APPROXIMATION THEORY HOMEWORK 1

TOMMENIX YU
ID: 12370130
STAT 31220
DUE WED MAR 29ST, 2023, 11PM

Discussed with classmates.

Exercise 1.

Proof.

(1) $\mathbb{R}^n$ space:

We decompose $x = x_k + x'$ where $x_k \in \Sigma_{k,n}$ such that x_k is supported on the k positions that $|x_i|$ is the largest. Thus, x' is in $\Sigma_{n-k,n}$. Moreover, we know by definition that

$$\sigma_{k,n}(x)_p \le ||x'||_p.$$

Now the heuristic is to note that if we just take off the k largest ones then the "proportion of change to the norm" is largest when the coordinates are same (since no infinite norm for p). So we deal with that kind of vectors first then extend for general vectors.

In addition, Jeremy says that $r > 0$ (via email) so we know by it's definition that $p > q$.

$\Big($If this $r \ge 0$ does not hold then we can find examples when $q \gg p$ and get counter example, e.g.

$$x = (1,1,1), k = 1, q = 100, p = \frac{10}{9}$$

and then

$$\sigma_{k,n}(x)_p = 2^{\frac{9}{10}} > 3^{\frac{1}{100}} = 1^{-r} \cdot ||x||_q$$

so result does not hold.$\Big)$

When $|x_i| = |t|$ for all i:

In this case, we know

$$||x||_p = n^{\frac{1}{p}}|t|; \qquad ||x'||_p = (n-k)^{\frac{1}{p}}|t|; \qquad ||x||_q = n^{\frac{1}{q}}|t|$$

and in this case it's obvious that the result hold since $(n-k)+k = n$ and $r + \dfrac{1}{p} = \dfrac{1}{q}$.

In this case $r \geq 0$ hence we have

$$\frac{||x||_q}{|t|} = (n-k+k)^{r+\frac{1}{p}} = (n-k+k)^r(n-k+k)^{\frac{1}{p}} \geq k^r(n-k)^{\frac{1}{p}} = k^r\frac{||x'||_p}{|t|} \geq \frac{k^r\sigma_{k,n}(x)_p}{|t|} \tag{0.1}$$

where we used the fact that $k, n-k \in \mathbb{N}^*$.

For general $x \in \mathbb{R}^n$:

If $x = 0$ then we are done. So let's assume $x \neq 0$ and define

$$|s| := ||x||_q n^{-\frac{1}{q}}; \qquad |t| := ||x||_p n^{-\frac{1}{p}}$$

that is, the "average" of $|x_i|$ under p or q norms. Then we have

$$\frac{|s|}{|t|} = \frac{||x||_q}{||x||_p} n^{\frac{q}{p}} \geq n^{\frac{q}{p}} \geq 1$$

since $p \geq q \geq 1$. (from linear algebra we get $||x||_p \leq ||x||_q$.)

Thus we have

$$\sigma_{k,n}(x)_p \leq ||x'||_p \leq \left(\frac{n-k}{n}\right)^{\frac{1}{p}} ||x||_p = (n-k)^{\frac{1}{p}}|t| \leq (n-k)^{\frac{1}{p}}|s| \leq k^{-r}n^{\frac{1}{q}}|s| = k^{-r}||x||_q$$

where the second inequality is because we've removed those largest k terms, so their "average" is larger than $|t|$ (or simply take degree p and it's obvious), the last inequality above is due to 0.1.

(2) Extention to l^p:

For an extension to l^p, we note first that since $x \in l^p$ it is in l^q. Moreover, it is in l^p means that the concept of "the largest k terms" is defined. Since the norms are invariant under permutations of x (absolute summable) so we can WLOG assume the first k terms of x has the largest absolute value. In other words

$$x = x_k + x'$$

where $x_k = (x_1, \dots, x_k, 0 \dots)$. Since $x \in l^p$ so we know the tail goes to 0, that is, for every ε there exists N such that $\left(\sum_{i=N+1}^{\infty} |x_i|^p\right)^{\frac{1}{p}} < \varepsilon$. So we denote the first $N = N(\varepsilon)$ terms in x to be $x^\varepsilon \in \mathbb{R}^N$.

