WHAT IS... THE MÜNTZ-SZÁSZ THEOREM?

ANDREU FERRÉ MORAGUES

Abstract. Weierstrass' Approximation Theorem (1885) implies that $\mathrm{Span}_{\mathbb{C}}\{1, x, x^2, \dots\}$ is dense in $C([0,1],\mathbb{C})$. It is natural to ask if all the powers of x are necessary for this to hold, and even if we can give necessary and sufficient conditions on the exponents, so that their span is dense in $C([0,1],\mathbb{C})$. This question was asked in a slightly more general way by Bernstein in 1912 and answered by Müntz in 1914. We will state and prove a special case of his theorem that uses approximation theory methods in the Hilbert space $L^2([0,1])$.

1. Introduction

Our story begin in 1885 with one of the most important results in approximation theory, due to Karl Weierstrass:

Theorem 1.1 (Weierstrass Approximation Theorem). *Let $f \in C([0,1],\mathbb{C})$ and let $\varepsilon > 0$. Then, there exists a polynomial $p \in \mathbb{C}[x]$ such that*

$$||f - p||_{sup} := \sup\{|f(x) - p(x)| : x \in [0,1]\} < \varepsilon.$$

In other words, $Span_{\mathbb{C}}\{x^n : n \in \mathbb{N} \cup \{0\}\}$ is dense in the Banach space $C([0,1],\mathbb{C})$ with the supremum norm.

As was stated in the abstract, the natural question of how many exponents are needed in order to have density in $C([0,1])$ had been asked by one of the greatest approximation theorists of the early twentieth century: S. Bernstein. In 1912 he conjectured the following: given a family of distinct real numbers $\{0 = \lambda_0 < \lambda_1 < \dots\}$, then $\mathrm{Span}_{\mathbb{C}}\{x^{\lambda_i} : i \in \mathbb{N} \cup \{0\}\}$ is dense in $C([0,1])$ if and only if

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty.$$

Bernstein himself had worked on this problem, and he had proved that the condition

$$\sum_{i=1}^{\infty} \frac{1 + \log \lambda_k}{\lambda_k} = \infty$$

is necessary for density, and also that

$$\lim_{k\to\infty} \frac{\lambda_k}{k \log \lambda_k} = 0$$

is sufficient.

In 1914, his conjecture was proven to hold by Müntz, using a method to compute distances in $L^2([0,1])$ using Gram determinants and a determinantal result due to Cauchy. Next we state a special case of his result, that will completely solve the problem posed in the abstract, but also contain a reduction that allows for a simpler proof showcasing the methods used in the original proof, as well as an extra simple proof of Golitschek. It should also be noted that part of the original proof of Müntz was simplified by Szász in 1916, which we will also make use of.

Theorem 1.2 (Müntz-Szász)**.** *Let $\{0 = \lambda_0 < \lambda_1 < \dots\}$ be a sequence of real numbers satisfying $\lim_{i\to\infty} \lambda_i = \infty$. Then, $Span_{\mathbb{C}}\{x^{\lambda_i} : i \in \mathbb{N} \cup \{0\}\}$ is dense in $C([0,1])$ if and only if*

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty.$$

The proof of this theorem will occupy most of the rest of these notes: we will first introduce auxiliary results needed in the proof, we will then offer the two promised proofs and end with generalizations and other possible proofs, as well as where to find them.

2. Preliminary results

We begin with an elementary result of convergence of infinite products, which will be used repeatedly:

Proposition 2.1. Let $0 < a_n < 1$ be a sequence of real numbers. Then,

$$\prod_{j=1}^{n} (1 - a_j) \xrightarrow[n\to\infty]{} 0 \text{ if and only if } \sum_{j=1}^{\infty} \frac{1}{a_j} = \infty.$$

Proof: By assumption on our sequence $(a_n)_{n\in\mathbb{N}}$, we can take logarithms to see that the LHS holds if and only if

$$\sum_{j=1}^{n} \log(1 - a_j) \xrightarrow[n\to\infty]{} -\infty.$$

Now, observe that for $x \in (-1, 0]$, we have that $\frac{x}{2} \leq \log(1 + x) \leq \frac{2x}{2+x}$, which can be readily checked with calculus and a simple computation of derivatives. Using this inequality on each term of the sum above, we immediately deduce the result of the proposition. □

Now we move to the results in Hilbert space necessary for the original proof of Müntz and Szász. We will not give proofs of the most classical results in Hilbert space, which can be found in Folland, or Rudin's real analysis textbooks, for example.

