf_{n,\pi} &= f(\xi_1)L(\varphi_n(x_1)) + \sum_{k=2}^{N_\pi} f(\xi_k)L(\varphi_n(x_k) - \varphi_n(x_{k-1})) \\ &= f(\xi_1)a_n(x_1) + \sum_{k=2}^{N_\pi} f(\xi_k)(a_n(x_k) - a_n(x_{k-1})). \end{aligned}$$

We will subtract and add to bring in α_0, remembering that $x_0 = 0$:

$$\begin{aligned} Lf_{n,\pi} &= f(\xi_1)a_n(x_1) + \sum_{k=2}^{N_\pi} f(\xi_k)(a_n(x_k) - a_n(x_{k-1})) \\ &= f(\xi_1)\alpha_0(x_1) + \sum_{k=2}^{N_\pi} f(\xi_k)(\alpha_0(x_k) - \alpha_0(x_{k-1})) \\ &+ f(\xi_1)(a_n(x_1) - \alpha_0(x_1)) + \sum_{k=2}^{N_\pi} f(\xi_k)((a_n(x_k) - \alpha_0(x_k)) - (a_n(x_{k-1}) - \alpha_0(x_{k-1}))) \\ &= f(0)\alpha_0(0) + f(\xi_1)(\alpha_0(x_1) - \alpha_0(x_0)) + \sum_{k=2}^{N_\pi} f(\xi_k)(\alpha_0(x_k) - \alpha_0(x_{k-1})) \\ &+ (f(\xi_1) - f(0))\alpha_0(0) \\ &+ f(\xi_1)(a_n(x_1) - \alpha_0(x_1)) + \sum_{k=2}^{N_\pi} f(\xi_k)((a_n(x_k) - \alpha_0(x_k)) - (a_n(x_{k-1}) - \alpha_0(x_{k-1}))) \\ &=: f(0)\alpha_0(0) + \sum_{k=1}^{N_\pi} f(\xi_k)(\alpha_0(x_k) - \alpha_0(x_{k-1})) + r(f, \pi, \xi, n). \end{aligned} \tag{2}$$

We next estimate the remainder, $r(f,\, \pi,\, \xi,\, n)$:

$$|r(f,\, \pi,\, \xi,\, n)| \le (\eta/2)\alpha_0(0) + \sum_{k=1}^{N_\pi} \|f\|(|a_n(x_k) - \alpha_0(x_k)| + |a_n(x_{k-1}) - \alpha_0(x_{k-1})|).$$

By now allowing n to increase, we can ensure that at each of the points x_k, $|a_n(x_1) - \alpha_0(x_1)| < \eta/(1 + 2 \cdot N_\pi \|f\|)$, so that $|r(f,\, \pi,\, \xi,\, n)| < \eta(\alpha_0(0) + 1)$. Here, the size of n depends on π and the sequence $\{a_n\}$ as well as $\delta/2$. From (2) we see that

$$Lf_{n,\pi} = f(0)\alpha_0(0) + RS(f,\, \pi,\, \xi,\, \alpha_0) + r(f,\, \pi,\, \xi,\, n), \quad \text{so that}$$

$$Lf - f(0)\alpha_0(0) - RS(f,\, \pi,\, \xi,\, \alpha_0) = Lf - Lf_{n,\pi} + r(f,\, \pi,\, \xi,\, n). \tag{3}$$

The left-hand side of (3) is independent of n, and we have seen that by taking n sufficiently large, the right-hand side of (3) is majorized by $\eta\|L\| + \eta(\alpha_0(0) + 1)$. Since f is continuous, we can, by letting the mesh size of π be small enough to make all $\left|RS(f,\, \pi,\, \xi,\, \alpha_0) - \int_0^1 f(x)\, d\alpha_0(x)\right| < \eta$, be assured that

$$\left|Lf - f(0)\alpha_0(0) - \int_0^1 f(x)\, d\alpha_0(x)\right| < \eta(2 + \|L\| + \alpha_0(0)).$$

Hence $Lf = \int_0^1 f(x)\, d\alpha(x)$, where $\alpha(x) := \alpha_0(0)\beta(x) + \alpha_0(x)$, and $\beta(x) = [0 < x \le 1]$.