Since every $N' > N$ works fine here we choose $N(\varepsilon) \geq k$. So now we have

$$\sigma_{k,\infty}(x)_p \leq ||x'||_p \leq \varepsilon + ||(x^\varepsilon)'||_p \leq \varepsilon + k^{-r}||x^\varepsilon||_q \leq \varepsilon + k^{-r}||x||_q$$

where the middle inequality is just using first part. Taking $\varepsilon \to 0$ we get our result.

(3): Extension to $p = 1$:

Since $r > 0$ this is impossible. One way to extend is to have $r = 0$ and in this case $q = 1$ as well. Then there's not much to consider though since the result becomes

$$\sum_{i=1}^{n-k} |x_\sigma(i)| := ||x'||_1 \leq k^0 ||x||_1 = ||x||_1$$

which is obvious since the left hand side has more positive terms in the summation.

We have seen why we keep $r \geq 0$, due to the counterexample above.

<u>Extension to $p = \infty$:</u>

We again use the notation

$$x = x_k + x'$$

in the same way as above.

This is doable since the left hand side becomes

$$\sigma_{k,n}(x)_\infty = ||x'||_\infty := |x_s|$$

by definition of infinite norm (we didn't need to show this above for $p < \infty$, we don't need strict equality here also but it's obvious in this case).

But now we know $r = \dfrac{1}{q}$ and

$$k^{-r}||x||_q \geq \frac{1}{k^{\frac{1}{q}}} \left(\sum_{i=1}^{k} |x_k|^q \right)^{\frac{1}{q}} = \left(\frac{1}{k} \sum_{i=1}^{k} |x_k|^q \right)^{\frac{1}{q}} \geq \left(|x_s|^q \right)^{\frac{1}{q}} = |x_s| = ||x'||_\infty$$

where the last inequality is because each non-zero term of x_k is larger or equal to x_s. Thus we're done.

$\square$

Exercise 2.

Proof.

(a):

Just plugging in we get

$$\int_0^1 e^{-2\pi ikx} f(x)dx = \int_0^1 e^{-2\pi ikx} \sum_{|m|<N} f_m e^{2\pi imx} dx \overset{\text{finite sum}}{=} \sum_{|m|<N} f_m \int_0^1 e^{-2\pi ikx} e^{2\pi imx} dx$$
$$= \sum_{|m|<N} f_m \int_0^1 e^{2\pi i(m-k)x} dx = f_k$$

since if $m - k \in \mathbb{Z}\backslash\{0\}$ then the integral is on the unit circle and it's result is 0.

(b):

$$P_N(f)(x) = \sum_{|k|<N} f_k e_k = \int_0^1 f(y) \sum_{|k|<N} e^{2\pi ik(x-y)} dy = f * D_N(x)$$

where

$$D_N(x) = \sum_{n=-N+1}^{N-1} e^{2\pi inx}$$

and we can compute

$$\sum_{n=-N}^{N} e^{2\pi in\theta} = e^{-2\pi iN\theta} \frac{1 - e^{2\pi i(2N+1)\theta}}{1 - e^{2\pi i\theta}} = \frac{e^{-2\pi iN\theta} - e^{2\pi i(N+1)\theta}}{1 - e^{2\pi i\theta}}$$
$$= \frac{e^{-2\pi i(N+1/2)\theta} - e^{2\pi i(N+1/2)\theta}}{e^{-\pi i\theta} - e^{\pi i\theta}} = \frac{\sin(2\pi(N+1/2)\theta)}{\sin(\pi\theta)} = D_{N+1}(\theta)$$

and thus

$$P_N(f)(x) = f * D_N(x) = \int_0^1 \frac{\sin(2\pi(N-1/2)(x-y))}{\sin(\pi(x-y))} f(y)dy.$$

(c):