Theorem 2.2 (Hilbert's projection theorem)**.** *Let $\mathcal{H}$ be a Hilbert space, and $M \subseteq \mathcal{H}$ be a closed linear subspace. Then, for all $f \in \mathcal{H}$, there exists a unique $g \in M$ such that $\inf\{||f - m|| : m \in M\} =: d(f, M) = ||f - g||$. Moreover, $f - g \in M^{\perp} = \{v \in \mathcal{H} : \langle v, m\rangle = 0$ for all $m \in \mathcal{H}\}$.*

Note that on a Hilbert space $\mathcal{H}$, $||v|| = \langle v, v\rangle^{1/2}$, where $\langle \cdot, \cdot\rangle$ is the inner product. We also remark that the unique $g \in M$ in the theorem above is called the *orthogonal projection of f onto M*.

Next we prove a result of Gram that allows us to compute distances between elements on a Hilbert space and finite dimensional subspaces (which are automatically closed).

Lemma 2.3 (Gram)**.** *Let $\mathcal{H}$ be a Hilbert space, $g \in \mathcal{H}$ and $V = Span\{f_1, \dots, f_n\}$ be a subspace of $\mathcal{H}$ with $\dim V = n$. Then,*

$$d(g,V)^2 = \frac{G(f_1,\dots,f_n,g)}{G(f_1,\dots,f_n)},$$

where $G(h_1, \dots h_r)$ is the determinant of the Gram matrix associated to the elements $h_1, \dots, h_r \in \mathcal{H}$. Namely,

$$G(h_1,\dots,h_r) := \det \begin{pmatrix} \langle h_1, h_1\rangle & \dots & \langle h_1, h_r\rangle \\ \vdots & & \vdots \\ \langle h_r, h_1\rangle & \dots & \langle h_r, h_r\rangle \end{pmatrix}$$

Observe that if the f_i's are linearly independent, then $G(f_1, \dots, f_n) \neq 0$.

Proof: By Hilbert's projection theorem, we can find the unique $f^* \in V$ satisfying $d(g,V)^2 = \langle g - f^*, g - f^*\rangle$. By assumption on V, there exist some uniquely define $c_1, \dots, c_n \in \mathbb{C}$ such that $f^* = \sum_{i=1}^n c_i f_i$. Recall that by Hilbert's projection theorem, $g - f^* \in V^\perp$, so we must have that the following equations hold:

$$\langle f_k, g - f^*\rangle = 0 \text{ for all } k = 1, \dots, n.$$

Moreover, we know that $d(g,V)^2 = \langle g, g\rangle - \langle f^*, g\rangle$, using bilinearity of $\langle\cdot,\cdot\rangle$ and the fact that $g - f^* \in V^\perp$. If we expand f^* in these equations, we obtain the following system:

$$\begin{aligned} c_1\langle f_1, f_1\rangle + \dots + c_n\langle f_1, f_n\rangle - \langle f_1, g\rangle &= 0 \\ \vdots \qquad\qquad \vdots \qquad & \\ c_1\langle f_n, f_1\rangle + \dots + c_n\langle f_n, f_n\rangle - \langle f_n, g\rangle &= 0 \\ c_1\langle f_1, g\rangle + \dots + c_n\langle f_n, g\rangle + d(g,V)^2 - \langle g, g\rangle &= 0 \end{aligned}$$

Therefore, since the vector $(c_1, \dots, c_n, 1) \neq \vec{0}$, it must be the case that

$$\det \begin{pmatrix} \langle f_1, f_1\rangle & \dots & \langle f_1, f_n\rangle & -\langle f_1, g\rangle \\ \vdots & & \vdots & \vdots \\ \langle f_1, g\rangle & \dots & \langle f_n, g\rangle & d(g,V)^2 - \langle g, g\rangle \end{pmatrix} = 0.$$