Application of the Special Case to the general case

We will construct an auxilary functional that maps non-negative functions to non-negative numbers, show that the functional can be extended to a linear functional with the same norm as L, and use the linear extension to show that if we follow the argument we used in the Special Case as much as we can and use Helly's Selection Theorem, we can complete the proof of existence of a suitable function of bounded variation.

We put

$$0 \le \Lambda^+(f) := \sup\{|Lh| : h \in C[0,\, 1],\ |h| \le |f|\} \le \|L\|\|f\|,$$

since $0 \le |Lh| \le \|L\|\|h\| \le \|L\|\|f\|$. By the definition, $|Lf| \le \Lambda^+(|f|)$ for all f.

If $f \ge 0$ and $g \ge 0$ we now show that $\Lambda^+(cf) = c\Lambda^+(f)$ if $c > 0$ and that $\Lambda^+(f + g) = \Lambda^+(f) + \Lambda^+(g)$ and which will show that Λ^+ is "as linear as it can be."

If $c > 0$ it is routine to check that $\Lambda^+(cf) = c\Lambda^+(f)$ (do it!).

Given $\epsilon > 0$ we find h and k with $|h| \le f$ and $|k| \le g$ such that $\Lambda^+(f) < |L(h)| + \epsilon$ and $\Lambda^+(g) < |L(k)| + \epsilon$. Then

$$\Lambda^+(f) + \Lambda^+(g) < |L(h)| + |L(k)| + 2\epsilon = |L(h) \pm L(k)| + 2\epsilon = |L(h \pm k)| + 2\epsilon, \quad \text{and} \quad |h \pm k| \le f + g.$$

Thus $\Lambda^+(f) + \Lambda^+(g) < \Lambda^+(f + g) + 2\epsilon$, and since $\epsilon > 0$ is arbitrary, $\Lambda^+(f) + \Lambda^+(g) \le \Lambda^+(f + g)$.

Next we suppose that $|\ell| \le f + g$. For $\eta > 0$ we put $h_\eta := \frac{\ell f}{f+g+\eta}$ and $k_\eta := \frac{\ell g}{f+g+\eta}$. Then $|h_\eta| \le f$ and $|k_\eta| \le g$. We also have

$$|h_\eta - h_\zeta| = |\ell| f \left|\frac{\zeta - \eta}{(f + g + \eta)(f + g + \zeta)}\right| \le |\zeta - \eta|, \quad \text{so we can let} \quad \eta \to 0.$$

By uniform convergence the function h that is zero when $f + g$ is zero and is equal to $\frac{\ell f}{f+g}$ when $f + g$ is positive is continuous, and k, defined similarly, is continuous as well. Moreover, $|h| \le f$, $|k| \le g$, and $h + k = \ell$.

Accordingly,

$$|L\ell| = |L(h + k)| \le |Lh| + |Lk| \le \Lambda^+(f) + \Lambda^+(g), \quad \text{so} \quad \Lambda^+(f + g) \le \Lambda^+(f) + \Lambda^+(g) \quad \text{and} \quad \Lambda^+(f + g) = \Lambda^+(f) + \Lambda^+(g)$$

when $f \ge 0$ and $g \ge 0$.

The extension of Λ^+ to a continuous linear functional is next; the proof of linearity of the extension is interesting. Let $\Lambda(f) := \Lambda^+(f^+) - \Lambda^+(f^-)$. Then if $g \ge 0$ we have $\Lambda(g) = \Lambda^+(g) \ge 0$, and for all f

$$-\|L\|\|f\| \le -\|L\|\|f^-\| \le -\Lambda^+(f^-) \le \Lambda(f) \le \Lambda^+(f^+) \le \|L\|\|f^+\| \le \|L\|\|f\|, \quad \text{so} \quad |\Lambda(f)| \le \|L\|\|f\|$$

and by the definition, $|Lf| \le \Lambda(|f|)$.

Since $(-f)^+ = f^-$ and $(-f)^- = f^+$, $\Lambda(-f) = -\Lambda(f)$ and hence $\Lambda(cf) = c\Lambda(f)$ for all f and all real c.