For convenience let's shift D_N to D_{N+1} here since when $N \to \infty$ this does not matter. We have

$$||P_N|| \geq \sup_{||f||_\infty = 1} \left\| \int_0^1 D_N(x-y)f(y)dy \right\|_\infty$$

so we can just pick f such that $||f||\infty = 1$ and

$$\left\| \int_0^1 D_N(x-y)f(y)dy \right\|_\infty \geq c \cdot \left\| \int_0^1 D_N(x-y)dy \right\|_\infty$$

and check that the latter goes to ∞. A natural choice is to pick $f = \text{sgn}(D_N)$ then inside is always positive, yet then f is not continuous. But since P_N is bounded (Riemann Lebesgue Lemma) it is continuous in f, so we pick a sequence of continuous $f_n \to \text{sgn}(D_N)$ then we know that for any ε there's some N that for all $n \geq N$ we have (note $D_N(y) = D_N(-y)$, then pick $x = 0$)

$$\left\| \int_0^1 D_N(x-y) f_n(y) dy \right\|_\infty \geq \left| \int_0^1 D_N(0-y) f_n(y) dy \right| \geq c \cdot \int_0^1 \left| D_N(y) \right| dy$$

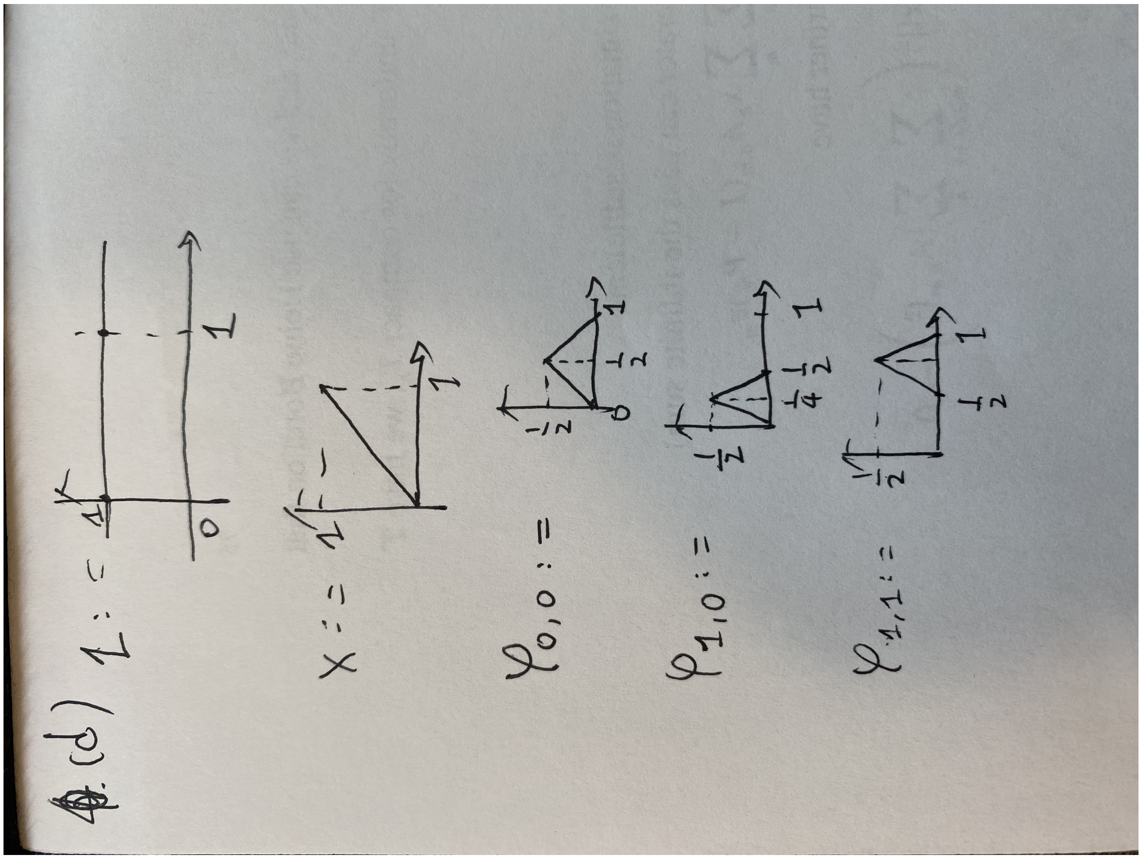
since $P_N(f_n) \to P_N(\mathrm{sgn}(D_N))$.