Splitting the last column as $(0, \dots, 0, d(g,V)^2) - (\langle f_1, g\rangle, \dots, \langle f_n, g\rangle, \langle g, g\rangle)$ and recalling the definition of the Gram determinant, we deduce that

$$d(g,V)^2 G(f_1, \dots, f_n) = G(f_1, \dots, f_n, g),$$

which completes the proof. □

We will use this theorem for the space
$L^2([0,1]) := \{f : [0,1] \to \mathbb{C} \text{ measurable } : \int_0^1 |f(x)|^2 \, dx < \infty\}$, where dx denotes the Lebesgue measure. To do this, we need

Theorem 2.4. *The space $L^2([0,1], dx)$ is a Hilbert space with the inner product*

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)} \, dx.$$

Observe that $L^2([0,1])$ is the space of measurable functions that satisfy $||f||_{L^2}^2 = \langle f, f\rangle < \infty$, as defined above.

Moreover, polynomials (and, in particular, all continuous functions) are dense in $L^2([0,1])$ *with the* L^2 *norm* $||\cdot||_{L^2}$*.*

Finally, we will need a result of Cauchy that will help us evaluate relevant determinants appearing in $G(f_1,\ldots,f_n)$ in the sequel:

Lemma 2.5 (Cauchy)**.** *Let* $n \in \mathbb{N}$ *and* $x_1,\ldots,x_n,y_1,\ldots,y_n \in \mathbb{C}$ *be such that* $x_i + y_j \neq 0$ *for all* i,j*. Then,*

$$D_n := \det \begin{pmatrix} \frac{1}{x_1+y_1} & \cdots & \frac{1}{x_1+y_n} \\ \vdots & & \vdots \\ \frac{1}{x_n+y_1} & \cdots & \frac{1}{x_n+y_n} \end{pmatrix} = \frac{\displaystyle\prod_{1\le i<j\le n}(x_j-x_i)(y_j-y_i)}{\displaystyle\prod_{1\le i,j\le n}(x_i+y_j)}.$$

We will offer two proofs of this lemma:

Proof 1: We follow the exposition in [2] for the first proof. First we obtain common denominators on each column of the determinant D_n, so that we can write

$$D_n = \frac{P_n}{\displaystyle\prod_{1\le i,j\le n}(x_i+y_j)},$$

where by the definition of determinant, P_n is a polynomial in $\mathbb{C}[x_1,\ldots,x_n,y_1,\ldots,y_n]$ of degree $n(n-1)$. Observe that if $x_i = x_j$ for $i \neq j$, or if $y_i = y_j$, then either two rows or two columns will coincide, so that $D_n = 0$. We conclude from this that $\prod_{1\le i<j\le n}(x_j-x_i)(y_j-y_i)$ must divide the polynomial D_n. Since the degrees coincide, there must exist a constant c_n depending only on n such that

$$D_n = c_n \frac{\displaystyle\prod_{1\le i<j\le n}(x_j-x_i)(y_j-y_i)}{\displaystyle\prod_{1\le i,j\le n}(x_i+y_j)}.$$

Now notice that $c_1 = 1$, but also that

$$\lim_{x_n\to\infty} x_n D_n = \det \begin{pmatrix} \frac{1}{x_1+y_1} & \cdots & \frac{1}{x_1+y_n} \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$

Similarly, we can deduce that $\lim_{y_n\to\infty}\lim_{x_n\to\infty} x_n D_n = D_{n-1}$. It also follows that

$$\frac{a_n D_n}{D_{n-1}} = \frac{c_n}{c_{n-1}} \frac{a_n \displaystyle\prod_{j=1}^{n-1}(x_n-x_j)\prod_{j=1}^{n-1}(y_n-y_j)}{\displaystyle\prod_{j=1}^{n}(x_n+y_j)\prod_{j=1}^{n-1}(y_n+x_j)}$$

so that taking the limit $a_n \to \infty$ and $b_n \to \infty$ shows that $c_n = c_{n-1}$, which implies that $c_n = 1$ for all $n \in \mathbb{N}$, completing the proof. $\square$