Since $f + g = (f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^-$ and Λ^+ is additive on non-negative functions, $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+ \Rightarrow \Lambda^+((f+g)^+) + \Lambda^+(f^-) + \Lambda^+(g^-) = \Lambda^+((f+g)^-) + \Lambda^+(f^+) + \Lambda^+(g^+)$ and so Λ is linear and continuous.

As in the Special Case we define

$$a_n(x) := L(\varphi_n(x)) \quad \text{(here, as before, } L \text{ acts on } \varphi_n(x) \text{ as a function of } t \in [0,\,1]).$$

In the Special Case we showed that the sequences $\{a_n(x)\}$ were decreasing and bounded below by 0, which gave us convergence to $\alpha_0(x)$. This monotonicity will be sorely missed, but its absence can be bypassed.

It is still true that whenever $0 \le x < y \le 1$, $\varphi_n(y) - \varphi_n(x)$ is a non-negative function of t. Therefore

$$|a_n(y) - a_n(x)| = |L(\varphi_n(y) - \varphi_n(x))| \le \Lambda^+(\varphi_n(y) - \varphi_n(x)) = \Lambda(\varphi_n(y) - \varphi_n(x)) = A_n(y) - A_n(x),$$

where $A_n(x) := \Lambda(\varphi_n(x))$, just as in the Special Case, since Λ fits that case. Hence for any partition π of $[0,\,1]$,

$$\sum_{k=1}^{n_\pi} |a_n(x_k) - a_n(x_{k-1})| \le \sum_{k=1}^{n_\pi} |A_n(x_k) - A_n(x_{k-1})| = \sum_{k=1}^{n_\pi} (A_n(x_k) - A_n(x_{k-1})) = A_n(1) - A_n(0),$$

since $A_n(x)$ increases. But $A_n(1) = \Lambda(1) \le \|L\|\|1\| = \|L\|$ and $A_n(0) = \Lambda(\varphi_n(0)) \ge 0$, so we have

$$\sum_{k=1}^{n_\pi} |a_n(x_k) - a_n(x_{k-1})| \le \|L\| < \infty.$$

Thus each a_n is a function of bounded variation on $[0,\,1]$ and $V[a_n] \le \|L\|$. We can put $a_n(x) =: P_n(x) - N_n(x)$, the Jordan decomposition of a_n, and we have $P_n(1) + N_n(1) \le \|L\|$, so each sequence $\{P_n\}$ and $\{N_n\}$ is a uniformly bounded sequence of increasing functions on $[0,\,1]$. By Helly's Theorem there is a subsequence $\{P_{n_k}\}$ that converges to an increasing function $P(x)$ on $[0,\,1]$. There is then (again by Helly's Theorem) a further subsequence $\{N_{n_{k_\ell}}\}$ that converges to an increasing function $N(x)$ on $[0,\,1]$. Therefore $a_{n_{k_\ell}}(x) \to P(x) - N(x) =: \alpha_0(x)$ on $[0,\,1]$, and α_0 is a function of bounded variation at most $\|L\|$.

We can now use the argument in the Special Case almost word for word. There are a few changes that have to be made. Instead of working with a_n and n we need to work with n_{k_ℓ} and ℓ. Of course, (1) remains the same (except for the change from n to n_{k_ℓ}), and (2) is the same, with the same *caveat*. In the *estimates* that follow (2) we have to replace $\alpha_0(0)$ by $|\alpha_0(0)|$, but we do not make this change in the *formulas* that contain $\alpha_0(0)$.

The presence of $\alpha_0(0)$ is "explained" by looking at L_0, because there the variations of each a_n were zero. In intuitive terms, there is still a jump at zero, even though there are no points to the left of zero in this "universe of discourse."

If we are given a function of bounded variation we can always replace it by one that is continuous from the left, and the difference will have bounded variation and will be zero except at countably many points.

Exercise: *Prove that if α is a function of bounded variation on $[0,\,1]$ that is zero except at countably many points, then for all $f \in C[0,\,1]$, $\int_0^1 f(x)\,d\alpha(x) = 0$.*

By using this exercise we can now assert the existence of a unique left-continuous function of bounded variation on $[0,\,1]$ that "represents" a continuous linear functional on $C[0,\,1]$.

This note uses ideas from Rudin's *Real and Complex Analysis,* pp 130-132, third edition, McGraw-Hill, 1987.