Thus we only need to show that the integral $\displaystyle\int_0^1 \left| D_N(y) \right| dy$ is divergent.

Since $\sin x \leq x$ we just bound D_N by $|D_N| \geq \frac{|\sin(2\pi(N-1/2)(x))|}{\pi x}$ and hence breaking the integral into small parts we get

$$\begin{aligned}\int_0^1 \left| D_N(y) \right| dy &\geq \frac{1}{\pi} \cdot \sum_{n=0}^{2N-1} \int_0^\pi \frac{|\sin(h+n\pi)|}{h+n\pi} dh \\ &\geq \frac{1}{\pi} \cdot \sum_{n=0}^{2N-1} \frac{1}{h+n\pi} \int_0^\pi |\sin(u)| du = O\left(\sum_{n=1}^\infty \frac{1}{n} \right) = \infty.\end{aligned}$$

Now we show f is not approximated by P_N, which is because $f = \lim_{n\to\infty} P_N f$ so bounded for some f, for the family P_n on $\mathbb{N}$, then by uniform boundedness we get that $||P_N|| < \infty$ is uniform in x. Thus $\sup ||P_N|| < \infty$, which contradicts with above. So we are done.

(d): 

(e):

We know that all the multiples of $1/2^n$ for all n uniquely defines a continuous function if they define one. But for any given continuous function f on $[0, 1]$, they already defines f, so it's sufficient to show that the f at all the multiples of $1/2^n$ for all n are uniquely defined by χ, and that this combination yields a continuous function.

Values at all the multiples of $1/2^n$ for all n are uniquely captured:

We note that all function except 1 is 0 at 0, so we let $\alpha_0 = f(0)$.

All functions other than 1 and x are 0 at 1, so we let $\alpha_x = f(1) - 1 \cdot \alpha_0 = f(1) - f(0)$.

The key from now on is to note that all multiples of $\frac{1}{2^n}$ are 0 for $\phi_{k,l}$ with $k \geq n$, and that for each $\frac{c}{2^n}$ with c even it's already fixed, while with c odd there's exactly one of the $\phi_{n,s}$ wavelets that's supported at that point. So we can find coefficients $\alpha_{n,k}$ for all $k \in \{0, \ldots, 2^n - 1\}$ in such a way that for

$$S_n = \alpha_0 + x \cdot \alpha_x + \sum_{i=1}^{n} \sum_{j=0}^{2^n-1} \alpha_{i,j} \phi_{n,j}$$

we have

$$f\left(\frac{c}{2^n}\right) = S_n\left(\frac{c}{2^n}\right)$$

with an induction we know that the value of all multiples of $\frac{1}{2^n}$ is perfectly suited by S_n. That is, all multiples of $\frac{1}{2^n}$ for all n is uniquely defined by a combination of χ.

So we are done if we know if the limit $\lim S_n$ is continuous.

$\lim S_n$ is continuous:

We show the limit of continuous functions are continuous by showing the convergence is uniform.

So we show $||S_n - f||_\infty < \varepsilon$ for all $n \geq N$, and we have that for all $a \in [0,1]$ there's some δ close $\frac{c}{2^n}$ to it and

$$|S_n(a) - f_n(a)| \leq \left|S_n(a) - S_n\left(\frac{c}{2^n}\right)\right| + \left|S_n\left(\frac{c}{2^n}\right) - f_n\left(\frac{c}{2^n}\right)\right| + \left|f_n\left(\frac{c}{2^n}\right) - f_n(a)\right|$$

where the first is by S_n piece wise continuous, f uniformly continuous, and approximation at $\frac{c}{2^n}$. Thus we are done.

Uniqueness

This is just because in our construct of S_n, each coefficient is uniquely defined, as $\frac{c}{2^n}$ is not supported for any $\phi_{n+1,k}$.

(f):

(This proof (for Lemma 0.2) below is direct from my homework from 381 last quarter, with some notational difference.)