Proof 2: We offer another proof using only elementary row and column operations. First, observe that using the same first step as in Proof 1, we can write

$$D_n = \frac{1}{\prod\limits_{1\le i,j\le n}(x_i+y_j)} \det \begin{pmatrix} \prod_{i\neq 1}(x_i+y_1) & \dots & \prod_{i\neq 1}(x_i+y_n) \\ \vdots & & \vdots \\ \prod_{i\neq n}(x_i+y_1) & \dots & \prod_{i\neq n}(x_i+y_n) \end{pmatrix}.$$

So performing the row operations $R_1 - R_2, \dots, R_{n-1} - R_n$ in the order they appear and factoring on the rows 1 through $n-1$ we find that

$$D_n = \frac{\prod\limits_{i=2}^{n}(x_i - x_{i+1})}{\prod\limits_{1\le i,j\le n}(x_i+y_j)} \det \begin{pmatrix} \prod_{i\neq 1,2}(x_i+y_1) & \dots & \prod_{i\neq 1,2}(x_i+y_n) \\ \vdots & \dots & \prod_{i\neq 1,2}(x_i+y_n) \\ \prod_{i\neq n-1,n}(x_{n-1}+y_1) & \dots & \prod_{i\neq n-1,n}(x_{n-1}+y_n) \\ \prod_{i\neq n}(x_i+y_1) & \dots & \prod_{i\neq n}(x_i+y_n) \end{pmatrix}.$$

Repeating similar procedures now leaving the $n-1$-th row intact, then the $n-2$-th, and so on, we get to

$$D_n = \frac{\prod\limits_{1\le i<j\le n}(x_j - x_i)}{\prod\limits_{1\le i,j\le n}(x_i+y_j)} \det \begin{pmatrix} 1 & \dots & 1 \\ (x_1+y_1) & \dots & (x_1+y_n) \\ \vdots & & \vdots \\ \prod_{i\neq n}(x_i+y_n) & \dots & \prod_{i\neq n}(x_i+y_n) \end{pmatrix}.$$

Now, applying the operations $R_n - R_{n-1}(x_{n-1}+y_1), \dots, R_2 - (x_1+y_1)$ in this order, we get to

$$D_n = \frac{\prod\limits_{1\le i<j\le n}(x_j - x_i)}{\prod\limits_{1\le i,j\le n}(x_i+y_j)} \det \begin{pmatrix} 1 & 1 & & \dots & & 1 & \\ 0 & (y_2-y_1) & & \dots & & (y_n-1) & \\ \vdots & & & & & \vdots & \\ 0 & (y_2-y_1) & \prod_{i=1}^{n-2}(x_i+y+2) & & \dots & & (y_n-y_1)\prod_{i=1}^{n-2}(x_i+y_n) \end{pmatrix}.$$

Thus, using basic properties of determinants, and repeating the argument for the smaller determinant we are left with, just after the final step on the display, we get the result. □

3. Golitschek's elementary proof of sufficiency

In this section we provide an elementary proof of one direction of the Müntz-Szász theorem in the version stated in the introduction. It is due to M.V. Golitschek, and can be found in [1].

Theorem 3.1 (Müntz-Szász). *Let $\{0 = \lambda_0 < \lambda_1 < \dots\}$ be a sequence of real numbers satisfying $\lim_{i\to\infty}\lambda_i = \infty$. Then, $Span_{\mathbb{C}}\{x^{\lambda_i} : i \in \mathbb{N}\cup\{0\}\}$ is dense in $C([0,1])$ if and only if*

$$\sum_{i=1}^{\infty}\frac{1}{\lambda_i} = \infty.$$

Proof: First observe that by Weierstrass's theorem, it suffices to show that for any $q \in \mathbb{N}\setminus\{\lambda_i : i \in \mathbb{N}\}$, we have that x^q is in the norm closure of $\mathrm{Span}_{\mathbb{C}}\{x^{\lambda_i} :$

$i \in \mathbb{N} \cup \{0\}\}$. We will show this constructively, as follows. We let $Q_0(x) := x^q$ and define $Q_n(x)$ for $0 < x \leq 1$ for $n \geq 1$ inductively by

$$Q_n(x) := (\lambda_n - q)x^{\lambda_n} \int_x^1 Q_{n-1}(t)t^{-1-\lambda_n}\, dt, \text{ for } n \geq 1.$$