We use the following theorems to show the result (whose proofs are long and hard, which I skip)

Theorem 0.1. *If $\{\mu_n\}$ is a tight family of probability measures on $C[0,1]$ whose limiting distributions converges to finite dimension distributions of some measure μ, then $\mu_n \rightharpoonup \mu$ weakly.*

Lemma 0.2. *Construct a sequence of random variables $\{X_n\}$ on $C[0,1]$ as follows. Let $X_1(0) = 0$, $X_1(1) \sim N(0,1)$, and define $X_1(t)$ for all other t by linear interpolation. In general, having defined X_k, define X_{k+1} as follows. If $t = j2^{-k}$ for some even j, let $X_{k+1}(t) = X_k(t)$. If $t = j2^{-k}$ for some odd j, let $X_{k+1}(t) = X_k(t)$+independent $N(0, 2^{-k-1})$. For all other t, define $X_{k+1}(t)$ by linear interpolation.*

Then, the limit of $X_n \to B$, the standard Brownian motion.

The proof of Lemma 0.2 will come in the end. But we first note that this lemma is equivalent to the fact that our construction of $S_n \to B$ with normal distributions as coefficients:

The reason is that the construction of "step n" in (e) above is exactly constructing X_n if the coefficients are normal distributed with coefficient $\frac{1}{2^{1+n/2}}$. This matches our assumption here in (f) and thus X_n is our S_n above for these sets of coefficients.

Our $\phi_{n,m}$ is exactly the "linear interpolation" in lemma 0.2, so Lemma 0.2 really is equivalent to $S_n \to B$ here.

Computing $\mathbb{E}[(f(x) - f(y))^2]$:

We use the fact that $f = B$ here and get (by definition of Brownian motion)

$$\mathbb{E}[(f(x) - f(y))^2] = E(N(0, |x - y|)^2) = |x - y|$$

since mean is 0, second moment is variance.

Now the rest is just hard toil of proving convergence to Brownian motion.

Proof. (Lemma 0.2)

The limit of finite dimensional distributions converge to that of a Brownian motion:

If the point t is a point that is a multiple of a degree of 2, that is, $t = c \cdot 2^{-s}$ for odd c, then we know that the pointwise sequence $X_n(t)$ is a constant after s terms by definition, so it's corresponding probability is

$$\mathbb{P}(X_n(t) \in A) = \mathbb{P}\left(\sum_{i=1}^{s} a_i N(0, 2^{-i-1}) \in A\right)$$

where a_i is some weight associated to the point's actual position in the interval. Hence, since it's the sum of independent Gaussians, it's still Gaussian.

Now we show that the coefficient a_i is really related to the standard Gaussian. That is, we notice that the a_i are $\frac{d}{2^i}$ where d is the difference between t and its closest $\frac{l}{2^i}$ for l odd. But this coincides exactly with the fact that the variance is additive ($\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y$) and hence we know that

$$\mathbb{P}\left(\sum_{i=1}^{s} a_i N(0, 2^{-i-1}) \in A\right) = \mathbb{P}\left(N(0, t) \in A\right)$$

which means we've checked difference of $X(s)$ and $X(0)$. But this is enough to generate all $X(t)$ and $X(s)$ by additivity of Gaussians.

This concludes for all points that can be finitely expressed. But for any point that is not a grid point (of degrees of 2s), we can list a sequence of odd multiples of 2 that is closest to it, then since the sum converges absolutely we can write out

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(t) \in A\right) = \mathbb{P}\left(\sum_{i=1}^{\infty} a_i N(0, 2^{-i-1}) \in A\right) = \mathbb{P}\left(N(0, t) \in A\right)$$

where the infinite sum is a sum of independent Gaussians that are well defined and converges, thus it's still a Gaussian. Moreover, the same decomposition is contained as in the finite end case since the sum converges.

But then the finite dimensional distribution is

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(t_1) \in A_1, \dots, \lim_{n\to\infty} X_n(t_m) \in A_m\right) = \int_{A_1\times\cdots\times A_n} dN$$

where

$$N = N(0, t_1) \times \cdots \times N(0, t_m)$$

is the multivariate Gaussian with the correct variance, so it's the same as that of a Brownian motion.

<u>Expectation $< \infty$:</u>

The expectation of the sup norm is the maximal difference of differences at each odd multiple of 2^{-k}. In other words, if there are l odd multiples at step n, then the difference is

$$\mathbb{E}\left[||X_{n+1} - X_n||_\infty\right] = \mathbb{E}\left[\max\left\{|Y_1|, \dots, |Y_l|\right\}\right]$$

where Y_i are iid with $Y_1 \sim N(0, 2^{-n})$, so we just have to compute the order distribution of normal distributions and show that they indeed forms a convergent sequence.

To do this, let $Y = \max\{X_1, \dots, X_n\}$, (where just for convenience we use X to stand for Gaussians) then we have by Jensen's inequality that

$$\exp(t\mathbb{E}[Y]) \leq \mathbb{E}[e^{tY}] = \mathbb{E}[\max e^{tX_i}] \leq \sum_{i=1}^{n} \mathbb{E}[e^{tX_i}] = n\exp\left(0 + \frac{\sigma^2 t^2}{2}\right)$$

where the last step is because X_i are iid and by Gaussian's moment generating function.

Thus take log on both sides we get

$$\mathbb{E}[Y] \leq \frac{\log n}{t} + \frac{\sigma^2 t}{2}$$

where to find minimum take $t = \dfrac{\sqrt{2\log n}}{\sigma}$ we get

$$\mathbb{E}[Y] \leq \sigma\sqrt{2\log n}.$$

But note if we were to compute a bound of $Z = \max\{X_1, -X_1, \dots, X_n, -X_n\}$ it suffices us to bound both the positive part and the negative part. But their bounds are the same since we're using the same moment generating function (mean is still 0). Thus this is the bound for what we want.

Now we check that for $\sigma = 2^{-n+1}$ we have

$$\mathbb{E}[Z] \leq \frac{\sqrt{2\log n}}{2^{n-1}} = o\left(\frac{n}{2^n}\right) = o\left(\frac{1}{n^2}\right) < \infty$$

and we are done (where the order sign really cares only when n large).

Conclude the limit is a Brownian motion:

By theorem in class, we only need to check 2 things in order to prove this: tightness of μ_{X_n} and weak convergence of finite dimensional distributions. By the first part of this question we are done with weak convergence of finite dimensional distributions.

But we have

$$\begin{aligned}\mathbb{P}\left(\lim_{n\to\infty}|X_n|\geq N\right)&\leq\mathbb{P}\left(\sum_{n=1}^{\infty}||X_{n+1}-X_n||_{\infty}\geq N\right)\\&=\mathbb{P}\left(\sum_{n=1}^{\infty}\max\left\{|Y_1|,\dots,|Y_l|\right\}\geq N\right)\to 0\end{aligned}$$

as $N\to\infty$ as right hand side is a convergent sequence, as is shown above. Hence they are a.s. Cauchy. Hence uniformly bounded.

Now, we only need to show uniform equicontinuity to show they are a tight family. To do so we note that at each point the local increment is the limit of a random walk with step size $N(0,1)$, since at each step the increment in slope is $\dfrac{N(0,2^{n-1})}{2^{n-1}}=1$. But using the fact that the probability of the limit of a random walk goes to ∞ is 0, we get that with probability 1 the local increment is bounded at each point. So it's equicontinuous.

Moreover, since $X_n\in C[0,1]$ which is a family from a compact set to a metric space, hence equicontinuity implies uniformly equicontinuity. So we are done.

So we know the convergent holds.

□

(Alternatively we can just create sequence to x and y with multiples of $\frac{1}{2^n}$ for all n, then use the coefficients of the limit to get same result. But I have this nice proof so why not.)

□

Exercise 3.

Proof.

For fixed $x \in K$, and we can just define $l : \mathcal{A}(K) \to \mathbb{C}$ such that

$$l_z(f) = f(z)$$

and see that it's linear bounded. It's linear by definition and bounded because it's continuous in f (remember that for operators continuous iff its bounded).

Now, we use Hahn Banach and first we check that

$$||l_z(f)|| = |f(z)| \leq ||f||\Big|_K = \max_{x\in\partial K} |f(y)|$$

and hence

$$||l_z|| \leq 1$$

by definition of operators.