We claim that for each $n \geq 0$ we can find $a_{n,i} \in \mathbb{R}$ satisfying

$$Q_n(x) = x^q - \sum_{i=1}^{n} a_{n,i} x^{\lambda_i}$$

Indeed, by induction on n, this clearly holds by definition when $n = 0$. Suppose now that it holds for $n > 0$. Then, by the inductive hypothesis, we can write

$$Q_{n+1}(x) = (\lambda_{n+1} - q)x^{\lambda_{n+1}} \int_x^1 \left(t^q - \sum_{i=1}^{n} a_{n,i} t^{\lambda_i} \right) t^{-1-\lambda_{n+1}}\, dt,$$

so integrating and using Barrow's rule yields

$$Q_{n+1}(x) = (\lambda_{n+1} - q)x^{\lambda_{n+1}} \left[\frac{t^{q-\lambda_{n+1}}}{q - \lambda_{n+1}} - \sum_{i=1}^{n} a_{n,i} \frac{t^{\lambda_i - \lambda_{n+1}}}{\lambda_i - \lambda_{n+1}} \right]_x^1$$

$$= x^q + \sum_{i=1}^{n} a_{n,i} x^{\lambda_i} \frac{\lambda_{n+1} - q}{\lambda_i - \lambda_{n+1}} + x^{\lambda_{n+1}}(\lambda_{n+1} - q) \left(\frac{1}{q - \lambda_{n+1}} - \sum_{i=1}^{n} a_{n,i} \frac{1}{\lambda_i - \lambda_{n+1}} \right).$$

Moreover, we have that $||Q_0||_{\sup} = 1$ and

$$|Q_n(x)| \leq |\lambda_n - q| ||Q_{n-1}||_{\sup} |x^{\lambda_n}| \left| \int_x^1 t^{-1-\lambda_n}\, dt \right| \leq \left| \frac{\lambda_n - q}{\lambda_n} \right| |x^{\lambda_n} - 1| ||Q_{n-1}||_{\sup}.$$

Therefore, taking supremums and applying this inequality repeatedly, we deduce that

$$||Q_n(x)||_{\sup} \leq \prod_{i=1}^{n} \left| 1 - \frac{q}{\lambda_i} \right|.$$

Thus, by Proposition 2.1, it follows that $||Q_n|| \to 0$ provided $\sum_{i=1}^{\infty} (1/\lambda_i) = \infty$. Putting all the properties of the Q_n's together, the proof of sufficiency follows. It is worth noting that if we let $q = 1$ and $\lambda_i = 2i$ we can obtain an approximation of $|x|$ on $[-1, 1]$ as is needed in the proof of Stone-Weierstrass, by polynomials in x.

4. Original proof by Müntz and Szász

We now proceed to give the full original proof of Müntz-Szász, with the extra conditions as in the introduction. First we need

Lemma 4.1. *Let $q, \lambda_0, \ldots, \lambda_n$ be distinct real numbers strictly above $-1/2$. Then, with the Hilbert space norm, we have that*

$$\delta = d(x^q, Span_{\mathbb{C}}\{x^{\lambda_0}, \ldots, x^{\lambda_n}\}) = \frac{1}{\sqrt{2q+1}} \prod_{i=0}^{n} \left| \frac{q - \lambda_i}{q + \lambda_i + 1} \right|.$$

We remark that this can be generalized to complex numbers, but we will not need that.

PROOF: As noted, $L^2([0,1])$ is a Hilbert space, and the functions $x^{\lambda_0}, \dots, x^{\lambda_n}$ span a finite dimensional vector space. Thus, by Gram's lemma, the distance we wish to compute is equal to

$$\delta = \sqrt{\frac{G(x^{\lambda_0}, \dots, x^{\lambda_n}, x^q)}{G(x^{\lambda_0}, \dots, x^{\lambda_n})}}.$$

In order to compute the determinants above, we observe that for any $a, b > -1/2$ we have

$$\langle x^a, x^b \rangle_{L^2} = \int_0^1 x^a x^b \, dx = \frac{1}{a+b+1}.$$

Thus, we obtain the equalities

$$G(x^{\lambda_0}, \dots, x^{\lambda_n}) = \det \begin{pmatrix} \frac{1}{\lambda_0+\lambda_0+1} & \cdots & \frac{1}{\lambda_0+\lambda_n+1} \\ \vdots & & \vdots \\ \frac{1}{\lambda_n+\lambda_0+1} & \cdots & \frac{1}{\lambda_n+\lambda_n+1} \end{pmatrix}$$

and

$$G(x^{\lambda_0}, \dots, x^{\lambda_n}, x^q) = \det \begin{pmatrix} \frac{1}{\lambda_0+\lambda_0+1} & \cdots & \frac{1}{\lambda_0+\lambda_n+1} & \frac{1}{\lambda_0+q+1} \\ \vdots & & & \vdots \\ \frac{1}{\lambda_n+\lambda_0+1} & \cdots & \frac{1}{\lambda_n+\lambda_n+1} & \frac{1}{\lambda_0+q+1} \\ \frac{1}{q+\lambda_0+1} & \cdots & \frac{1}{q+\lambda_n+1} & \frac{1}{q+q+1} \end{pmatrix}.$$

Consequently, using Cauchy's lemma with $x_i = \lambda_i$ and $x_j = \lambda_j + 1$, and $x_{n+2} = q$, $y_{n+2} = q + 1$ for the second case and simplifying, we get the original claim. □

Onto the original proof of

Theorem 4.2 (Müntz-Szász). *Let $\{0 = \lambda_0 < \lambda_1 < \dots\}$ be a sequence of real numbers satisfying $\lim_{i\to\infty} \lambda_i = \infty$. Then, $Span_{\mathbb{C}}\{x^{\lambda_i} : i \in \mathbb{N} \cup \{0\}\}$ is dense in $C([0,1])$ if and only if*

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty.$$

PROOF: We begin by recalling that for $f \in C[0,1]$ we have

$$||f||_{L^2} = \int_0^1 |f(x)|^2 \, dx \leq ||f||_{\sup},$$

which shows that a subset being dense in $C([0,1])$ implies that it is dense in $L^2([0,1])$. Therefore, the span of x^{λ_i}'s being dense implies that x^q is in the L^2-closure of this span, for any $q \in \mathbb{N} \setminus \{\lambda_i : i \in \mathbb{N} \cup \{0\}\}$. This can happen if and only if

$$d(x^q, \mathrm{Span}_{\mathbb{C}}\{x^{\lambda_0}, \dots, x^{\lambda_n}\}) \xrightarrow[n\to\infty]{} 0.$$

Since $\lim_{j\to\infty} \lambda_j = \infty$, this will happen if and only if

$$\limsup_{n\to\infty} \prod_{j=0}^{n} \left| \frac{q - \lambda_j}{q + \lambda_j + 1} \right| = 0.$$

Now, for j big enough, we can write the product (except for the first few factors) as $\prod(1-\frac{2q+1}{\lambda_j+q+1})$, which we know converges to 0 if and only if

$$\sum_{j=1}^{\infty}\frac{2q+1}{\lambda_j+q+1}=\infty.$$

Observe that given $a>0$, if λ_j is big enough, the following inequalities are trivially true:

$$\frac{2}{\lambda_j+a}\geq\frac{1}{\lambda_j}\geq\frac{1}{\lambda_j+a},$$

so that the condition above happens if and only if

$$\sum_{j=1}^{\infty}\frac{1}{\lambda_j}=\infty.$$

This shows the first direction. For the other one, using similar inequalities for λ_j, notice that if $q\geq 1$, we can write, for any given $i_0\in\mathbb{N}$

$$\left|x^q-\sum_{i=i_0}^{r}a_ix^{\lambda_i}\right|=\left|\int_0^x\left(qt^{q-1}-\sum_{i=i_0}^{r}a_i\lambda_it^{\lambda_i-1}\right)\,dt\right|\leq\left|\left|qx^{q-1}-\sum_{i=i_0}^{r}a_i\lambda_ix^{\lambda_i-1}\right|\right|_{L^2}$$