Thus we can extend l_z to a linear functional $l'_z : C(\mathbb{C}) \to \mathbb{C}$ such that it's norm is also 1.

Thus, we have that for all $f \in \mathcal{A}(K)$ (i.e. f on K is analytic)

$$f(z) = l_z(f) = l'_z(f)$$

and by Riesz representation it is the dual product of f and some element in $C(\mathbb{C})^* = \mu_B$ the space of bounded measures. That is, we can find $\mu : \mathcal{B}(\mathbb{C}) \to [0, M]$ such that

$$l'_z(f) = \langle l'_z, f\rangle_{Dual} = \int_{\mathbb{C}} f\, d\mu$$

for all f. Now we fix some $f \in \mathcal{A}(L)$ and since any continuous continuation outside K yields f in $C(\mathbb{C})$, so we pick f that is supported only on $K^\varepsilon := \{x \in \mathbb{C} | d(x, K) \leq \varepsilon\}$.

Thus, we have

$$l'_z(f) = \int_{\mathbb{C}} f\, d\mu = \int_{K^\varepsilon} f\, d\mu = \int_K f\, d\mu + \int_{K^\varepsilon\backslash K} f\, d\mu \to \int_K f\, d\mu$$

as $\varepsilon \to 0$ since there's essentially "no points" in $K^\varepsilon\backslash K$, i.e. every point is not in it as $\varepsilon \to 0$. Thus we continue with only integral on K.

Now since K is compact we can assign some path inside K to each point on it's boundary such that (a) they do not intersect and (b) they left out only a measure 0 set in K under μ. Then since f is fixed we can define a new measure $\mu' = \mu'(f, \mu)$ such that

$$\int_{t\in\partial K} f(x) d\mu' = \int_{h(t):\text{path}} f(x) d\mu$$

and 0 else where.

Thus, we have found a suitable measure supported on ∂K such that

$$f(z) = \int_{\partial K} f(x) d\mu'$$

and the only thing left to show is that is a probability measure.

But if we go back to μ then take $f = 1$ everywhere we get

$$1 = f(z) = \int_{\mathbb{C}} f(x) d\mu = \int_{\mathbb{C}} d\mu$$

but by our construction of μ' it's total measure is the same as μ for each f, thus for each μ' we have

$$\int_{\partial K} 1 d\mu' = \int_{\mathbb{C}} 1 d\mu' = \mu'(\mathbb{C}) = 1$$

hence a probability measure. And we are done.

□

Exercise 4.

Proof.

Assume that

$$\operatorname{sgn}\left[(u_1 - f)(x)\right] \neq \operatorname{sgn}\left[(u_2 - f)(x)\right]$$

then there is a neighborhood of x, $N := N_x$ such that this holds because both $u_1 - f$ and $u_2 - f$ is continuous. But then

$$\int_N |u_1 - f| + |u_2 - f| d\mu > \int_N |u_1 + u_2 - 2f| d\mu$$

where the inequality is strict since the sign flip means cancellation always happen inside N.

But now we have

$$\begin{aligned}
\int_X |u_1 - f| + |u_2 - f| d\mu &= \int_{X\setminus N} |u_1 - f| + |u_2 - f| d\mu + \int_N |u_1 - f| + |u_2 - f| d\mu \\
&\geq \int_{X\setminus N} |u_1 + u_2 - 2f| d\mu + \int_N |u_1 - f| + |u_2 - f| d\mu \\
&> \int_{X\setminus N} |u_1 + u_2 - 2f| d\mu + \int_N |u_1 + u_2 - 2f| d\mu \\
&= 2d\left(\frac{1}{2}u_1 + \frac{1}{2}u_2, f\right)
\end{aligned}$$

yet since M is convex $\frac{1}{2}u_1 + \frac{1}{2}u_2 \in M$ and hence

$$d\left(\frac{1}{2}u_1 + \frac{1}{2}u_2, f\right) < d(u_1, f) = d(u_2, f)$$

contradict to the fact that u_1 and u_2 are best approximations.

□