using Jensen's inequality, and increasing the area (from x to 1). Therefore, we can approximate in the uniform norm the polynomial x^q as well as we like, with elements of $\mathrm{Span}_{\mathbb{C}}\{x^{\lambda_i}:i\in\mathbb{N}\cup\{0\}\}$ provided that

$$\sum_{i=i_0}^{\infty}\frac{1}{\lambda_i-1}=\infty.$$

Note that we can pick i_0 so that $\lambda_i-1>0$ for all $i>i_0$. Thus, using the if and only if result for $L^2([0,1])$ we proved above for arbitrary positive sequences, together with the final observation that the equality for the sum above happens if and only if we remove the -1 from it, completes the proof. $\square$

5. Further extensions and comments

We end these notes by pointing out that several other proofs of Müntz-type results exist in the literature. To name a few, one can do it through a method of measures and zeros of analytic functions, as well as a method of divided differences (due to work by Feller, Hirschmann, Widder and Gelfond (published in 1968)). Good surveys on these include the book [3], which includes a lot of history and particular cases of this Theorem, as well as many generalizations, and proofs via polynomial inequalities, and [4], which also discusses some generalizations to countable compact sets.

Next we state a generalization of the result we presented due to Borwein and Erdélyi:

Theorem 5.1 (Borwein, Erdélyi). *Let $(\lambda_i)_{i\in\mathbb{N}}$ be a sequene of distinct, positive real numbers. Then, $Span_{\mathbb{C}}\{1,x^{\lambda_i}:i\in\mathbb{N}\}$ is dense in $C([0,1])$ if and only if*

$$\sum_{i=1}^{\infty}\frac{\lambda_i}{\lambda_i^2+1}=\infty.$$

For a proof, we refer to [5]. In another paper, the same authors generalize this result to compact subsets of $[0, \infty)$ with positive Lebesgue measure.

Replacing the interval $[0, 1]$ by a general interval $[a, b]$ is non-trivial, and some partial results were established by Clarkson, Erdős and Schwartz, using complex analysis methods. Formulations of the theorem with complex exponents were explored by Szász in his paper in 1916, but a complete characterization was not given. Namely, he showed:

Theorem 5.2 (Szász, 1916)**.** *Let $(\lambda_j)_{j\geq 0}$ be complex numbers. If $\lambda_0 = 0$ and $Re(\lambda_i) > 0$ for all $i > 0$, then $Span_{\mathbb{C}}\{x^{\lambda_i} : i \geq 0\}$ is dense in $C([0, 1])$ whenever*

$$\sum_{i=1}^{\infty} \frac{Re(\lambda_i)}{1 + |\lambda_i|^2} = \infty.$$

Moreover, if

$$\sum_{i=1}^{\infty} \frac{1 + Re(\lambda_i)}{1 + |\lambda_i|^2} < \infty,$$

the density result no longer holds. In particular, if

$$\liminf_{i\to\infty} Re(\lambda_i) > 0,$$

then the density result holds if and only if the first sum considered is infinite.

Finally, we mention that a version of the result holds for nilpotent Lie groups. We refer to [6].

References

1. Golitschek, M.V.; *A short proof of Müntz's Theorem*, Journal of Approximation Theory **39**, 394-395 (1983).
2. Achieser, N.I.; *Theory of Approximation*, Dover, Mineola, New York (2003)
3. Borwein, P., Erdélyi, T.; *Polynomials and Polynomial Inequalities*, Springer, Graduate Texts in Mathematics, vol. 161, New York, NY, (1995)
4. Almira, J.M.; *Müntz Type Theorems I*, Surveys in Approximation Theory, **3**, pp. 152-194 (2007)
5. Borwein, P., Erdélyi, T.; *The Full Müntz-Szász Theorem in $C[0, 1]$ and $L_1(0, 1)$*, J. London Math. Soc. **54** 102-110 (1996)
6. Cook, C. D.; *Szász-Müntz Theorems for Nilpotent Lie Groups.* (1996). LSU Historical Dissertations and Theses. 6328. `http://digitalcommons.lsu.edu/gradschool_disstheses/6238`